Human-Agent Knowledge Fusion: Collaborative sensemaking with explainable and tellable AI

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy

David Braines

June 2024

Cardiff University

School of Computer Science & Informatics

Abstract

Augmenting human cognitive activities with Artificial Intelligence (AI) powered machine agents shows promising potential, with new cloud services released regularly. However, rapidly using these in traditional applications requires technical skills beyond typical users. Developers build or extend applications to harness these services, often delaying availability to these users. Chatbot-style conversational interfaces attempt to address this but favour simple interactions. To support richer solutions, I propose knowledge sharing through co-construction of task-relevant information between humans and machine agents. Specifically, shared knowledge supporting multiple modalities and a range of specificity, from rapidly foraged and fluid information to more formally defined knowledge. Moreover, users should be able to invoke relevant cloud services, quickly establishing a level of trust appropriate to those services. By fusing knowledge through co-construction, we can move beyond simple conversational interactions or bespoke applications common for machine agent integration today, enabling faster and richer collaboration mechanisms.

This thesis introduces Human-Agent Knowledge Fusion (HAKF) as a conceptual framework to support co-construction of multi-modal knowledge, and support human-agent teams in task-specific and time-constrained problem-solving activities. Specifically, HAKF highlights the need for explainable AI to establish trust rapidly, and tellable AI for fluid knowledge exchange. An open-source instantiation of HAKF, Cogni-sketch, is defined, enabling experimentation for: (1) human-led information foraging, sensemaking and storytelling for open source intelligence analysis, and (2) information fusion from machine agents and data feeds, alongside human analysts. Results from (1) show that users successfully completed the task, concurrently progressing multiple sensemaking activities. Results from (2), featuring fusion of machine vision and object identification, demonstrate co-construction of knowledge from machine agents for consumption by human users.

Through HAKF and Cogni-sketch I show the potential for powerful but flexible solutions, enabling task-relevant problem-solving activities between human and machine agents, ranging from information gathering and organisation through sensemaking and storytelling.

Contents

List of Figures

List of Tables

List of Acronyms

HAKF	Human-Agent Knowledge Fusion. iii–vi, ix, x, xviii, xxi, xxiii–xxix, xxxi, 6–12, 16–21, 26, 58, 60, 63, 64, 67, 68, 75, 76, 78–99, 101–103, 105–115, 121, 130, 131, 133–135, 145, 152–154, 160, 165, 168–171, 174–176, 181–184, 214, 237, 239–244, 246, 250, 252, 278
HAT	Human-Agent Teaming. iv, xix, xxv, 5, 6, 9, 18, 20–23, 28–31, 39, 45, 49, 57, 59, 61, 62, 64, 65, 75, 79, 82, 89, 91, 99, 154, 168, 183, 241, 255
HCI	Human-Computer Interaction. 24, 32, 42
ITA	International Technology Alliance. xxii, xxvii, xxx, xxxi, 4, 134, 239, 278, 279, 284
JSON	JavaScript Object Notation. 122, 126, 253, 286, 309
ML	Machine Learning. xxviii, 2, 22, 33, 35, 39, 40, 45, 58–61, 79, 84, 89, 94, 136, 162, 167, 278
MVP	Minimum Viable Product. 112, 246, 247
NER	Named Entity Recognition. 47, 116, 186
NIS	Network and Information Science. xxx, 279
NLP	Natural Language Processing. 116, 139, 177, 186

OSINT Open Source Intelligence. x, xxv, xxxi, 6, 18, 48, 49, 62, 66, 70, 97, 174, 184–186, 188, 197, 199, 201, 202, 212, 228, 230, 237, 243, 246, 284, 286, 289

RAG Retrieval Augmented Generation. 254

SA Situation Awareness. 40, 47, 48, 61, 163

SDK Software Development Kit. 107, 115, 133, 149, 157, 159, 171

SME Subject-Matter Expert. 17, 61, 68, 69, 79

SSU Shared Situation Understanding. 47, 48

SU Situation Understanding. xviii, xxi, 5, 6, 8, 11, 13, 15–17, 20, 40, 47, 48, 56, 65, 79, 83, 97, 103, 108, 110, 111, 133–136, 139, 140, 143, 151, 153–155, 170, 242, 244, 245, 278

SUS System Usability Scale. 175, 214, 215, 222, 227–231, 238, 244

UI User interface. ix, 23, 38, 95, 113, 116, 117, 121, 127, 144, 205, 248, 254

VUI Voice user interface. 1

XAI Explainable Artificial Intelligence. xxiv, 10, 17, 20–22, 32–40, 61–63, 65, 66, 68, 69, 71, 74, 79, 85, 97, 150, 160, 161, 168, 241, 246

Glossary

Affordance	An affordance is one or more qualities or properties for an object that defines its possible uses or makes clear how it can or should be used. In the context of human users and machine agents these affordances are defined as typical characteristics or capabilities for human beings when compared to computer processes, and vice-versa [39].
Artificial Intelligence	From the many definitions of Artificial Intelligence (AI), one that particularly resonates with the content of this thesis is: "Artificial Intelligence (AI) is the automation of activities that we associate with human thinking, activities such as decision-making, problem-solving, learning, and the study of the computations that make it possible to perceive, reason, and act" ([12]).

Coalition Situation Understanding	Coalition Situation Understanding (CSU) is Situation Understanding (SU), extended to a coalition environment where "The purpose and actions of a coalition are contingent on gaining and maintaining models of an environment and events" ([115]).
Cogni-sketch	An experimental platform for exploring different aspects of Human-Agent Knowledge Fusion (HAKF) with human users and machine agents. Defined as part of this research (in Chapter 4) and applied in various use cases within this thesis.
Explainability	A flow within Human-Agent Knowledge Fusion (HAKF) that provides a greater level of transparency into a conclusion or output from a machine agent or human user. Amongst human users this is a familiar concept and is often invoked through *why?* questions and appropriate responses [21].
Explainable Artificial Intelligence	Explainable Artificial Intelligence (XAI) systems "Deliver accompanying evidence or reasons for outcomes and processes; provide explanations that are understandable to individual users; provide explanations that correctly reflect the system's process for generating the output..." ([111]).

Human-Agent Knowledge Fusion	Human-Agent Knowledge Fusion (HAKF) is a Human-Agent Teaming (HAT) conceptual architecture that supports co-construction of task-relevant knowledge, allowing all agents to contribute and consume task-relevant information and knowledge [21]. Defined within this thesis (in Chapter 3) and formalised as the motivating basis for Cogni-sketch.
Human-Agent Teaming	Human-Agent Teaming (HAT), also known as Human-Machine Teaming (HMT), is "A relationship—one made up of at least three equally important elements: the human, the machine, and the interactions and interdependencies between them" ([78]). (This abbreviation is used for Human-Agent Team and Human-Agent Teaming).
LLM	A Large Language Model (LLM) is a very large transformer-based model trained on large volumes of unlabelled text using self-supervised methods.
Machine Learning	Machine Learning (ML) is the field of study that gives computers the ability to learn without being explicitly programmed [129].

Natural Language Processing	Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things [35].
Open Source Intelligence Analysis	Open Source Intelligence (OSINT) Analysis is "The collection, processing, analysis, production, classification, and dissemination of information derived from sources and by means openly available to and legally accessible and employable by the public in response to official national security requirements" ([131]).
Sensemaking	Sensemaking is an interconnect set of tasks that "consist of information gathering, re-representation of the information in a schema that aids analysis, the development of insight through the manipulation of this representation, and the creation of some knowledge product or direct action based on the insight" ([113]). More literally it is the act of making sense of an environment, usually through the analysis and consumption of various data sources.

Shared Situation Understanding	Shared Situation Understanding (SSU) is a state in which multiple agents seek a common Situation Understanding (SU). It involves aligning the mental models of different agents [140].
Situation Awareness	Situation Awareness (SA) is the "perception of the elements in the environment within volume of space and time, the comprehension of their meaning and the projection of their status in the near future" ([47]).
Situation Understanding	Situation Understanding (SU) is an ability to "explain how the current situation, or elements thereof, came to be as they are, and it often involves an additional ability to predict how the current situation may develop or evolve in the future" ([140]).
Tellability	A flow within HAKF where new information is conveyed from one of the human users or machine agents, often to impart useful local or task-relevant knowledge [21].

List of Publications

The research reported in this thesis in based on the previously published journal and conference papers listed below. Many of these can be downloaded at the Distributed Analytics and Information Science (DAIS) International Technology Alliance (ITA) Science Library[1]. See Section 1.4 for the definitions of the research contributions for this thesis and a mapping of those contributions to the publications listed below.

Primary Articles

The six publications listed here are those that primarily contribute to the material in this thesis. All are collaborative publications that I have led and provide material for much of the core content of this thesis. A brief description is given for each paper, outlining how that published work contributes to this thesis, and in which chapter(s) the material can be found. My contributions to each publication are defined in the summary for each of these papers.

P1. *D. Braines, A. Preece, and D. Harborne. (2018).* ***Multimodal Explanations for AI-based Multisensor Fusion****. In NATO SET-262, 2018.* [26] This paper defines a series of conversational explanations which provide worked examples for different types of explanations from machine agents (some visual, some textual, and some that disagree across the modalities).

[1]The DAIS ITA Science Library is a list of all publications from the DAIS ITA research program and can be found at `https://dais-legacy.org/science-library/`

It also motivates the need to explicitly account for the different roles of human users within such systems, and the different considerations that apply to each. This early work provided useful examples to meet specific human-agent interaction needs but limits the interaction type to textual conversation. This led to the insight that a broader interpretation could give more flexibility and extensibility and, along with [29] (paper P3), identified the need for an environment such as Cogni-sketch, based on the emerging HAKF concept. I created the initial implementation of the conversational solution that is the basis for this paper, along with the conceptual models to support the solution, whilst my co-authors created the components that generated the traffic-related inputs to the system. This work is reported in Chapter 5 (Section 5.4) of this thesis.

P2. *D. Braines, R. Tomsett, and A. Preece. (2019)* ***Supporting User Fusion of AI Services through Conversational Explanations****. In 22nd International Conference on Information Fusion (Fusion). IEEE.* [29]
Building on paper P1, this paper summarises early work exploring the potential for rich and interactive explanations from machine agents and how they can be served in the form of a textual conversation. This directly inspired the formalisation of HAKF as a broader concept to enable such capabilities to be more readily defined and implemented without being limited to only conversational interactions. The work builds on the earlier North Atlantic Treaty Organisation (NATO) publication [26] (paper P1) and aligns heatmap explanations of images to a simple conversational model and, in turn, to five different levels of a common model for information fusion. In addition to the contributions from paper P1, I led the mapping to the different levels of the information fusion model and defined example questions for different user types within the system. Relevant aspects of this work are reported in Chapter 5 (Section 5.4) of this thesis.

P3. *D. Braines, E. Lee, G. Pearson, and A. Preece. (2019).* ***Exploring the future of Explainable AI solutions with military stakeholders****. In Proceedings of the 3rd Annual Fall Meeting of the DAIS ITA, 2019.* [22]
This paper summarises the planning, execution and results of a Design Thinking (DT) workshop on the topic of Artificial Intelligence (AI) and Explainable Artificial Intelligence (XAI) that was held with military representatives from across the U.K. armed forces. The purpose of the workshop was to identify perceived needs of AI assistants in future operational settings and was an important motivation for a unifying and underlying conceptualisation such as HAKF, most notably for the human users and machine agents to be able to operate more fluidly. I co-led the workshop with my co-authors, transcribed the results and led the post-workshop topical analysis. This work is briefly outlined in Chapter 3 (Section 3.2) and Appendix C along with some of the key findings.

P4. *D. Braines, F. Cerutti, M. R. Vilamala, M. Srivastava, L. Kaplan, A. Preece, and G. Pearson. (2020).* ***Towards Human-Agent Knowledge Fusion (HAKF) in support of distributed coalition teams****. In AAAI Fall Symposium Series, AI in Government & Public Sector, 2020.* [21]
This paper introduces the concept of HAKF and defines the tellability and explainability flows, motivating their value for human users and machine agents. It provides a worked example of HAKF instantiated in an early version of the Cogni-sketch environment that builds on other collaborative research [146], demonstrating dynamic event definition and detection in a multi-modal sensor processing environment with a hybrid team of human users and machine agents. My co-authors created each of the services that are integrated, whilst I embedded them into the Cogni-sketch environment and aligned their inputs and outputs to the HAKF tellability and explainability flows to demonstrate knowledge co-construction in a multi-agent setting. This can be found in Chapter 5 (Section 5.2) of this thesis.

P5. *D. Braines, A. Preece, C. Roberts, and E. Blasch. (2021).* ***Supporting Agile User Fusion Analytics through Human-Agent Knowledge Fusion****. In 24th IEEE International Conference on Information Fusion.* [27]
This paper brings together two separate use cases to show the breadth and flexibility of HAKF and the support provided within the Cogni-sketch environment at this stage of my research. The first reports findings from a long-running pilot exercise with an OSINT analyst to use the environment for intelligence analysis and sensemaking and is reported in Chapter 6 (Section 6.3) of this thesis. The second use case within this paper is an evaluation involving real-time video processing and event detection capability aligned with definition of logical inference rules to explore Human-Agent Teaming (HAT) capabilities and interactions within Cogni-sketch and is reported in Chapter 5 (Section 5.3) of this thesis. I supported the analyst during the pilot OSINT exercise and analysed the results, and I built the agents for the real-time event processing, with my co-authors working closely with me as advisors and users.

P6. *D. Braines, and A. Preece.* ***Open Source Intelligence: Sensemaking evaluation for Human-Agent Knowledge Fusion****. 2024 (unpublished). [25]*
This paper reports in detail the formal experiment with twelve human subjects to evaluate the Cogni-sketch environment for sensemaking. Specifically, testing whether novice users can successfully undertake OSINT analysis and report their results by constructing task-relevant knowledge within the environment. The analysis of the user activities is based on a mapping of Pirolli and Card sensemaking loops [113] to Cogni-sketch events. Key extensions to the core environment to support this exercise are reported, along with details of the social media data collection that was carried out to provide a dataset for participant exploration. This paper is unpublished,

with my contribution being the definition and execution of the experiment followed by the three-way analysis of the results as reported in Chapter 6 (Section 6.4) of this thesis.

Related Articles

The five additional publications listed here are those which are relevant to this thesis but not considered primary. Again, my contributions to each publication are identified in the summary for each paper as well as a mention of the chapter(s) in which any content is included or mentioned. Unless otherwise stated, I only include within this thesis the material that I directly contributed. Generally, these papers report on example use cases that have been demonstrated within the Cogni-sketch environment, or relevant supporting research that informed the design of HAKF.

R1. *R. Tomsett, A. Preece, D. Braines, F. Cerutti, S. Chakraborty, M. Srivastava, G. Pearson, and L. Kaplan. (2020).* ***Rapid Trust Calibration through Interpretable and Uncertainty-Aware AI***. *Cell Press Patterns, Vol 1 Issue 4.* [150]
My contribution to this paper is focused on human-factors challenges for communicating contextually relevant information and appropriate metadata, including uncertainty. This includes recommendations for researchers, some of which were addressed when building the Cogni-sketch environment, as reported in Section 4.3.1. This paper also includes a variant of HAKF with a particular focus on communication of interpretation and uncertainty information in the explainability flow. *Rapid trust calibration* is covered in this thesis in both the background material (in Chapter 2) and with the relevant roles extended in Sections 3.3.4 and 4.3.1.

R2. *D. Braines, J. Stockdill-Mander, and E. Lee. (2020).* ***The Science Library: Curation and Visualization of a Science Gateway reposi-***

tory. *Concurrency and Computation: Practice and Experience: e6100.* [28]
This paper describes in detail the implementation of a science gateway for the DAIS ITA research program, and why provenance and other supporting information are key to establishing trust and confidence in the accuracy of information before publication. I created the implementation described in the paper with my co-authors providing requirements as well as subsequently using it to maintain the corpus of publications. Cogni-sketch is extended with a custom palette with meaningful semantics and a series of custom panes to achieve this task. The reported solution exercises the Cogni-sketch environment and provided a useful improvement to the knowledge management role for the team (the co-authors on this paper). This exercise provided more insight into required capabilities for tellability as mentioned in Section 3.4 and serves as a useful illustrative example of the intended flexibility. It is reported briefly as one of the examples in Section A.1.1 of this thesis.

R3. *A. Preece, D. Braines, F. Cerutti, J. Furby, L. Hiley, L. Kaplan, M. Law, A. Russo, M. Srivastava, M. R. Vilamala, and T. Xing. (2021)* ***Coalition Situational Understanding via Explainable Neuro-Symbolic Reasoning and Learning***. *In Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III, SPIE DCS, 2021.* [114]
This paper reports on a specific use case for human operator assistance in understanding fast moving multi-modal data sources. My contribution was in helping to define some of the details for this operational context and how the different information flows can be mapped to tellability, explainability and other required capabilities within HAKF. This work is briefly mentioned in Chapter 5 (Section 5.2) of this thesis, with a description of how typical machine agent assistance is provided for human users.

R4. *E. Blasch, T. Pham, C-Y. Chong, W. Koch, H. Leung, D. Braines and*

T. Abdelzaher. (2021). ***Machine Learning/Artificial Intelligence for Sensor Data Fusion – Opportunities and Challenges****. IEEE Aerospace and Electronic Systems Magazine (Volume: 36, Issue: 7).* [16]

In this paper I and my co-authors summarise our perspectives for data fusion systems and the role for AI or Machine Learning (ML) support within these kinds of environments. My contribution covers models of sensemaking and decision making (such as Observe, Orient, Decide, Act (OODA) and Direct, Collect, Process, Disseminate (DCPD)), motivates the need to explicitly consider the role of the human users, and describes how HAKF can support this. I also outline key social considerations for machine-generated explanations (mainly arising from [95]). Some of these aspects have already been reported in the primary articles listed earlier but, in particular, the discussion of social considerations for machine generated explanations as reported in this paper can be found in Section 5.1.

R5. *D. Millar, D. Braines, E. Blasch, D. Summers-Stay, and I. Barclay. (2021)* ***Semantically-guided acquisition of trustworthy data for information fusion****. In 24th International Conference on Information Fusion (Fusion). IEEE.* [94]

This paper shows the potential for using graph analytic methods, based on a semantic vector space, to identify potentially related entities within large complex graphs based on their structural similarity. A common issue with this technique is the lack of meaning that is articulated by the machine agent that performs this processing; it simply identifies clusters and can articulate structural features, but not what they mean in context. My contribution to this research was to show how the Cogni-sketch environment can be used to capture this important local knowledge from human users, as a form of co-creation with the machine agent processing. Overall, this can lead to improved explanations from the machine agent, by capturing human insight through the definition of named *meaningful paths* that correspond to relevant real-world impact that is captured in the structure of the graphs. This work serves as an example of a different kind of use case for both the explainability and tellability flows in HAKF and is reported as one of the examples in Section A.1.3 in this thesis as well as providing a clear motivation for rich knowledge representation in Section 3.4.1.

Summary

In summary, these six primary (P) and five related (R) publications provide material for a substantial portion of this thesis and are included in the following chapters:

Location	**Publications**
Chapter 3	P3, P4, R1, R2, R5
Chapter 4	P4, P5, P6, R2, R5
Chapter 5	P1, P2, P4, P5, R3, R4
Chapter 6	P5, P6, R2, R5

Table 3: Primary and related publications mapped to thesis chapters

Acknowledgements

I would like to start by thanking Dr Peter Waggett who has always supported my research and enabled me to pursue this PhD. His predecessor, Dr Dave Watson CBE, was also very supportive and continued to take an active interest after his retirement. For my International Business Machines (IBM) colleagues, a particular thank you to Helen Stanton and Sue Johnson for their valiant attempts to keep me focused over an extended period, whilst many other things were vying for my attention.

The 5-year DAIS ITA, and the previous 10-year Network and Information Science (NIS) ITA research programs, enabled me to work with many inspirational individuals from academia, industry and government. The seed to pursue a part-time PhD was sown during the latter half of the NIS ITA and grew directly out of our research investigating the role of Controlled Natural Language (CNL) in agile knowledge management. My guide and mentor in this was Dr David Mott from IBM Research U.K. and he, more than anyone else, directly inspired me towards pursuing a PhD. My collaborators also played an important role by exhibiting their deep knowledge and passion for pushing the boundaries of that knowledge through research. I was privileged to collaborate with Dr Paul Smart and others at Southampton University, Gavin Pearson, Peter Houghton, Simon Bray and others from Defence Science and Technology Laboratory (Dstl), and Dr Tien Pham, Dr Lance Kaplan and others from DEVCOM Army Research Laboratory (ARL). Dr Dinesh Verma from IBM Research U.S. was also a supporter of my work, providing valuable advice and mentoring. All these colleagues helped me develop my research grounding that would eventually lead to this PhD.

I must also single out Professor Alun Preece from Cardiff University who I first worked with investigating and applying CNL technology. As our research evolved into new fields through the DAIS ITA, he became my PhD supervisor as well as ongoing research collaborator, and fellow technical area leader for the DAIS ITA research program. His support and guidance to me have been invaluable over the last 10 or so years, and I look forward to continuing our collaboration into the future.

For the intelligence analysis aspects of this PhD, I had invaluable guidance and trade-craft advice from Dr Colin Roberts and Professor Martin Innes at the Cardiff Crime and Security Research Institute (CSRI). They were continually supportive of my efforts to create and release the Cogni-sketch platform, especially shaping its support for OSINT analysis. I was also privileged to collaborate with the late Paul Sullivan (Sully) who, in his role as an independent military advisor, brought substantial first-hand operational experience and lots of valuable practical insight relating to intelligence analysis and so much more. Also, to Erik Blasch from Air Force Research Laboratory (AFRL) who was both a research collaborator and co-author, as well as a supporter of this work, especially the HAKF approach and the ability to rapidly and flexibly support human users in evolving settings.

Finally, the biggest thanks are for the support from my family throughout this PhD, including finding our feet during a global pandemic with three children and two adults working and schooling from home. They have always been there to support me and make space for me to have deep thinking and writing time. Finding quality time to pursue my research, to design and run experiments, and to write up the results alongside a full-time job was very challenging and their support was invaluable.

This research was partly sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. In this respect, the views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Chapter 1

Introduction

The ability for human users and machine agents to interact and exchange information is already possible for limited interaction types such as question answering, and for basic text and audio modalities. These usually occur in closed settings with predefined outcomes such as providing requested information, applying analytic techniques to input data, or calling other services. For non-technical users the most commonly encountered examples are voice or chat assistants, or predefined functions in software products. As machine agent capabilities improve, the ability to take task-relevant factors into account and provide more dynamic services will become increasingly important [120].

Chat interfaces are a common example of human interaction with machine agents for non-technical people today, and some analysis of their capabilities and near-term potential are discussed here before identifying a richer basis for collaboration that must be supported to progress beyond these simple chat-based interactions to a deeper and more useful relationship between human users and machine agents.

Voice assistants (or smart speakers [147]) enable users to experience basic but useful interactions with machine agents [108]. They provide a Voice User Interface (VUI) [99] and are typically implemented based on simple turn-taking conversations. Users typically request direct actions from the assistant such as setting a timer or requesting music to be played. Whilst the domain of discourse is potentially large, the types of interaction are relatively few [121], even when considering extensions that can expand the reach of the agent but don't usually add new types of interaction [37]. However, sometimes a multi-turn dialogue is used to request additional information [2].

Even with these limitations, basic interactions such as these are useful, rendering the agents easy to understand and engage with, and driving adoption in common settings. However, misinterpretations or poorly judged interjections [155] can annoy or deter the human users, undermining their confidence, and reminding them of the simplicity and fragility of the experience [86].

Contrasting typical voice activated assistants with online text-based chatbots we find some cases where the complexity of the interaction can be increased, for example, with multi-turn dialogues to populate the textual equivalent of online forms [58]. In some cases, these chatbots can hand over to a human operator to handle more complex situations, providing the human operator with relevant details of the agent-based chat so far [77].

In addition, human users can also interact with machine agents within applications. These interactions will typically take the form of predefined functions such as an embedded Machine Learning (ML) feature to perform entity detection in pictures, or to propose text completion when writing an email. For developers an Integrated Development Environment (IDE) can provide powerful but highly specialised AI assistance from systems such as GitHub Copilot [158]. Systems such as these are prompted by the human developer user to produce source code proposals that can be accepted (typically after making modifications to account for context) into their codebase. This can increase the productivity of the human developer and provide a good education opportunity, assuming the developer is paying attention and not just accepting the code without review [11].

In some game environments human players can interact with machine agents in the form of AI assistants or opponents [3]. These agents have limited abilities and can only operate within the game environment, although within this fixed setting they can achieve impressive results [55].

In all these examples of human-agent interaction a non-expert human user can only interact with these machine agents in a predefined modality or environment. Contrast this with human-to-human interaction where we can use language (verbal or written) to exchange information, define tasks or goals, create new knowledge or insights, and much more. All of this is achieved in a fluid and extensible manner appropriate to the setting, and agreed or understood by the participating human users, with misunderstandings able to be resolved using the same techniques.

The simple human-agent interactions listed previously are however, just the first step towards more complex and rewarding interactions [24]. These can include increased collaboration, a deeper feeling of team-working and eventually support for extensible interactive problem-solving [93].

This thesis considers the potential for a future environment in which human cognitive activities can be better supported by machine agents in a dynamic and extensible manner, without the need for time-consuming technical integrations into existing tools or platforms, and for these machine agents to access broader sets of task-relevant information and knowledge, such as written, drawn, or otherwise contributed material from other users as well as access to relevant online sources.

To support such a capability the human users must be able to operate in an information environment that the machine agents have access to, and both human users and machine agents must be able to dynamically share new artefacts into that environment. They must be able to create modifications to existing information, create or delete links, meta-data or contextual information. The vision is that all agents can work together to collaboratively co-construct knowledge and task-relevant information to more quickly achieve their collective goals. Specifically, this co-constructed knowledge will need to support multiple modalities and a range of specificity, from rapidly foraged and fluid information to more formally defined knowledge. In this context human users should be able to invoke relevant cloud services, quickly establishing a level of trust appropriate to those services and contributing relevant results back into the shared knowledge graph. By fusing knowledge through co-construction, we can move beyond simple conversational interactions or bespoke applications common for machine agent integration today, enabling faster and richer collaboration mechanisms to support more challenging problem-solving settings such as collaborative sensemaking.

The research reported in this thesis was funded by the Distributed Analytics and Information Science (DAIS) International Technology Alliance (ITA) [110] which had a particular focus on *rapidly evolving* situations in a *coalition context* operating at the *edge of the network* with *limited resources and communications* [159]. These focus areas and target capabilities can be found throughout this thesis, for example in the form of coalition operations involving collaborating human users and machine agents, and the need to rapidly construct task-relevant solutions to support these coalitions securely but efficiently in environments with limited resources. In such settings the ability to predict the applications that may be needed and define them in advance are limited, hence the desire for a flexible and extensible knowledge sharing environment.

1.1 Motivation: Supporting sensemaking

The primary motivation for this thesis is recognition of the potential for richer and more powerful teams of human users and machine agents working together to achieve a common goal. Today the ability to interact with machine agents is typically limited in style, modality and environment, with the need to predefine the capabilities and design specific tools. By investigating relevant techniques to unlock the potential power of human and machine agents can we identify ways in which they could interact in a more unconstrained manner, especially in support of larger and more challenging goals?

Rather than considering conversation specifically as an interaction modality, or the ability to invoke functions via technical Application Programming Interface

(API)s or more user-friendly Graphical User Interface (GUI)s, instead the goal is a shared conceptual knowledge space where human users and machine agents can contribute and consume information, building on the contributions of others. This ability to carry out human-agent knowledge co-construction can support both the simple interaction mechanisms that we see commonly today (as described in the previous section), but also provide a solid basis for the implementation of more challenging operational contexts such as collaborative problem-solving and joint investigation, sensemaking and Situation Understanding (SU). These constitute examples of higher-level operational needs for which powerful HAT environments can be applied, but they are by no means the only such modes that can be supported. There are also clear opportunities for using the underlying insights to support other high-level goals such as knowledge modelling, interactive information assurance and provenance, and collaborative domain mapping.

There are many cases where SU and sensemaking activities are successfully undertaken today, but usually with a variety of tools and systems ranging from traditional pen, paper and whiteboard techniques, through electronic tools that are familiar to trained users, but typically not well integrated into any wider environment, and in some cases through custom built systems and bespoke solutions. The need for accurate sensemaking varies in value and fidelity depending on the operational context, and the cost of achieving and maintaining SU will vary also [68]. A key factor motivating the research throughout this thesis has been the need for machine agents to operate in a dynamic environment, meaning that agility and flexibility of both the human users, machine agents, and their operating environment is essential.

It is also the case that many problem-solving activities that are undertaken by human users could benefit from machine assistance, and an improved ability to communicate and interact in a more natural format could underpin such collaborations. This thesis explores specific techniques to allow human users to identify and record information relevant to their sensemaking goal, and how machine agents can also contribute task-relevant capabilities. Both sensemaking and SU are well researched areas and there are several methodologies and frameworks that have been defined and are used by professionals working in this space. The potential for machine assistance here is well known, with specific but usually narrow examples, and this thesis investigates the potential for broader capabilities to support improved HAT and how it can be applied to sensemaking.

Taking OSINT analysis as a specific example within the broader sensemaking field brings concerns around veracity, authenticity, speed of access, volume of information, and sharing of sources and findings with other analysts [5]. Building hypotheses, explanations and stories to explore the domain are all important capabilities that are especially needed in this setting. In the literature there are several publications that specifically call out the potential for machine-assisted sensemaking (e.g., [14]), but it is noticeable that there is currently no single unified approach proposed to enable this more generally. While tools do exist to support sensemaking, some tend to focus either on data and information visualisation while other tools can be used to capture insights and hypotheses from human users as they explore the information. Further tools provide capabilities for simple augmentation of information, or capture and sharing of visualisations, and some provide simple predefined analytics that can be performed.

Relevant factors for enhanced human-agent collaboration

This broad description of the problem space and potential solutions is useful scene-setting, but for meaningful advances to be made it is necessary to be more specific about factors that will be considered throughout this thesis. The general goal of supporting dynamic collaboration between human users and machine agents in a problem-solving setting comes together in the definition of Human-Agent Knowledge Fusion (HAKF) as defined in this thesis. Figure 1.1 shows a mind map with annotated links that provides more structure to the broad problem space described so far, showing the set of relevant factors for HAKF. In this figure the nodes are named (within each ellipse) with a short additional description underneath. The directed links are represented as arrows and named using a typical style that allows the graph to be read by moving from node to node via the links. For example: *Human users exploit machine agents* (where human users are usually domain experts not computer scientists, and machine agents are usually cloud-based services), or *Knowledge Fusion requires multi-modal information formats and supports problem-solving.*

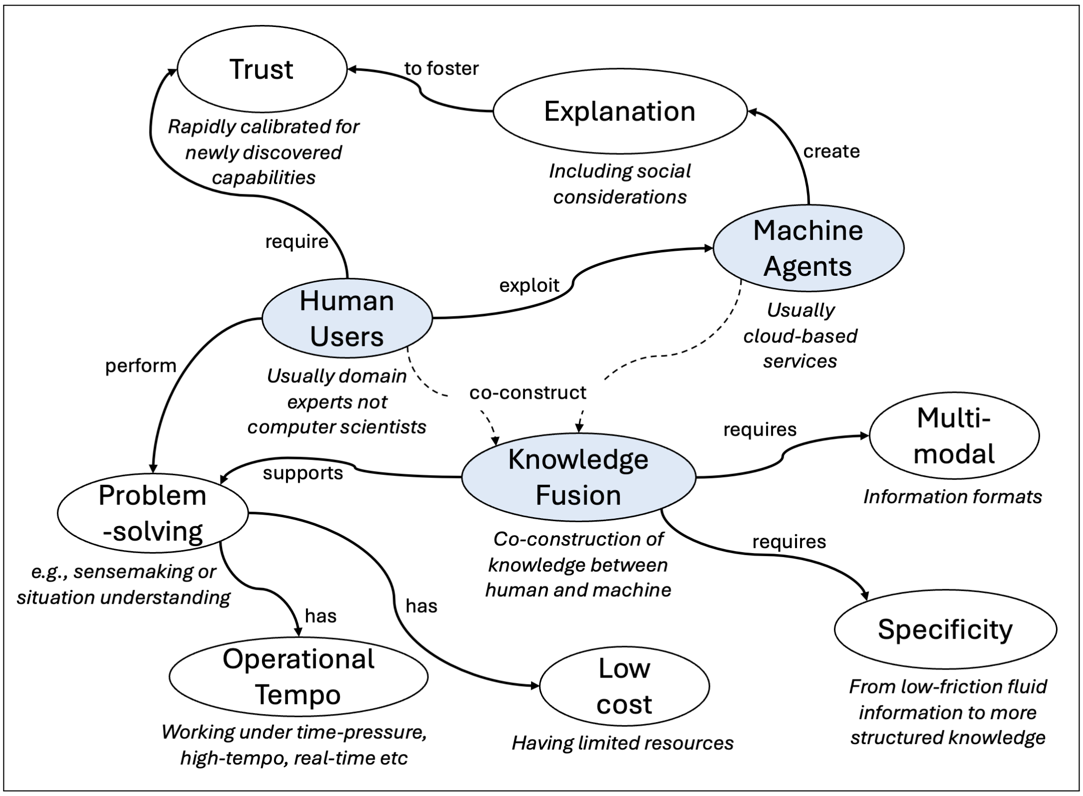


Figure 1.1: Relevant factors for enhanced human-agent collaboration

The topics shown on the mind map in Figure 1.1 are listed below with a brief commentary for each item, and these collectively form the overall scope of focus for this thesis and will be refined into more specific proposals in later chapters. Two flows are highlighted: between *human users*, *machine agents* and *knowledge* *fusion*, and these both take the form of co-construction activities and are defined more thoroughly in Chapter 3, but they represent two key forms of interaction that allow task-relevant knowledge to be contributed.

The relevant factors that define the enhanced human-agent collaboration scope of focus for HAKF, and therefore this thesis, are:

- **Human users** "Usually domain experts, not computer scientists" - these human users typically lead the task, attempting to progress or resolve a particular challenge.

- **Machine agents** "Usually cloud-based services" - these may be simple or complex capabilities provided by machine agents and can be independent/autonomous or triggered as required by the human users.

- **Knowledge fusion** "Co-construction of knowledge between human and machine" - a core capability that enables task-relevant information and knowledge to be collaboratively and iteratively created by both human users and machine agents.

- **Explanation** "Including social considerations" - details relating to something else that has been created within the co-constructed knowledge graph, often (but not exclusively) created by machine agents for artefacts they have previously contributed.

- **Trust** "Rapidly calibrated for newly discovered capabilities" - required for human users to collaborate and machine agents to be useful, fostered through outputs and behaviours. Trust values can be provided by others or built directly through experience.

- **Problem-solving** "e.g., sensemaking or SU" - a typical but challenging class of problems that might be undertaken.

- **Operational tempo** "Working under time pressure, high-tempo, real-time etc" - this indicates that the problem being solved has some urgency or time-critical aspect and is not a theoretical problem for which time constraints do not apply. Sometimes referred to as *mission speed* or similar by practitioners.

- **Low cost** "Having limited resources" - related to the operational tempo, this indicates that there are typically limited resources available, so problem-solving activities must take this into account.

- **Specificity** "From low-friction fluid information to more structured knowledge" - for challenging tasks such as sensemaking there are a variety of processes that are relevant, and they require different information specificity; generally, more fluid information is lower cost to create than more structured knowledge.

- **Multi-modal** "Information formats" - explicitly identifying the need to support multiple modalities for information (image, video, audio, tabular and more).

In combination these relevant factors constitute the key considerations for achieving a more advanced set of HAT capabilities and in combination they form the basis for HAKF.

1.2 Human-Agent Knowledge Fusion (HAKF)

Ahead of a thorough definition in Chapters 3 and 4 respectively, here is a brief outline of the HAKF concept and Cogni-sketch implementation to help the reader better understand how these two technical contributions come together to form the basis for the research reported in this thesis.

Building on the relevant factors and their relationships, as reported in Section 1.1, Figure 1.2 shows a high-level proposal for HAKF as a concept to support this kind of fluid task-relevant information and knowledge co-construction, enabling rich collaboration between agents. This responds to the need for the agile integration of human users and machine agents into dynamic and responsive teams that can work together within the conditions identified as relevant factors in Figure 1.1. HAKF is designed to support this deep interaction, comprising bi-directional knowledge and information flows through co-construction, and to support meaningful communication between human users and machine agents [26].

The term *knowledge* within the HAKF acronym is carefully chosen and is used to specifically capture the fact that *knowledge* could be shared between agents via this approach. In different situations this knowledge may instead be *information* but with additional context this will become *knowledge*, and vice-versa, and sometimes when there is no shared context at all it might all be just *data*. Often the operating context of the different agents will determine the difference between knowledge, information and data, rendering the distinction contextual rather than universal. For this reason, I have chosen Knowledge (K) as the relevant term in HAKF to express the upper end of this contextual range.

There are two main flows indicated for HAKF as shown in Figure 1.2:

- Tellability: a flow where new information is conveyed from one of the users, often to impart local task-relevant information or knowledge that could improve the performance of the system overall. Typically, this flow is from human users but both human users and machine agents can perform tellability.

- Explainability: a flow that provides a greater level of transparency into a conclusion or output from an agent. This is often invoked through *why?* questions and appropriate responses. Typically, this flow is from machine agents (and can encompass XAI techniques), but human users can also perform explainability.

HAKF systems with explainability aim to increase *trust*, or more specifically human-agent confidence (through transparency), and systems with tellability can

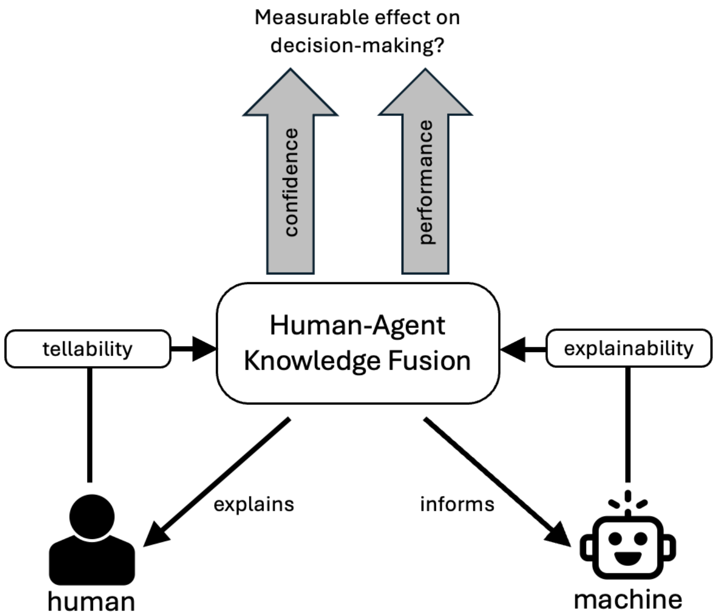


Figure 1.2: Human-Agent Knowledge Fusion (HAKF) - a high-level concept to support co-construction between human users and machine agents.

increase the *operational tempo* via machine agent performance (through rapid and explicit customisation or configuration of those agents within the same environment).

The ability to create a hybrid environment that can drive improved confidence alongside increased generic machine agent performance or agility could be a powerful tool for many knowledge-intensive applications such as sensemaking and SU. Whilst HAKF is a high-level concept, the twin flows of tellability and explainability capture the mechanisms to underpin the general co-construction of knowledge or information between human users and machine agents. The top part of Figure 1.2, relating to the potential for increased performance and confidence leading to a measurable effect on decision-making performance are discussed in Chapter 3 when the topic of HAKF is defined in more detail.

Cogni-sketch is a tool that represents an instantiation of the HAKF concept and is a middle ground between human users and machine agents, enabling the simple representation of information and knowledge, through machine inference or human intuition in a knowledge graph-based environment to support co-construction. Machine agents are available in several forms but can be located as *functions* which can be invoked by human users to fulfil different tasks. Cogni-sketch is designed to be flexible and extensible through palette extensions, new machine agent functions, visualisations, extensions and more. Ahead of a thorough introduction and description later in this thesis, an example of typical Cogni-sketch usage is shown in Figure 1.3. This example is included here to give the reader an early idea of the high-level form that this environment takes in a similar way to the brief description of HAKF. In this example the graph shows details of some aspects of my research and literature review since I found the Cogni-sketch platform a useful resource for activities like that during my research. This kind of usage also helped me to understand the limits of traditional systems and the potential for improvements on these.

In this example the assistance of machine agents was not directly required, but the ability to have a highly visual layout under my control, with an extensible palette and the ability to provide meta-data for nodes and links was very valuable. Other uses cases were developed to investigate the requirements for machine agent integration and are reported later in this thesis. As the PhD has progressed and the graph has grown in size and complexity the search function, as well as the ability to easily create different projects for different perspectives, coupled with GitHub integration for version control of Cogni-sketch knowledge graphs have all been both pragmatic and simple advances. For other use cases a wide variety of example machine agents for processing of different data modalities and purposes based on local libraries and cloud-based services have been created. Appendix A has links to videos that track the progress of these activities, and more, throughout the development period of the Cogni-sketch platform, and Chapter 4 provides a detailed description of all aspects relevant to Cogni-sketch.

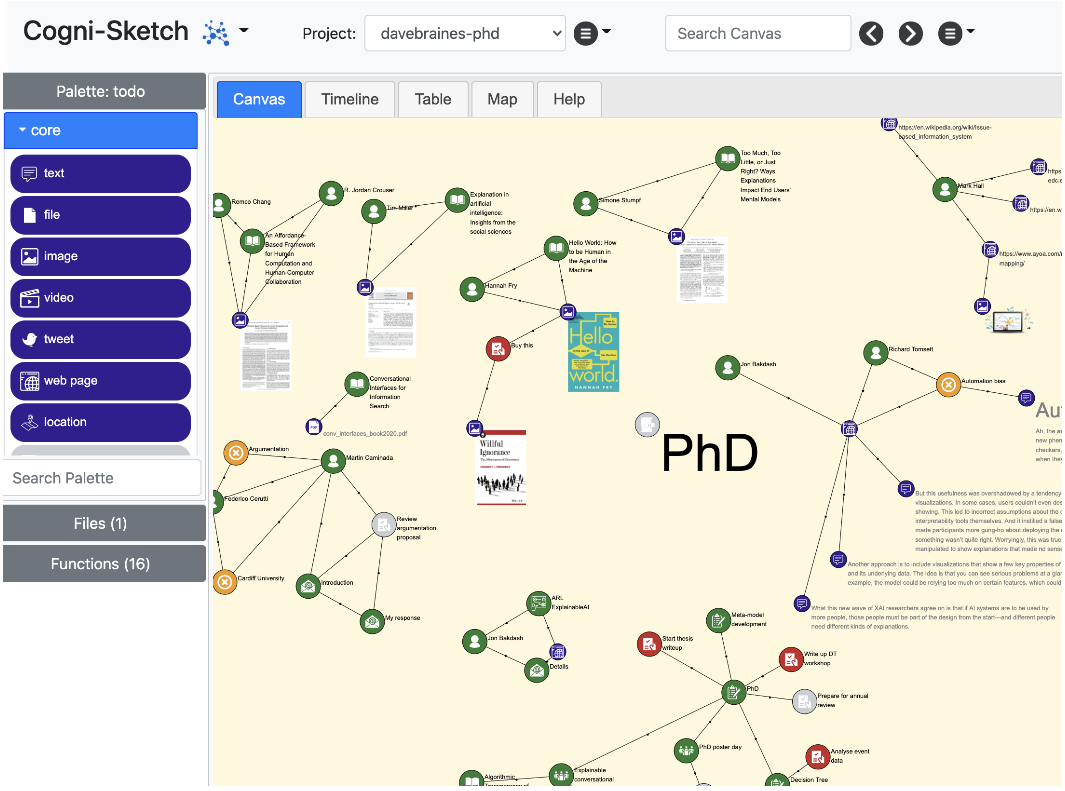


Figure 1.3: Tracking PhD progress using the Cogni-sketch environment.

1.3 Research goals

The research goals for this thesis are summarised below and expressed in the form of an overall research question, with three more focused sub-questions covering more specific aspects. These focus primarily on the ability for both human users and machine agents to operate in a common environment, working together to collaboratively build and refine shared task-relevant information and knowledge for an active operation with operational tempo and low-cost motivations. The overall question sets the scope of this thesis, with the three sub-questions each motivating a specific aspect of the research and outputs spanning a variety of topics including: human creativity, machine assistance, sensemaking and SU. The question of novelty and how feasible such an approach could be is also considered.

Research Question: *To what extent can **human** users and **machine** agents* *operate in an **open unified information environment**, able to **consume** and **create** contextually relevant **information** in a variety of **modalities**, in support of **problem-solving** goals?*

To achieve more specificity and drive particular areas of investigation for this thesis, this main research question can be split into three further sub-questions, each of which contributes to the overall research question, but is more focused and able to be pursued and answered by the research reported in this thesis. These are listed below:

RQ1: Can human users create multi-modal and semantically meaningful information into an environment that is easily processable by machine agents? (tellability).

RQ2: Can such an environment support relevant machine agent capabilities to assist human users in their goals and provide additional task-relevant information? (explainability).

RQ3: Can human users pursue shared understanding using sensemaking techniques in such an environment?

Throughout this thesis these three specific research questions are referenced and for ease of use their individual *RQx* identifier is used. *RQ1* and *RQ2* capture the human user and machine agent needs and will require capabilities, specifically to ensure that human creativity can be adequately supported, and machine assistance credibly provided without substantial additional effort or intervention. *RQ3* considers a specialised use case to contain the scope of focus and shape an experiment with human users.

In addition to these specific research questions, it is also important to ensure a focus on novelty and feasibility within this opportunity space where there are multiple existing solutions for parts of the problem space, but no attempt to create a unified environment to support all these related needs. Together these research questions aim to define the scope for a novel and feasible human and machine communication and collaboration environment. Within this environment specific information intensive problems such as sensemaking or intelligence analysis can be pursued, in search of repeatable and straightforward techniques to foster SU.

These research questions must also account for key factors for both human users and machine agents, but which may impact each of these agents differently. For example, human users can use visual layout and style to help their reasoning and help convey or gain understanding, especially of loosely related information, e.g., through visual clustering [51]. It is important that these kinds of features are not prevented in any solution as it would likely limit the human users in their effectiveness. Likewise for machine agents, the ability to easily identify the type and semantic meaning of different content types may enable logical inferencing to be carried out by a suitably qualified machine agent, and the chosen representation format should be able to support this level of detail in the information should the human users choose to use it.

Feasibility of the solution is an important counter-balancing perspective, so for example, any ability for the human users to take advantage of machine agent reasoning should not place a high-burden on those human users to produce detailed semantically defined knowledge conforming to a potentially complicated predefined ontology. They must be able to still quickly sketch outline information in some cases whilst perhaps providing more detailed knowledge in other cases. The ability to do this must not be constrained in time sequence; both human users and machine agents should be able to revisit any part of the shared knowledge space and make modifications or improvements whenever needed.

It is important to recognise that there will often be a tension between the desired precision of the machine agents and the need for informality or less structure from the human users [144], for example to move quickly, or operate in an initial exploratory mode where such precision may not yet be known and could cause friction or delay if it must be obtained before anything meaningful can be recorded. Such behaviour should be acceptable (and indeed desirable), with support for the human user(s) to augment any captured information with additional semantics later in the process as they become available, and if they are relevant to the task.

1.4 Thesis contributions

Using these three research questions to frame the problem space that is in focus for this research, the following main contributions can be found within this thesis:

RC1: **Definition of HAKF** as an under-pinning concept to support the agile human-agent collaborative environment that is outlined in this thesis (specifically to support *RQ1* and *RQ2*), comprising the tellability and explainability flows, and how these can support collaboration through increased confidence and context-aware knowledge and configuration.
The HAKF concept has been published in [21] and is described in Chapter 3.

RC2: **Creation of an operational instantiation** of the HAKF concept for human-agent collaboration, enabling experimentation and general usage. Specifically, this resulted in the development and open-source release of the Cogni-sketch environment along with specific plugins to implement various machine agent capabilities (*RQ2*), human problem-solving and visualisation capabilities (*RQ1*), and features specific to sensemaking and SU (*RQ3*).
Cogni-sketch is introduced and defined in [21] and is described in detail in Chapter 4.

RC3: **Evidence that inexperienced human users can successfully use HAKF** as embodied in Cogni-sketch for sensemaking and communicate their findings. This is achieved through the execution of a formal human user experiment into the use of the Cogni-sketch environment for a **simulated sensemaking exercise** based on analysis of social media sources

(*RQ3*). This includes the answering of two predefined intelligence questions, making use of the Pirolli and Card [113] sensemaking loops including foraging, sensemaking and storytelling to communicate the conclusions of each participant (*RQ1*).

The results from this formal experiment with 12 participants is reported in Chapter 6 (Section 6.4) and is reported in [25].

RC4: **A methodology for integration of machine agents** into a HAKF-based environment, specifically through co-construction of machine generated information into the knowledge graph. This takes the form of various explanations and other contributions from machine agents operating within the Cogni-sketch environment (*RQ2*) in a variety of SU demonstrations (*RQ3*), with development and integration of these machine agents carried out by the author of the tool and this thesis as well as other collaborators, with the latter helping to validate the simplicity and extensibility of the architecture for the insertion of machine agents.

The methodology and subsequent details of these implementations are reported throughout Chapter 5 and have been published in [27] and [21] along with earlier published material that outlines the original conversational approach [29].

As is the case for the research questions, these contributions are referenced throughout this thesis, and for ease of use their individual *RCx* identifier is used.

In addition to the four listed main thesis contributions, the following additional contributions are also noted as novel and valuable. They are reflected in various publications produced as part of this research as indicated below:

- **A wider set of methods for human-agent collaboration** (*RQ1*, *RQ2*) in support of SU (*RQ3*) as reported in [21].
- **Insights from a Design Thinking (DT) workshop** with Subject-Matter Expert (SME)s, focused on XAI and the role of both AI systems and ex-

planations (*RQ2*) in a military context, as reported in Section 3.2. These insights have been refined into various categories and some helped to inform both the conceptualisation of HAKF and aspects of the design for Cogni-sketch.

- Use of the Cogni-sketch experimental environment in a **real OSINT analysis pilot exercise** (*RQ3*), used by an expert intelligence analyst using real open source data over a 3-month period during which improvements and extensions were identified alongside the successful use of the Cogni-sketch platform to support the investigation, as reported in Section 6.3.

1.5 Thesis structure

Chapter 2 provides background material in the form of a literature review for research relevant to HAKF and the application to sensemaking, considering factors important to both human users and machine agents and for communication in general.

Chapter 3 introduces the concept of HAKF to support HATs collectively solving problems, identifying specific aspects that must be supported in any implementation. A DT workshop with military stakeholders is also described, and how this helped to inform the required capabilities for HAKF.

Chapter 4 outlines the experimental Cogni-sketch platform as an instantiation of HAKF. This chapter starts with a brief assessment of existing relevant tools and techniques before defining the scope of Cogni-sketch, how it supports the required capabilities of HAKF and the various extension points for customisation and integration of machine agents.

Chapter 5 is focused on machine agents and their ability to make task-relevant contributions based on their processing or analysis. Broadly these contributions align to examples of the explainability flow and are expressed through a pilot and evaluation as well as some conversational explanations to demonstrate a variety of behaviours.

In **Chapter 6** the focus moves to the ability for human users to successfully perform sensemaking through the tellability flow, and the creation of relevant material in a form that is visually and cognitively useful to human users. A pilot exercise with an intelligence analyst informed the design and execution of a subsequent formal experiment to measure sensemaking behaviours and outcomes for 12 human participants. Results from both the pilot and experiment are analysed and reported.

Finally, **Chapter 7** summarises the contributions, presents a brief timeline of the HAKF research activity, summarises some additional example use cases, and proposes future potential extensions and further areas of research. It also highlights recent advances in Large Language Model (LLM) technology that are highly relevant but not covered in this thesis since they occurred after completion of the research reported here.

There are also three appendices:

Appendix A contains additional detail for the Cogni-sketch platform and links to further resources such as code, documentation, and video demonstrations.

Appendix B contains the full set of data obtained from the human sensemaking experiment described in Chapter 6.4 along with a qualitative assessment of the artefacts created by each of the 12 participants during the experiment.

Appendix C contains some additional useful information relating to the DT workshop reported in Chapter 3, Section 3.2.

Chapter 2

Background

2.1 Introduction

Taking the specific research goals and general motivation to pursue HAKF as described in the previous chapter there are several areas in the literature that are relevant to these aims. The field of Human-Agent Teaming (HAT) has a variety of highly relevant publications that provide insights and approaches relevant to the overall goal of providing a *knowledge fusion* environment in which *human users* and *machine agents* can collaborate, with particular focus on information exchange and co-construction that would be at the heart of any such solution. The ability to support *multi-modal* knowledge and information and do so in a *low-cost* setting that does not require significant effort from either human users or machine agents is also important, and the ability for knowledge graphs to support these capabilities is discussed.

Also relevant are the topics of *collaboration* more generally, and *trust*, especially in cases of rapid trust calibration for sets of agents that are brought together and expected to reach *operational tempo* quickly. From a machine agent perspective, the role of *explanation* is important and the research field of XAI explores the different capabilities that are offered, the forms that they take and the degree to which their value can be influenced by human explanation styles as defined in the social sciences.

Finally, the application of the above to *problem-solving* use cases such as sensemaking and Situation Understanding (SU) are highly relevant with approaches and techniques defined in the literature. For human users to be effective the need for a common operating environment was previously proposed, as well as the ability to support different *specificity* of data/information/knowledge, ranging from more fluid low-cost data or information to more structured knowledge. In addition to these aspects there are a small number of publications relating specifically to the role of HATs for sensemaking. Some of these provide a summary of requirements or attributes desirable for future solutions, and these are summarised accordingly and revisited in later chapters.

The method employed to undertake the literature review reported in this chapter was primarily two-fold. The first part of the process occurred during the early phase of the research that informed the HAKF concept and corresponding required capabilities. This arose naturally from a series of investigations and subsequent publications into different aspects of human-agent collaborative systems, incorporating various aspects of well-known research areas such as XAI and HAT as well as relevant works from the social sciences. These individual reviews were thorough but limited in scope to the publications in question. As the concept of HAKF emerged more formally the second phase of the literature review was undertaken; a more thorough and extensive review of the literature, spanning the sections and sub-sections reported in this chapter. This was originally undertaken prior to the more detailed definition of the HAKF concept, early in the research phase of the project. This identified a broader set of relevant literature around a number of key topics. This broader scope informed the HAKF concept and was revisited again during the final writing phase of the thesis, to ensure that full coverage of the key areas was achieved, and to identify and new and relevant publications that had been released during the research period. The basic search method used academic publication search engines such as Google Scholar, Research Gate, and the Cardiff University LibrarySearch as well as following citations in identified relevant papers. This second phase of the literature review started with the topics already identified in the first phase, but scanning more broadly for new interconnecting works, further developments in experimentation and/or experimental results, and to ensure that the latest developments in rapidly developing fields such as XAI were captured. The results are reported in the remainder of this chapter.

2.2 Human-agent teaming

Human-Agent Teaming (HAT) is a term used to define the interaction between human users and machine agents [78], where the machine agents can be virtual and disembodied, or more physical such as robots [92]. Machine agents can be implemented using a wide variety of technologies and solutions, but of particular interest in this field are AI and ML agents [128]. A motivating hypothesis is that human and machine teams can together deliver improved performance [40, 41] but it is also important to observe that care is needed to avoid unintended side-effects or biases from either type of agent having a detrimental effect [42].

Within the scope of this thesis the focus is on the virtual and disembodied agents, with physically embodied robotic systems out of scope. This exclusion is driven mainly by the focus on integration at the level of knowledge and communication rather than physical capabilities and the navigation of a spatial environment, but all the capabilities outlined in this thesis could also be applied, with extensions, to supporting a physical robotic setting if needed. With this restriction in mind this literature review does not focus at all on robotic systems or any related publications or considerations. The machine agents that are within scope are envisaged to either operate autonomously with their human team-mates or are invoked in specific contexts by those human users.

An important consideration for any co-construction of knowledge is the ability for the behaviour of the machine agents to be modified or configured by the human users either directly, or as a side-effect of their provision of task-relevant information into the environment. This desired autonomous capability draws in aspects from the HAT and multi-agent systems literature.

2.2.1 From information exchange to co-construction

It is common in the literature to consider communication between human users and machine agents, but typically this is envisaged as a conversation or some similar form of message passing between the agents [60]. However, a different and potentially more powerful and extensible co-construction approach may be feasible, and by using this the more traditional forms of agent-to-agent communication can be supported as a by-product. The goal is for agents (both human and machine) to exchange information, and a co-construction approach can be used, as presented in the research from Kopp et al [79] investigating the importance of joint co-construction and understanding mental states. This envisages the act of communication between agents being the incremental joint co-construction of a shared knowledge space or mental model [10]. This is a re-imagining of more traditional inter-agent communication mechanisms and can support a broader and less controlled set of information exchange between agents by those agents simply contributing to the shared and co-constructed space.

This co-construction approach can be applied to any number of interacting agents with each reading and writing to the shared conceptual space. It is reminiscent of the earlier *blackboard architecture* [62] for collaborating AI agents and provides a powerful and scalable basis to support other collaboration forms, for example by enabling more specific UIs, such as conversation, by navigating the relevant parts of the collaboratively constructed graph to build the content of the conversation.

It can also support a range of psychological behaviours observed in humans that are also directly applicable to machine agents, (e.g., following a joint goal, supporting others in achieving that goal, trusting that other agents will mutually to adhere to these, and using language to give evidence of understanding [57]). This perspective arose not only from the from the position that “Building computer systems that are able to converse autonomously and coherently with a human is a long-standing goal of AI and Human-Computer Interaction (HCI)" ([79]), but also the realisation that "we must start to re-consider the hallmarks of cooperative communication and the core capabilities that we have developed" ([79]).

With this co-construction approach the authors recognise that there are large discrepancies between human-agent communication and natural human-human communication, noting that differences in effectiveness and flexibility are particularly notable, along with the remarkable capability for resilient, robust and efficient communication between humans. Building on Grice [57], they start with cooperation as a fundamental prerequisite for human communication along with more recent reinforcement of this through the human psychological infrastructure of shared intentionality [148]. Specifically, the authors claim that the "two crucial mechanisms that allow humans to achieve mutual understanding in a dialogue are the primacy of *joint co-construction* as the stepwise construction of a joint activity, and the primacy of *mentalizing* as the ability to perceive, understand, and predict an interlocutor's relevant mental states" ([79]) (emphasis added).

With this approach it is also possible to explicitly recognise, and to some degree mitigate, the issue with absolute meaning in knowledge representation, enabling each agent to consume the shared information representation according to their own conceptual model, which might be a simplified version of a bigger model, and may be specifically aligned to the capabilities of that agent. This aligns strongly with the social constructivism perspective [154] from the pragmatics of human communication, recognising that meaning may not be universal and is constructed by the receiver based on their subjective perception of the context and can enable different agents with partial models of the world to operate effectively. For co-construction, the authors conclude that mechanisms are needed to enable agents to "cooperatively and incrementally co-construct a successful interaction with a human user" ([79]).

The form in which information is co-constructed or exchanged is also important, and this includes considerations of specificity in addition to support for different modalities as other considerations. For humans this should be in a manner that does not prevent or suppress their creativity or other inherent human capabilities, and for machine agents this must be in a format that does not require substantial or costly processing, and ideally contains minimal ambiguity. Humans often revert to informal methods when attempting to solve problems, for example just thinking in their head, using pen and paper for sketching, or maybe using simple electronic tools such as MS-Word or MS-PowerPoint. If support systems require too much precision, complexity or overhead then it can be a disincentive to human users, even if as a result of using them they can be subsequently supported more easily by machine agents. For example, in [134] the authors compare digital sketching of designs with pen and paper sketches and find that the digital tooling can get in the way and reduce available human capacity for thinking about the problem when compared to pen and paper.

To support a rich and flexible co-construction basis for information exchange there are a number of possible implementation styles, but a typical approach, given these requirements, is the use of knowledge graphs [48] to capture the relationships between information. The node-and-link structure of knowledge graphs also provides an ideal opportunity for the incremental aspects of co-construction, with different agents able to contribute new knowledge on top of existing knowledge provided by other agents, either as new nodes, links or data properties. Knowledge graphs can be augmented through schemas or ontologies to convey semantic information and thereby support reasoning [34] by machine agents with access to those schemas or ontologies and the ability to perform inferences based on the defined semantics. Whilst this provides a powerful additional benefit for machine agents it is important that human users are not prevented from easily accessing or understanding the knowledge graph data, for example by enforcing a need to first understand a complex ontology. There are other possible information representation formats, and the physical implementation is a separate decision to the conceptual approach used. Given the close alignment of knowledge graph capabilities to the general knowledge fusion desire for HAKF, and the ability for both human users and machine agents to access and contribute to the knowledge, the capabilities of knowledge graphs are a strong match and highly relevant.

When considering scalability issues with knowledge graphs, the work from Pienta et al [112] is useful. They propose techniques for handling large graphs, usually by navigating to an appropriate sub-graph that is suitable for processing, through graph sampling, filtering, partitioning or clustering. This requirement is especially relevant for the human team members, whereas the machine agents are typically able to process larger graphs without overload, although limits still apply, albeit far higher than for the human users. It is interesting to note that all these sub-graph creation techniques can be performed by machine agents and could either be coded specifically into a system and triggered on demand by the human users, or they could be more general capabilities provided by an autonomous machine agent able to apply them on behalf of the human users, without a specific request, but instead based on their perceived contextual need.

2.2.2 Collaboration

Building on a core co-construction approach [79] to support inter-agent communication, in this section the closely related topic of *collaboration* is reviewed, with specific consideration for the similarities and differences between human users and machine agents that wish to collaborate.

A useful approach to conceptualising typical human-agent collaboration is the terminology of affordances [53] (see [52] for the application of affordances to technology in general), and specifically an affordance-based framework for human computation and human-computer collaboration as proposed in [39]. An affordance is one or more qualities or properties for an object that defines its possible uses or makes clear how it can or should be used. In the context of human users and machine agents these affordances are defined (in [39]) as typical characteristics or capabilities for human beings when compared to computer processes, and vice-versa. Additionally, it is noted that "there exist affordances in both directions. Both human and machine bring to the partnership opportunities for action, and each must be able to perceive and access these opportunities for them to be effectively leveraged. These affordances define the interaction possibilities of the team and determine the degree to which each party's skills can be utilized during collaborative problem-solving" ([39]).

Typically, humans exhibit affordances related to visual perception, visuospatial thinking [136], socio-cultural awareness, creativity and domain knowledge. Machines on the other hand are given affordances that relate to large-scale data manipulation, collecting and storing large amounts of data, efficient data movement, and bias-free data analysis (with the authors noting that the final category only includes new biases from the machine agent directly; any biases introduced by the humans in the data, training or implementation will already be present). This concept of affordances originated in the world of physical design and usability [101, 100, 107], and was used as terminology for how a physical object advertises or presents its use. It is however very useful for characterising different processes, situations or services to determine how much they benefit from human verses machine affordances.

As machine agent capabilities continue to advance some of the traditional human affordances can be approximated or mimicked by machine agents so they become applicable to both agent types, albeit still dominant for humans for now at least[1].

When considering the need for human and machine agents to collaborate, especially in a fundamentally open and co-constructive environment, as outlined in

[1]The work reported in this thesis predates the advent of Large Language Model (LLM)s but the capabilities of machine agents using these could represent a clear example of this sharing or blurring of traditional human affordances.

the previous section, the concept of affordances can provide a useful abstraction for identifying which contributions are best performed by which agents. More specifically, it becomes clear what the human users and machine agents can each contribute, and therefore which tasks or goals they can best contribute to. This is loosely related to the roles of the human users [149] and can be seen as an important secondary characteristic to be considered alongside the more fundamental role they are fulfilling within the system. The consideration of affordances can therefore help to identify what is within a particular agent's scope, and how they should express it. These affordances in the context of HATs also enable a two-way flow of information that is key to collaboration and aligned with the blackboard architecture pattern discussed earlier. The human users must be able to perceive the affordances of the machine agents to make use of them, and the machine agents should recognise and respect the capabilities of the human users to avoid operating in situations where the humans instead should be responsible. The latter is usually achieved through programming, training or configuration of the machine agent, but with a co-construction approach it is possible that such information could be drawn dynamically from the knowledge graph.

The set of affordances for a particular team of human users and machine agents define the interaction possibilities for that team and can determine whether and how the various skills can be applied during challenging tasks such as collaborative sensemaking or problem-solving. By expressing capabilities in terms of affordances the machine agents are able to contribute a wide range of technical capabilities that can be filtered in the context of particular problems and presented as a short list of options to the human users or triggered autonomously, thereby mitigating potential overload for the human users when trying to harness input from machine agents.

In addition to affordances there are other perspectives relevant to considerations of collaborating between human users and machine agents. One of these is the different thinking styles for humans, and especially the difference between thinking *fast* (system 1, intuitive and reactionary), and *slow* (system 2, methodical and deliberate). Originating from Kahneman [70] this work has been more recently applied specifically to AI agents [19]. In the context of HATs, it is often clear exactly what capabilities the machine agents can bring, but those can be narrow and/or brittle. Humans tend to be better at generalizability, robustness, explainability, causal analysis, abstraction, common-sense reasoning, ethics reasoning, as well as integration of learning and reasoning supported by both implicit and explicit knowledge. These capabilities come with corresponding human limitations such as lower speed and accuracy, as well as the inadvertent potential introduction of biases.

As well as providing a useful perspective relevant to collaboration, and being additive to the earlier discussion on affordances, Booch et al propose a set of research questions within their work [19]. The latter three of these are directly relevant to the joint and incremental co-construction basis for HATs, and the final question applies to any subsequent implementation. The questions are:

- "How do we define abstraction / generalization mechanisms that are guided by a notion of attention and pass from the raw data level to a more abstract level? How do we know what to forget from the input data during the abstraction step? Should we keep knowledge at various levels of abstraction, or just raw data and fully explicit high-level knowledge? What does it mean for knowledge to be explicit: is it related to the presence of metadata, structured knowledge graphs, or language-related entities?" ([19])
- "In a multi-agent view of several AI systems communicating and learning from each other, how to exploit/adapt current results on epistemic reasoning and planning to build/learn models of the world and of others?" ([19])
- "What architectural choices best support the above vision of the future of AI?" ([19])

For the final question the authors propose that it is likely that architectural support for their vision of AI is aligned to a multi-agent system with individual agents having specific skills and focusing on particular problems, and that these agents can act asynchronously, and independently contribute to building models of the world (as well as models of other AI systems, and models of self), and that these agents can be combined in many ways. This again aligns strongly with the join incremental co-construction mechanism [79] and can simplistically be viewed as a rich but powerful example of a blackboard architecture. The considerations of system 1 vs system 2 thinking are directly relevant to some of the problem-solving domains such as sensemaking and are revisited later in this chapter.

Finally, in the context of collaboration, Bradshaw et al discuss in detail the idea of 'Making Agents Acceptable to People' (2004) [20] with a specific focus on both the social and technological factors that can help to improve collaboration in such systems. They lay out a number of principles and policies in support of this broad goal, calling out factors such as the ability for agents to adjust their autonomy based on the context of the task and the needs for the team, and the ability for notification and collaboration policies to support agent interactions between members. Whilst the use of dynamic policies such as these is not a major focus of the research reported here, it is important to distinguish between autonomous and directed agents and the ways in which they might interact with their human teammates, with the potential for a flow between these two states in more advanced systems such as those needed for successful sensemaking.

2.2.3 Trust

Finally, trust is a broad issue even just considering the scope of HAT, and there has been substantial work in this area (e.g., [88]). It not a central topic for this thesis, so is only covered in a particular context, and that is the specific goal of supporting *rapid trust calibration* [150]. This is the process of establishing and adjusting trust between human and machine agents quickly, effectively and accurately. An assessment, or trusted declaration, of the reliability and competence of the machine agent is required and this may be processed in conjunction with the human user's perception and confidence in that particular machine agent's capabilities, with the declared trust level being potentially modified by each human user or other machine agent accordingly. This is modelled on human techniques for building trust via reputation or based on performance. These clear and transparent metrics for the machine agent capabilities and performance can serve as the basis for initial human user trust levels. For the co-construction multi-agent architecture to be viable there must be mechanisms available for agents to determine and declare their trust in each other, and the ability to do so rapidly and dynamically in the context of the task or problem is important.

We return to the context of HATs later in Section 2.4.4 where we consider cases of HAT when applied specifically to sensemaking. This material is intentionally located at the end of that section so both the component topics of HAT and sensemaking have been thoroughly introduced before considering their combination.

The proposal for a co-construction basis for knowledge and information sharing between agents aligns well with a multi-agent human and machine collaboration environment, and a knowledge graph with an appropriate level of corresponding semantic information may be a useful basis for implementation. Affordances represent a mechanism for human users and machine agents to advertise their capabilities, and for decisions to be made about their applicability to different tasks. The need to support a wide variety of cognitive styles (e.g., system 1 and system 2 for example) is also an important consideration and motivates the need for a flexible and extensible approach.

These and other factors will be important considerations for typical HAT aspects. An extensible knowledge fusion core that can support human users and machine agents with multi-modal information in a range of specificities at an operational tempo, and with the ability to rapidly form trust between agents, seems credible.

2.3 Explanation

The field of Explainable Artificial Intelligence (XAI) has seen a large amount of activity in recent years, with different techniques described [46], some studies and approaches for evaluations of their effectiveness for human users [97], and several code libraries and frameworks being released to assist with building XAI solutions. However, in addition to these core XAI techniques there is important research which reminds us that the need to provide explanations, and the mechanisms used to convey them, has been well studied in the social sciences over many years [95]. Earlier in this chapter, when considering human-agent interactions and how they can be influenced by human-human communication styles, mention was made of relevant work in [57] and others.

For explanations specifically there is an extensive review of the relevant social science literature from Miller [95] that provides substantial examples and motivation to the XAI community. He identifies key perspectives and approaches from this substantial and long-running body of relevant social science research that has investigated how humans explain to each other, and the different motivations and styles for this. Specifically, he identifies that there are valuable and largely untapped bodies of research in philosophy, psychology, and cognitive science relating to how people define, generate, select, evaluate, and present explanations, and he draws conclusions from a detailed review of over 250 publications on that topic. Miller places XAI firmly at the intersection between three key research areas: (1) AI, (2), social science, and (3) HCI, and proposes that XAI represents just one specific form of human-agent interaction, but many similar interactions could also benefit from considering these social science factors.

The main conclusions from this detailed survey of the social sciences all relate to the manner and style of explanations that should be served, rather than the content they contain, or the techniques used to obtain them, with the latter being the majority focus of the XAI field.

The specific recommendations from this work are that:

- Explanations are **contrastive**; they are sought in response to counterfactual cases, which are named foils.
- Explanations are **selected**; in a biased manner to serve the context from the environment or the interaction.
- **Probabilities** (probably) **don't matter**; while truth and likelihood are important in explanation and probabilities really do matter, referring to probabilities or statistical relationships in explanation is not as effective as referring to causes.
- Explanations are **social**; they are a transfer of knowledge, presented as part of a conversation or interaction, and are thus presented relative to the explainer's beliefs.

Miller also notes that the above findings are based on an analysis of *everyday explanations* which relates to why specific events occurred or decisions were made, rather than explanations of more general concepts such as scientific explanations. He also reports that biases and social explanations occur in human explanations, and these should not be discounted when explanations are generated by machines because their interpretation by human users is important. It is also important to note that an effective explanation is more than just causal attribution; they are contextual and the explainer and explainee must work together to establish what context the explanation is occurring in, and therefore what form is best, including a potential negotiation about this between the agents.

There are several relevant factors for explanations and interpretations, spanning ethical considerations about models and their data sources, as well as adversarial and manipulative concerns for ML models generally. These include the informativeness and transparency of XAI solutions and whether they are inherently transparent or require post-hoc explainability methods [84]. These cover considerations of the models themselves and their internal structures, as well as the style and content of explanations that can be provided, enabling consideration of how they may be interpreted by human users. Related work [103] investigates a range of techniques spanning feature attribution, visualisation (e.g., saliency maps), the use of simplified or surrogate models and influence functions. It also explicitly considers opportunities for interfaces and interactions between these techniques and the benefits that can arise. These publications and many more provide underlying mechanisms (and factors motivating these) that a particular explanation may use to communicate relevant information to the user, and in any implementation these actual explanations may be delivered via an interaction that can use guidance from Miller [95] to determine the form or style in which the explanation is served.

Returning to the need for post-hoc explanations: In cases where the internal processing of a model is unable to provide transparent explanations directly, then post-hoc explanation techniques can be used [66]. These are techniques that involve using results from the model or service in some way, usually through multiple invocations with different combinations of parameters, to attempt to determine which features or other aspects of the input most influence the result. For example, in the case of image classification, this may be the generation of a saliency map that highlights the parts of the image that most strongly influenced the classification in the form of an overlay heatmap onto the image. Widely used examples of surrogate models as a post-hoc method for providing explanations are Local Interpretable Model-Agnostic Explanations (LIME) [124] and SHapley Additive exPlanations (SHAP) [87] which can provide local and global explanations for individual classifications or more general model features respectively. These are commonly used techniques that might feature (alongside or instead of many additional similar capabilities) as the details of an explanation provided by an XAI system.

However, whilst the desire to provide good XAI solutions is high, and there are many techniques proposed in the literature as briefly described above [126, 124, 84, 103], recent experiments have found that providing XAI explanations alone does not necessarily improve human decision-making [4]. In this study the presence and accuracy of the AI agent does improve human decision-making measurably, but the addition of an explanation for that same AI agent has no additional effect. Other factors such as the value of the explanation and the manner in which it was delivered are not reported, so the ability to interactively explore an explanation (e.g., in the style suggested in [95]) may change this outcome, as may experimenting with other factors relating to the content or style/modality of the explanations given.

Finally, it is important to note that whilst we have talked about explanations so far as a general concept, any actual explanation will need to account for the purpose of the explanation, or the context in which it will be processed. One factor relevant to the identification of that context is the role of the user to whom the explanation is being provided, and specifically how the interpretation of that explanation may occur. In related work, we have asked the question: *Interpretable to whom?* [149] and explored typical roles for human users in the context of a ML system, and specifically one that can provide explanations. The ability to provide an explanation, and thereby create an interpretation by human users, is not universal, and the techniques used to convey relevant information should take this role into account. In this work we propose that the needs for an explanation are driven at least partly by the user role, and that different levels of detail will be required between creator, operator and executor user roles (and others). The understanding of user role needs combined with the social considerations from Miller [95] suggests that explanations must be agile and may contain a variety of information and detail, and that they may be built up incrementally rather than having all information available in advance.

2.3.1 Terminology for aspects of explanation

There is a variety of terminology in use with the field of XAI and there are efforts to standardise this for common understanding and communication. For example, Arrieta et al [6] differentiate these key terms:

- Understandability/intelligibility: whether a model can reveal its function, without sharing internal structure or algorithmic details.
- Comprehensibility: the ability for an algorithm to represent learned knowledge in a human understandable form.
- Interpretability: the ability to provide meaning in understandable terms to a human.
- Explainability: associated with the notion of explanation as an interface between agents. Measures whether an explanation is possible, and if so to what degree.
- Transparency: a model is considered to be transparent if by itself it is understandable.

They declare that understandability (rather than explainability) is the key general concept when considering any of these terms at a higher-level of detail, but for alignment with the broader literature this thesis uses XAI and specifically explainability when discussing the ability for agents to explain themselves to others. The terms defined in the list above are used in this thesis where relevant and aligned to these definitions.

The broader literature is surveyed for claims about what needs are sought to be fulfilled by using XAI [6]. These stated needs are varied but largely built around the sharing of additional information to achieve some secondary potentially valuable effect related to communication, trust or learning.

These needs are summarised as:

- Trustworthiness: whilst this is a valid goal for a human-agent system, the provision of an explanation may not be enough to achieve trustworthiness, and indeed trustworthiness may also be achieved in other ways.

- Causality: the agent providing the explanation may not be aware of, or able to compute causality between variables.
- Transferability: by providing explanations, the agent can reveal additional information that could help to assess the transferability to adjacent problems.
- Informativeness: this is a standard and generally applicable reason for providing explanations, but the additional information that can be provided may not align with the human user requirements.
- Confidence: specifically, the algorithmic confidence in a classification (or other) output from a model.
- Fairness: provision of details, for example about the training data, to enable assessment of whether the model can be considered 'fair' in any given context.
- Accessibility: to enable consumers of the model to more easily understand what is happening with the model, especially for complex algorithms.
- Interactivity: to allow users a level of interaction with the model that can reveal relevant information as requested rather than all in advance.
- Privacy awareness: specifically, to enable assessment of whether any privacy issues arise because of the model algorithm.

These various terms are useful to review (noting that there is some inconsistency within publications) and serve to define the scope of relevant factors for XAI-related research and capabilities, and to provide precise vocabulary to be used throughout the thesis for this topic.

2.3.2 Principles for design and interaction

Separately, in [36] the focus is on human-XAI interaction, and suggested design principles for explanation UIs. Interestingly there is no proposal made that it may be possible to support all of these within a single architecture or implementation. Specifically, they identify seven different human-XAI interaction styles based on a systematically obtained set of XAI publications that mention UIs/interaction:

- Information transmission: to present users with accurate or complete explanation about AI behaviour (transparency).
- Dialogue: to facilitate natural and iterative conversation about AI behaviour (transparency, scrutability).
- Control: to support rapid convergence towards desired AI behaviour (effectiveness).
- Experience: to manage expectations about AI behaviour (satisfaction, trust, persuasiveness).
- Optimal behaviour: to adjust human behaviour despite limitations of fully understanding the AI behaviour (efficiency).
- Tool use: to facilitate learning from AI behaviour about a given domain (effectiveness).
- Embodied action: to establish a joint understanding with the AI for an effective collaboration within a domain (effectiveness).

They also propose four design principles for interaction with XAI UIs:

- Complementary naturalness: consider complementing implicit explanations with rationales in natural language.
- Responsiveness through progressive disclosure: consider offering hierarchical or iterative functionalities that allow follow-ups on initial explanations.

- Flexibility through multiple ways to explain: consider offering multiple explanation methods and modalities to enable explainees to triangulate insights.
- Sensitivity to the mind and context: consider offering functionalities to adjust explanations to explainees' mental models and contexts.

Whereas in [111] authors from National Institute of Standards and Technology (NIST) suggest that there are four different principles related to the explanation itself: the *meaningfulness* of it, the *accuracy* of it, a recognition of the *knowledge limits* of the AI system and recognition that it should only *operate within these limits*.

Whilst it is natural to focus on a well-designed implementation for digesting explanations it is also important to account for issues that may arise with human interactions, however well they are designed. Kaur et al observe that "Often, ML-based systems and interpretability tools are designed with seamless interaction and effortless usability in mind. However, this can engage people's automatic reasoning mode, leading them to use ML outputs without adequate deliberation" ([71]). This forms a key part of their analysis of techniques to support ML explanations to human users and their proposal to view explanations as a form of sensemaking.

These explanation-related capabilities are driven by different motivations and collectively represent a comprehensive set of perspectives. The ability to contextually model these and bring relevant factors into a socio-technical system is clearly valuable as evidenced by the many publications focused on the definition of XAI and the corresponding demonstrations and examples of usage. Typical XAI solutions today are specific to the problem space rather than being more general capabilities that can be applied in multiple settings.

The need for explanations within a HAT built around knowledge fusion resonates well with these more general observations, rather than the specific capabilities provided by individual XAI techniques. The ability to explain is assumed to be possible, but the way the explanation is provided is important, with the ability to provide that explanation into a co-constructed knowledge environment being a useful broad mechanism to support incremental contributions, contributions from different agents, and some of the social interaction styles suggested by Miller [95]. XAI literature focuses mainly on the need for machine agents to provide explanations, and specifically for the results arising from AI or ML processes that may be inherently unexplainable, often referred to as *black boxes*. The ability to provide contrastive explanations, to potentially generate multiple possibilities to then choose from, to include optional probability information, and to serve these in an interactive style are all important considerations and can be supported through a flexible core knowledge representation component.

2.4 Sensemaking and intelligence analysis

Intelligence analysis [64] is a specific form of sensemaking and requires the analysis of a wide variety of different data sources, often in pursuit of decision advantage [138] (being ahead of your adversary) and can underpin Situation Awareness (SA), Situation Understanding (SU) and other higher-level concepts as described below. This is a continuous process that involves (at least) the collection, processing, exploitation and dissemination of task-relevant information [90] that can inform decision-making. Intelligence analysis is often performed by teams of analysts but may also be a solitary activity. It continues to evolve as both data sources, techniques and software capabilities are created that can support human analysts [7] in their task. Usually, the results of intelligence analysis are reported to a decision-maker who is typically a different person to the analyst, but depending on the size and complexity of the task the decision-maker may also be the analyst.

Intelligence analysis[2] involves the consumption of existing data, and recogni-

[2] An easily readable and useful overview of intelligence analysis and the potential for AI

tion of salient (task-relevant) information from a wider set of potentially irrelevant data. It focuses on building a model of what is already known or is currently happening (insight [96]). The typical goal of intelligence analysis and sensemaking more generally is to inform some higher-level capability to predict future actions or outcomes so that plans or mitigations can be developed (foresight [96]). Generally speaking, sensemaking progresses from a fluid and relatively unconstrained process through to a more structured and rigorous process as time passes, however it is important to recognise that this is not a linear activity. The ability to move between fluidity and rigour at any stage, especially for human users, is very important. Often the human analysts are working under time pressure and with limited resources and information; this time pressure can lead to issues with cognitive biases, errors or miscommunications, but the need to harness potentially valuable human intuition and insight from the process must not be hindered. If machine agents are involved, the human users and machine agents can share information, with one operating on the outputs of the other, and potentially with machine agents creating new information that may then further inform the human analysts or modify their activities [44]. Importantly, "sensemaking is not about truth and getting it right. Instead, it is about continued redrafting of an emerging story so that it becomes more comprehensive, incorporates more of the observed data, and is more resilient in the face of criticism" ([156]). This fluid and emerging nature of sensemaking is an important consideration both for human and machine agents and the processes that might be relevant, and for the manner in which sensemaking is considered.

There are many models and processes proposed for sensemaking, but they are typically iterative and constructive, and there is often no obvious start or end point for the process: "sensemaking begins and ends based on triggering events, available data, analyst knowledge or previous experiences, or external factors such

assistance is available in the Dstl biscuit book titled "Human-centred ways of working with AI in intelligence analysis" ([45]).

as deadlines" ([44]). As the process moves from fluidity to rigour (and back) newly identified information is used to confirm or disprove/modify existing hypotheses, information or knowledge.

2.4.1 Principles and needs for intelligence analysis

There are a small number of recent publications that review the state-of-the-art for machine-assisted intelligence analysis, and in some cases define principles for further investigation, or unaddressed requirements.

Nine principles for information interaction and collaborative interpretation

Attfield et al identify nine principles [8] relating to "information interaction and collaborative interpretation", specifically observing that "these principles have implications for the design and evaluation of training, culture, processes and technology relevant to sensemaking".

These principles are derived from a thorough literature review considering sources from: organisational studies, computer supported collaborative work, Naturalistic decision making (NDM) and HCI. The overall scope for this work is the recognition that "The Future Operating Environment (FOE) is likely to be one in which operational combat units are required to be increasingly mobile and geographically dispersed with more decentralised Command and Control (C2) structures" ([8]). The authors note that "the current C2 structure does not support agility", and "the current C2 hierarchy is linked to the military culture". The latter point is very important since any attempt to improve agility in this space must recognise the overall hierarchical setting for the human users and be able to provide agility and fluidity within this, but also be capable of broader operation to support future evolution and diverse settings.

Overall, the paper outlines how the need for devolution to support increased agility drives the need for local sensemaking.

The principles are listed below, with a summary for each:

1. **Provide sufficient cues for sufficient sensemaking**
   This principle relates mainly to uncertainty and disambiguation; seeking patterns and clues in the data to seed ideas that might lead to more data collection or analysis. This could drive the selection of specific *cue patterns* to try to formalise the understanding of the data and is often achieved by people seeking expected cues within the environment.

2. **Support low-cost information workflows**
   This principle is about the ability to provide information processing workflows quickly and easily.

3. **Represent information quality and provenance**
   The ability to represent information provenance and quality is key to understanding the overall situation that is being modelled.

4. **Promote expertise and domain knowledge**
   The idea that experts have a library of template frames to support them in their sensemaking, and that these could be credibly operationalised into suitable tools for different situations. The idea that mental reasoning tasks can be operationalised into perceptual tasks or physical procedures.

5. **Allow time to acquire data/information to build an evidence-based and coordinated situation picture**
   The need to create time and space away from the high-tempo (system 1) operation to allow deeper (system 2) thinking [19].

6. **Use strategies for negotiation of sense**
   This principle is focused on ensuring that input and insight can come from many users in different ranks and roles and talks about incentivisation systems for producing content or identifying errors. This comes across as a

form of gamification; incentivising the creation of new content or insights based on some notional value or reward.

7. **Where appropriate, use strategies for frame enumeration and elimination**
   This helps avoid common cognitive pitfalls such as confirmation bias, it is recommended for analysts to assess multiple frames against their hypotheses, and eliminate those that do not fit. Humans struggle to do this at scale and often only focus on any one possible frame at any point.

8. **Provide explanatory context for actions, orders and requests**
   The observation that requests for action (e.g., orders) are better received if they contain contextual information to provide a sense, or rationale, for the request.

9. **Minimise the costs of achieving and maintaining common ground**
   The observation here is that achieving and maintaining common ground between agents is expensive, so techniques to minimize the cost are encouraged. Specifically: "Aim to reduce these costs through the use of standardised terminology, protocols and procedures" ([8])

These nine requirements resonate strongly with the focus topics for this thesis as shown earlier in Figure 1.1. *Low-cost* information workflows correspond entirely to that topic, whilst information quality and provenance are a form of *explanation* and can easily be represented as part of the *knowledge fusion* core. Related to provenance, the consideration of rationale is also important, especially in terms of seeking to obtain some degree of shared understanding [98]. There is also recognition of the *operational tempo* and the need for different levels of *specificity*. Due to the strong match between these nine principles and the focus of this thesis, this list is revisited later with a specific summary of how each of these nine are directly supported (in Section 6.2), as well as a set of suggested small extensions that would provide even better coverage (in Section 6.5.2).

Related work investigating a framework for systems thinking practice [38] proposes an approach for operational research practitioners to engage in complex situations. This applies at a much broader level than the previous set of nine principles from Attfield et al but provides a good example of the need for systems and approaches to be compatible with a much higher level of conceptualisation. This reflective practice approach is based on an intersection between the practitioner, their method and the philosophy under which the complex problem is being progressed. This is an interesting formalisation for capturing the need for methodology being aligned with the beliefs, values and biases of the practitioner based on their worldview and accounting for the experiences, skills and preferences. Any tooling to support flexible activities in this context must be able to adapt to different philosophies and practitioners and perhaps also be able to switch between a set of different methods.

Requirements for HATs in intelligence analysis

Also relevant is the paper on "Human-Machine Teaming in Intelligence Analysis: Requirements for developing trust in machine learning systems" ([76]) from Knack et al, based on interviews and focus groups with experts to understand technical and policy considerations. This work is mentioned where relevant throughout this section and the following two specific recommendations for future research are noted, both of which are partly addressed by this thesis: "(1) to identify technical and policy considerations for more advanced use of ML in HAT (such as non-classification use cases), and (2) to develop methodologies for understanding the analyst workflow to guide ML application development, and embed behavioural and decision science into software engineering practices" ([76]).

Analysis of expertise in intelligence analysis

Finally, Hepenstal et al [63] carried out a set of structured interviews with six intelligence analysts using Cognitive Task Analysis (CTA) to identify the kinds of expertise found in intelligence analysis, with the goal being to support the design of Human-centred AI solutions. Examples of the likely benefits from the inclusion of AI agents included "the potential to aid analysts when making decisions, for example, by speeding up their analysis, improving accuracy, or focusing their attention upon the most important information" ([63]). This work explicitly recognised that intelligence analysts typically work in challenging and uncertain environments, usually under tight time constraints and are accountable for their recommendations. They need primarily to deliver the outputs required by their decision-maker but must also consider the situation and the analytic requirement and whether they are appropriately matched. The typical intelligence analyst setting is described to be: "an intellectually demanding problem space with no clear or obvious answers. Their experience and expertise (technical, subject, procedural, and disciplinary) are key and they use it instinctively" ([73]).

Given the accountability requirements, if human analysts can take advantage of machine agent assistance, they must have appropriate and accurate explanations from the machine agents, otherwise they cannot explain the evidence that underpins a claim and articulate why they are recommending a particular decision. This transparency was discussed in Section 2.3 and can enable the human user to inspect the path to the outputs. Specifically, Hepenstal et al state "If a system simply provides a result to an analyst, without transparency of the reasoning involved, then the analyst cannot use their expertise effectively, for example, to form an understanding of potential patterns of interest, lines of inquiry that have been explored, and lines that could be augmented, or the nature of arguments used. Nor can they learn from the system or develop expertise that would inform future analysis tasks" ([63]). This latter point about the analysts needing to learn better skills through application of their expertise is key, and any dependence on machine agents that are not appropriately understood may degrade the capabilities and growth of the human analyst as an unwelcome side-effect.

The structure used to drive the interviews was to consider 6 steps for each of the interviewees: (1) the driver, what signal triggered the analysis, (2) recognition of the type of analysis, (3) intuition, (4) following lines of inquiry, (5) insights, (6) claims. All of these are informative to the goal of this thesis, especially 1, 3 and 4 which can most readily be supported by machine agents. The results of the interviews identified a number of possible opportunities for these analysts, most of which fall into the category of directed machine agent services invoked by the human users, and align well to some of the capabilities described later in this thesis (e.g., Named Entity Recognition (NER), filtering of data, similarity identification, monitoring and alerting, automatic proposal for lines of inquiry, system challenges to hypotheses, automatic generation of reports). Whilst the suggested set of capabilities aligns well with expectations from other sources, there was no explicit discussion of the need for human users to configure these services based on the context of their usage.

2.4.2 Related situational concepts

Beyond intelligence analysis and sensemaking is the pursuit of Situation Awareness (SA) and Situation Understanding (SU) and the concepts of Shared Situation Understanding (SSU) and Coalition Situation Understanding (CSU). This thesis is mainly focused on sensemaking and its application to intelligence analysis with considerations for the role of machine agents in the process. However, these related concepts within the broader field of problem-solving are all relevant and are briefly expanded below:

- **Situation Awareness (SA)**: is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future. SA is a more immediate and focused concept than SU and is often associated with real-time decision-making in high-stress environments [47, 123].

- **Situation Understanding (SU)**: this builds on SA and is the process of comprehending and interpreting situations to support subsequent decision making [141]. To achieve SU requires the gathering and processing of contextually relevant information to build an accurate understanding of the world, typically via SA. SU is usually obtained via a more comprehensive and in-depth analysis of the situation over a longer period. It involves not only perceiving and interpreting the current state of the situation but also projecting that state into the future [140].

- **Shared Situation Understanding (SSU)**: is a state in which multiple agents seek a common SU. It involves at least communication of information relevant to the SU obtained by one or more agents and may also require some alignment of the mental models of the participating agents [139], especially in terms of their separate interpretation of the meaning of the situation and its contents [140].

- **Coalition Situation Understanding (CSU)**: is a specific variant of SSU noting that the team members trying to achieve SSU are from different organisations (i.e., they are members of a coalition) and therefore may have particular security requirements or concerns, and these may vary in intensity depending on the specific relationship and/or operation at that point in time [114].

- **Data fusion**: is defined by the JDL/DFIG (Joint Directors of Laboratories / Data Fusion and Information Group) and defines six levels of data fusion, at increasing levels of processing, ranging from low-level to high-level [15, 17, 18, 142] and can be applied to all the situational concepts described above (also, see OODA [137]). For a specific analysis of how this relates to the material reported in this thesis refer to [16].

Also, relevant is the concept of OSINT [54] which is a form of intelligence analysis that specifically draws information from publicly available open sources. It is defined as the "systematic collection, processing, analysis and production, classification and dissemination of information derived from sources openly available to and legally accessible by the public in response to particular government requirements serving national security" ([131]). This publication also advises that "information, even plenty of it, without processing, analysis and production, classification and dissemination is not authoritative intelligence, now less than ever" which serves as a useful reminder that provenance information, explanations and the resulting trust between human users and machine agents are essential considerations as well as the raw data itself. OSINT sources can present specific opportunities (large volume and high velocity) but also require careful attention to check for common pitfalls and issues especially as mis- and dis-information proliferate (validity, veracity and more) [131]. The scientific distinction between the opportunities and pitfalls is mainly focused on issues of trust [76] and reliability as well as the ability to process data quickly.

2.4.3 Models of sensemaking

There are several sensemaking models and small number were considered during this review, based on a combination of prevalence and adoption. Another important consideration for inclusion was adaptability, to ensure that HATs can be supported, rather than considering more structured analytic techniques that might only support the use of predefined machine services.

The following subsections describe the two main methods considered.

Klein's data-frame theory of sensemaking

The data-frame theory of sensemaking [73] proposes that frames can be used to explain data by fitting that data into a frame of reference, often using particular slots. This is a key mechanism for moving from data to information (and eventually to knowledge) and is usually context specific, sometimes with potentially multiple different valid interpretations of the data in the context of different frames.

Klein views data as “aspects of the world which a sensemaker experiences” and the frame is “a representation in the mind of the sensemaker which accounts for the situation and allows the data to fit” ([73]). Typically, a frame will involve the inclusion of lots of different data, and the existence of frames and the ability to choose which one to use comes from the expert behaviour of the analyst based on their experiences gained previously in doing this task. Frames can highlight causal relationships and can be the beginnings of a narrative structure to begin to tell the story of that data. Like all sensemaking processes it is cyclical, with new data being fitted into frames, or causing frames to be rejected and data to be moved elsewhere. Frames can also guide the search for new data, most commonly when slots are identified but have not yet been filled, and in doing so the story becomes stronger as the frame is more completely filled.

The frame acts as a mechanism for interpreting data but can also be used as an explanation when needed and typically seeks to account for multiple data within a wider and more integrated/complete picture. This inherent value of the frame is also extended beyond the data it contains when gaps are identified and used to task further collection, or support inference of possible values etc.

Klein describes the fluid nature of frames and their multiple uses as follows:

> As we encounter a situation a few key elements, or anchors, invoke a plausible frame as an interpretation of the situation. Active exploration guided by the frame then elaborates it or challenges it by revealing inconsistent data. By extending further than the data, a frame offers an economy on the data required for understanding, but also sets up expectations. Hence a frame can direct information search and in doing so reveal further data that changes the frame. An activated frame acts as an information filter, not only determining what information is subsequently sought, but also affecting what aspects of a situation will subsequently be noticed. ([73])

This data-frame theory of sensemaking is a widely used method across multiple communities sometimes intuitively rather than formally, for example doctors framing data about patient conditions, pilots understanding location and heading, or a ship's captain seeking understanding of an approaching aircraft (using position, heading and affiliation to infer intent). Each of these examples suggests an expert and therefore a repertoire of frames from which they can select and this is a key distinction between experts and novices (who may have no frames, simple frames, or incorrect frames).

It is proposed that experts and novices can reason using the same procedures, but experts have a richer set of frames to use and therefore have a substantial advantage. Experts can also use background knowledge and expectations or assumptions to fill gaps in frames but may sometimes forget the distinction between specific data and assumptions. By using data-frames the expert is typically framing and re-framing dynamically considering new data that is emerging. Klein also observes that "frame activation depends upon the sensemaker's stance, including factors such as workload and motivation, and their current goals" ([73]). Frame-activation is often done instinctively within the analyst's head, but there are opportunities for tooling to support this process if it does not add too much overhead and cognitive friction.

Similar techniques can also be used to provide more advanced capabilities to novice users and enabling them to develop their knowledge and expertise by using the techniques and associated frames. For example, techniques for enhancing the development of 'understanding capability' within a military setting have been investigated in [151], with potential for enhanced effectiveness if computational support could also be provided in a manner that does not hinder human creativity in such processes.

The Pirolli and Card model of sensemaking

The Pirolli and Card model of sensemaking [113] is a rich and well-defined set of processes and interconnected loops that explicitly attempts to capture the non-linear nature of the sensemaking process and uses intelligence analysis as a specific example to motivate the flows and component processes. The authors observe that intelligence analysis is a form of expert behaviour and therefore (like Klein [73]) they observe that it is expected that those experts will have a set of reusable patterns or schemas around which they have built their approach.

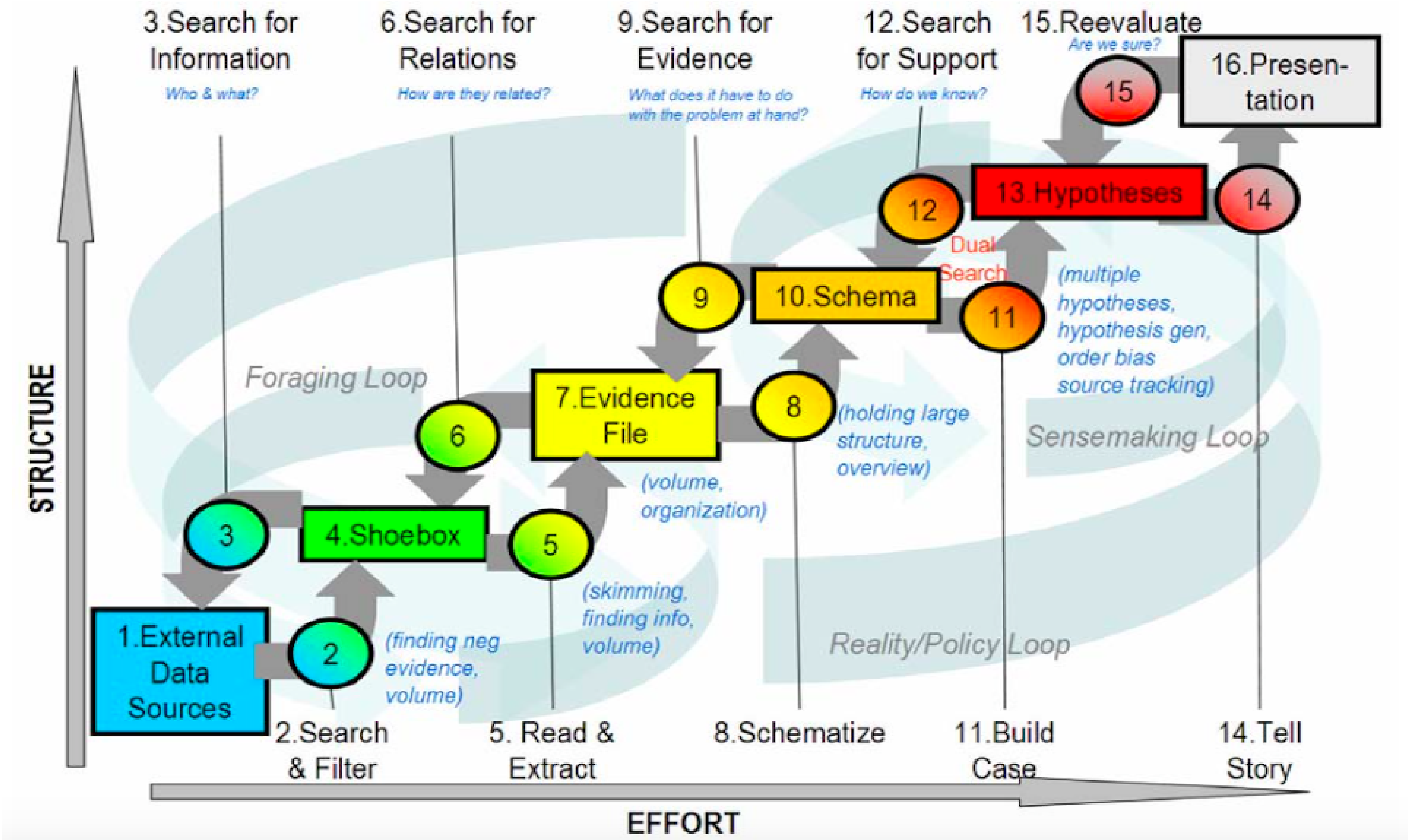


Figure 2.1: The Pirolli and Card sensemaking process for intelligence analysis (recreated based on [113]).

This is a widely used methodology within the field and has the specific benefit of fluidity across processes. It does not attempt to define a rigid analytic process specifically but represents a high-level progression of analysis from data sources through foraging to sensemaking and then storytelling as shown in Figure 2.1.

Within the diagram in Figure 2.1 the rectangles represent the approximate data flow from bottom-left to top-right. The circles associated with these rectangles represent the process flow, with the descriptions of each located at the top or bottom of the figure and indicating whether those processes are typically top-down (from theory to data) or bottom-up (from data to theory). The processes and data are arranged by degree of effort and degree of information structure which tend to increase in tandem from low-structure, low-effort foraging, through sensemaking, to high-structure high-effort storytelling. In addition to the data and processes there are also a set of interconnected loops: first, a foraging loop that seeks information through searching and filtering, and then extracting that information into a low-level schema, and then a sensemaking loop that involves iterative development of a mental model from a schema that best fits the evidence, and finally a reality/policy loop that enables the entire process to stay aligned with the outside world as processes, policies and technology evolves.

For the core foraging and sensemaking loops, and the broad goal of the whole diagram, the authors observe that the attention of the analysts is on both evidence and hypotheses and the interplay between them. For the lower-cost and lower-structure foraging region the authors propose that multiple types of activity are typically undertaken and these span exploration, enrichment and exploitation (usually into higher-level structures).

For schematization it is proposed that this is often done initially in the mind of the analyst (or perhaps with informal tools like pen and paper) and the authors explicitly call out the lack of easily used tools for this creative task. Whilst tools to support schematization do now exist (See Section 4.2.4 for an assessment of relevant tools) they are typically brittle and narrow, and often require extreme schematization in the form of an ontology and therefore can be too cumbersome for analysts to easily use. These tools are also generally not integrated into the other parts of the sensemaking process. The authors do however observe the opportunity for computer assistance at this point, through visualisation or analytics, or both, and observe that the analysts may be schematizing multiple small stories at this stage, to answer specific questions in pursuit of a larger overall goal.

In general, the data flow is from information to schema to insight and product, with the effort and structure increasing, and typically the volume of material decreasing as it flows from data to information to knowledge. It is important for analysts to be able to traverse this data and process space non-linearly and recognise that new findings from any level may influence existing findings or the meaning of information already collected. It is also critical that the generation of alternative hypotheses is undertaken, and the seeking of disconfirming data, as well as other defensive techniques to mitigate against common human biases.

This model was derived based on an empirical descriptive study using CTA [133] and a verbal task description protocol, with a set of interviews with intelligence analysts to derive common patterns across that community. This approach was used because of the observation that experts will typically build reusable patterns and some of these were already known and shared by the community. This CTA exercise was an attempt to unearth the full set of schemas and formalise the findings into a broad map of the patterns and processes. They found evidence of schemas for organisation of information to specifically aid planning, evaluation and reasoning, very much in line with the similar expectations of Data Frame Theory [73] and many of these schemas focused on common central concepts of people, organisations, tasks, and time.

This paper [113] also serves as a call to the community for points of opportunity within the component processes where tooling or other forms of assistance could be provided. They state that "the notional model presented (in Figure 2.1) provides an organization for identifying new technologies for improving the production of novel intelligence from massive data" ([113]). Specifically, they observe that many of these opportunities could be "aimed at expanding the working memory capacity of analysts by offloading information patterns onto external memory" ([113]), with a view to easing the burden on the human analysts, especially in the context of ever-increasing data [13]. Many of these opportunities remain unaddressed, and this work was a key motivation for the sensemaking use case focus for this thesis, based on our[3] earlier research in this area (e.g., [116, 118]).

General sensemaking processes

Also relevant are more general analytic styles or principles, such as:

- **Observe, Orient, Decide, Act (OODA)** - Usually described as a circular process, the purpose of this method is to recognise that there is a continual process of learning as new information is gathered, and this can be a feedback loop. Typically, the main consideration of the loop is speed and therefore responsiveness, and how the tempo can be increased. It is also commonly noted that the process is often not a simple circular flow [105], and therefore caution is needed with operationalisation [125].

- **Direct, Collect, Process, Disseminate (DCPD)** - Similar to OODA this cycle is typically considered a circular flow and is focused more on intelligence analysis and SU with specific steps for collection of information and dissemination of resulting knowledge [109].

- **Naturalistic decision making (NDM)** - A model for understanding how people make decisions in non-trivial real-world settings, including "dynamic

[3]In general in this thesis I use the terms we/our to describe the reported research as it formed part of a collaborative endeavour. However the content reported in this thesis is the sole output and contribution of the author, with any specific cases of directly collaborative work called out explicitly where they occur, with a citation or footnote indicating the collaborators.

and continually changing conditions, real-time reactions to these changes, ill-defined tasks, time pressure, significant personal consequences for mistakes, and experienced decision makers" ([72])

- **Recognition primed-decision (RPD)** - Earlier work from Klein et al that recognised that experienced decision-makers working under uncertainty and stress can instinctively recognize a plausible Course of Action (COA) as the first one they should consider [74]. This method is based on satisficing [4] [122] rather than optimizing and can account for a large variety of contextual factors rather than being a strictly defined process, explicitly acknowledging the value of intuition based on experience through situation recognition and mental simulation [74].

These are only briefly summarised in the short list above as the focus in this section is on sensemaking and the specific need to support a fluid process that allows the participants to move between different stages at different times, in order to contribute relevant information and knowledge at any stage. Typical structural analytic techniques require analytical (system 2) thinking [19] and are often better supported by more prescriptive methods and tools designed to counter biases that analysts will otherwise bring to the process, regardless of how well trained they are. Therefore, the principles arising from these should be considered in the design of systems to support analysis and problem-solving in order to achieve higher-level goals such as sensemaking.

2.4.4 Sensemaking for HAT

There is a small body of work, some located within visual analytics, that has been investigating the role for HATs specifically for sensemaking, for example Wenskovitch et al [157] argue that the role of both human user and machine agents need to be re-calibrated as team-mates rather than just humans who invoke

[4] A satisfactory or adequate result, rather than the optimal solution.

decision support services. This is similar to the observation from Knack et al [76] that ML agents should be designed from the outset to be integrated into the intelligence analyst's toolset and workflow.

Wenskovitch et al [157] focus on both transparency of the machine agents (a form of explanation) and the ability to improve human reasoning efficiency. Although they focus specifically on visual analytic techniques and interfaces there is a lot of commonality with the general role of sensemaking that is considered here. They explicitly state the need for bidirectional communication in human-agent systems and that such two-way flows are vital to providing a system in which the machine agents can operate more like team-mates. They note that "Bidirectional communication is important for productive human-agent interaction that can benefit all team members and, consequentially, improve overall team performance. The human benefits from a machine that clearly communicates its capabilities, limitations, and actions both explicitly (e.g., transparent display design) and implicitly (e.g., consistent predictable behaviour)" ([157]), and "The machine benefits from the human's corrective feedback by learning and improving its performance. The human's increased understanding of machine capabilities and improved machine performance from corrective feedback leads to appropriate trust and reliance on the machine. Consequently, there is an overall improvement in human-agent system performance" ([157])[5]. However, they also believe that the pursuit of machine agents operating as team members will require the human users (regardless of their role) to forfeit some control of their current capabilities to the machine agent. It is not clear why this would be required, and that aspect of their general position is therefore not considered further here.

The authors of [157] also identify several types of role that can be performed by agents within the system, and that these can eventually be performed by human users or machine agents when suitable technology and techniques exist

[5]Both of these statements are close to the definition of HAKF (introduced in Section 1.2), although the published HAKF definition predates these descriptions by 2 years.

to support this rich interaction. The roles are explorer, investigator, teacher and judge.

This is interesting and relevant work and similarly spans multiple areas of the literature but has been located within the literature review in this section as it strongly correlates with sensemaking. Their vision for a HAT is similarly grounded in the literature reported in Section 2.2 and they seek not only to propose techniques to achieve sensemaking but also that a mechanism for intelligible communication between human users and machine agents is needed, especially for them to work as team-mates rather that simple services.

In similar work Dorton et al [44] suggest two research questions that should underpin any approaches to collaborative human-AI sensemaking for intelligence analysis, along with the observation that algorithms, (ML) features and outcomes matter when considering the machine agents within the system and the ability for human users to select relevant machine agents from a set, and for a particular operational context. Their two proposed research questions are:

- "How can analyst sensemaking be used to affect AI/ML performance? Since analysts benefit from effective AI/ML and suffer from ineffective AI/ML, it is important to investigate the mechanisms by which a human analyst can influence the effectiveness of an AI/ML tool"
- "Can we develop a notional framework for Human-AI sensemaking, where human analysts augment AI/ML performance, and AI/ML tools are used to augment analyst sensemaking processes?"

Like Wenskovitch et al [157], through these research questions they are describing a two-way flow of information between human users and machine agents and similarly noting that the outputs from the machine agents could positively affect the human analysts and modify their behaviour, and that the human analysts can also influence the machine agents to improve their outputs relevant to the context of the problem. Knack et al specifically observe that the way "an analyst treats an output from a ML model is highly context-specific, and the lack of technical explainability of many ML systems is widely acknowledged" ([76]), and part of the co-construction approach and bi-directional flow proposed for HAKF is to mitigate at least the first part of this issue.

Dorton et al [44] focused on two different models of sensemaking, one of which is data-frame theory [73] and note that all models generally favour an iterative and constructive approach. They contrast sensemaking to analytical thinking and they observe that valuable human capabilities such as the ability to generate hunches and pursue information in a way that uses their biases can be useful for sensemaking whereas generally these same capabilities are not useful for analytical thinking. This relates to the system 1 / system 2 perspective reported by Booch et al [19] where they suggest that a good environment for human use must allow the speed and unpredictability of system 1 but also support the rigour and precision of system 2.

This work (Dorton et al, [44]) is another example of recent research that is closely aligned to the intersection of related research interests that is the focus of this thesis. The authors propose a need for a human-agent sensemaking framework and suggest that it must attempt to overcome brittleness (by allowing machine agents to become team-members, and human users to configure the machine agent behaviours) and trust (a common issue with current systems that use ML processes that may not be appropriate to the current situation or data). They refer to the different valuable capabilities of human users vs machine agents (referred to in this thesis as affordances, as described in Section 2.2), noting that machine agents can at least process large volumes of data, whilst human users bring better handling of context, the ability to overcome complexity and the ability to harness their system 1 thinking in the form of intuition and insight. They also propose that in a fluid bi-directional communication environment such as this the analyst is likely to develop a more nuanced understanding of the machine agent(s) and be more likely to explain the outputs, largely because the machine agent was configured directly by the human user's actions and any relevant collected data and information. The authors conclude by noting that "There is still much work to be performed to develop more fluid and meaningful interactions between humans and AI/ML in intelligence analysis" ([44]).

Whilst mainly focused on SA, John et al define four principles for maintaining and recovering SA [68] which are directly applicable to considerations of HAT settings for sensemaking. These include: (1) the ability to naturally alert human users to any changes in the information they have already assessed or accepted, (2) the need for unobtrusive notifications, (3) the need for summary descriptions of each change, and, related to unobtrusive notifications, (4) the need for careful thought about the user environment and the avoidance of visual overload. These arose from the observation of expert users in complex and noisy environments but are directly relevant to the goals of this thesis and drive some of the requirements for collaboration.

2.5 Gap analysis

The analysis of relevant background research in this chapter has identified several gaps in the existing body of literature, especially when considering the combination of HAT techniques to embody machine agents using XAI for challenging problems such as sensemaking for intelligence analysis. These are identified using *[Gap n]* in the text below, and are mapped to the RQs and RCs in Table 2.1 at the end of this section, along with a reference to where in the thesis the supporting material can be found.

Much of the existing material relating to the processes of sensemaking, and specifically the role that machine agents may perform in that space, is based on interviews with SMEs and the use of techniques such as CTA. No experimental results with fine-grained behaviour analysis were identified for such sensemaking activities, and there is a substantial opportunity to provide useful experimental result data here *[Gap 1]*. Such experimental results could be beneficial for investigating specific focused questions and could enable a comparative baseline to be established for use in future experiments. The research reported in Chapter 6 culminates in such an experiment and provides an analysis of this, based on the Pirolli and Card sensemaking model [113] applied to an OSINT sensemaking scenario with a set of novice human users. There are also a small number of publications that present specific requirements on the research community for different aspects of sensemaking, especially when progressed in the context of HATs. These have been reported as part of this literature review and are revisited in later chapters to report on relevant progress in this thesis against those requirements.

From a HAT perspective there is a substantial gap in consideration of a conceptual basis for supporting advanced human-agent interactions. Typically, the focus is on information exchange between agents, but some recent work has proposed a co-construction approach for information and knowledge co-creation that can be applied more broadly to support typical HAT activities. The application of this co-construction-based approach to problem-solving domains has not been covered *[Gap 2a]*, and the potential for a co-construction approach in this space could underpin some of the social science perspectives required by the XAI community.

Also relevant is earlier work on affordances, and how they can enable human users and machine agents to implicitly advertise their task-relevant capabilities in typical HAT settings *[Gap 2b]*. These affordances provide a strong and well-defined conceptual basis for general capabilities but their application to directly enable collaboration between human users and machine agents in environments beyond visual analytics (such as sensemaking) is not reported, nor the extensions required to achieve this.

Taking this co-construction-based approach to information and knowledge exchange, and harnessing affordances to enable key flows for human users and machine agents Chapter 3 proposes HAKF as a high-level conceptual architecture to support task-relevant interactions and sharing of information and knowledge.

The concept of XAI is extensively covered in the literature and has been a focused area of research in recent years. As can be seen from the literature review, there is substantial focus on specific techniques and their applicability and value in different situations, but there is little to explore the ways in which explanations can be surfaced to human users, accounting for their needs. In particular Miller [95] observes that there is a substantial and long-running body of relevant research from the social sciences that has investigated how humans explain to each other, and the different motivations and styles for this. Whilst this work can be seen as a call to the XAI community to account for the consumption of explanations, as well as their composition there remains a gap in the literature in responding to this challenge *[Gap 3]*. Chapter 5 provides details relating to the style of contribution of explanations, specifically the ways in which these are delivered by machine agents into the environment, and how this underpinning approach, built around co-construction, can support many higher-level interaction styles such as conversation, exploration etc. It is usually these latter examples that are the focus of the current literature, along with specific XAI techniques, rather than any specific research investigating a more fundamental co-constructionist approach to underpin these.

Moving to more practical considerations, there is no platform identified that is available for easy use (e.g., available via open-source software) which supports exploration in areas proposed by all the gaps identified above *[Gap 4]*. Substantial technical effort would be required to demonstrate progress towards closing these gaps, and this would typically be with individual solutions for each case. The ability to run a formal experiment for sensemaking that encompasses these capabilities would also therefore be a substantial endeavour. Considerations for the operational tempo, and the need to support different levels of specificity, from low-cost and fluid information to more structured knowledge are important, along with the ability to move between these forms and formats easily. The material in Chapter 4 describes the open-source Cogni-sketch platform, created to instantiate the HAKF principles outlined in Chapter 3 along with some details of the implementation and a link to the open-source repository and associated plugins.

Finally, whilst there were several resources available providing the results of structured interviews and CTA with analysts, there was no attempt to identify potential functions for HATs to support typical professional roles in the future *[Gap 5]*. The material reported in Section 3.2 reports on a Design Thinking (DT) workshop carried out specifically to elicit such information and which, in conjunction with the gaps identified above, informed the relevant factors and required capabilities for HAKF to enable the creation of Cogni-sketch.

Many of the gaps reported above could be considered minor within that field of research, but when considered as a whole, and in the context of a sensemaking use case that draws together human users and machine agents working as a HAT to progress problem-solving tasks, the gap in both research literature, available tooling, and experimental results is substantial. Addressing these gaps and progressing understanding at the intersection of them is the scope of this thesis.

Table 2.1 below shows the research contributions (RCs) and research questions (RQs) as defined in Chapter 1, and the gaps identified in the earlier gap analysis summary, along with a reference to where in this thesis the material relating to that gap (and therefore the corresponding research questions and contributions) can be found.

The brief descriptions for each of the terms in Table 2.1 is listed below. Refer to Sections 1.3 and 1.4 for full definitions of each of the RQs and RCs:

- RC1 - Definition of HAKF
- RC2 - The Cogni-sketch open-source platform
- RC3 - Evidence of human user sensemaking
- RC4 - Integration of machine agents

- RQ1 - Human creativity
- RQ2 - Machine assistance
- RQ3 - Shared understanding
- Gap 1 - Experimental sensemaking results
- Gap 2 - Co-construction and affordances for HAT
- Gap 3 - Social considerations for XAI
- Gap 4 - Platform for efficient experimentation
- Gap 5 - HAT support for future expert roles

Contribution	**Question**	**Gap**	**Where addressed**
RC1	RQ1, RQ2	Gap 2	Chapter 3 (Section 3.3)
		Gap 5	Chapter 3 (Section 3.2)
RC2	RQ1, RQ2, RQ3	Gap 4	Chapter 4
RC3	RQ1, RQ3	Gap 1	Chapter 6
RC4	RQ2, RQ3	Gap 3	Chapter 5

Table 2.1: Mapping of gaps to research questions and contributions

2.6 Chapter Summary

This chapter has reviewed the various research areas that are relevant to the support of HATs working together to collaboratively progress challenging problems such as sensemaking and SU. The abilities for both the human users and typical machine agents have been investigated, with a particular focus on the ability for machine agents to provide explanations and explanation-related information in the form of XAI techniques. An important consideration is the way such explanations can be surfaced, and there is an important call from within the community to account for relevant work arising from the social sciences, especially in terms of the manner and style in which explanations are delivered. There is substantial literature in the field of problem-solving and sensemaking more specifically, and several of the main techniques were reviewed, alongside the observation that much of the insight from this area is driven by structured interviews and techniques such as CTA rather than direct experimental results.

The conclusion drawn from this literature review, and the identified gaps and opportunities, form the basis for the proposed approach described in Chapter 3, with Chapter 5 drawing specifically on the literature relating to explanations and XAI to motivate a pilot and subsequent evaluation for specific types of explanations. Finally, Chapter 6 uses a key sensemaking technique from this literature review to underpin a long-running pilot with an intelligence analyst, and subsequently design and execute a formal experiment with human users to measure their performance in completing a sensemaking task based on OSINT analysis.

Chapter 3

Human-Agent Knowledge Fusion (HAKF)

3.1 Introduction

This chapter describes the process that was followed to validate the need for an approach to enable fluid and extensible interactions between human users and machine agents, with a particular use case of sensemaking to provide a set of testable capabilities. It includes identification of the set of *required capabilities*, building on the earlier *relevant factors* identified in Section 1.1. This conceptual framework is known as HAKF and supports co-construction of information and knowledge to support such exchanges. This must be lightweight and flexible enough for human users, but formal enough for machine agents, enabling any instantiation of this to support experimentation and exploration of different relevant scenarios.

To briefly revisit the research questions for this thesis as described in Section 1.3: The conceptual framework proposed in this chapter directly addresses **RQ1** on human creativity (*"Can human users create multi-modal and semantically-meaningful information into an environment that is easily processable by machine agents?"*) and **RQ2** on machine assistance (*"Can such an environment support relevant machine agent capabilities to assist human users in their goals and provide additional task-relevant information?"*). It is the ability to support both human users and machine agents, and the various characteristics that are important to them, which underpins the definition of HAKF reported in this chapter. Information and knowledge must be multi-modal and able to be presented at a range of specificities, and the human users must be able to rapidly assess the capabilities of the machine agents in order to form trust (or otherwise) in them based on their performance or provided reputation.

The overall goal for HAKF is to enable rapid exploitation of task-relevant information and knowledge to inform decision-making activities. This must be realistically achievable in an operational tempo suitable for the task and support a variety of human user types as well as a range of machine agent capabilities. The operational tempo can vary from urgent and fast-paced (system 1), to more relaxed (system 2) and the different thinking and analysis styles of the human users must be accommodated across these. The human users may be novices or domain experts, but they will typically not be computer science experts or programmers capable of directly integrating machine agents themselves.

The goal of this chapter is to identify a system architecture to enable demonstrable synergy between human users and machine agents seeking to gain actionable insight and foresight in an ambiguous and rapidly developing operational settings.

3.2 Seeking stakeholder input through design thinking

In the early stages of this research a customised enterprise Design Thinking (DT) [81, 67] workshop was organised involving several research collaborators. A variety of military stakeholders were invited to formally gather their insights and requirements into AI and XAI systems [22]. There were 19 Subject-Matter Expert (SME) participants: 9 were serving military officers from the U.K. Army and U.K. Navy, 3 were U.K. government scientific advisors, with the remainder being other participants from U.K. industry. The focus of the workshop was the investigation of potential applications for AI and XAI capabilities across three different use-cases and personas defined by the military personnel as the first exercise in the workshop.

3.2.1 Participants and planning

The SMEs who attended the session represented a cross section of roles from the U.K. armed forces and U.K. industry who were highly skilled in their professional capabilities but who were not experts in computer science, AI or XAI. These represent the ideal *human users* for the focus of this thesis as defined in Chapter 1, specifically that they are "Usually domain experts not computer scientists". The military roles tended to the more senior ranks with a Major, Lt Colonel, Lt Commander and Commander present in the workshop. To help bridge the gap between AI expertise and military expertise we ran a short primer session at the start of the all-day workshop to brief them on the relevant high-level details for these technologies and embedded a researcher within each of the teams to help them understand the potential (and realistic limits) for AI assistance in future settings. The goal was not to teach AI and XAI but instead to set expectations on the art of the possible at a high level, suitable for the less technical domain experts within the workshop. The goal was to elicit valuable and realistic potential future opportunities in this space, but without unrealistic expectations of what can achieved from technology in that future time frame.

Prior to running this DT workshop, we considered alternative formats to source ideas for potentially valuable AI capabilities in the future, and reviewed these with colleagues, including:

- **User survey** - good for requirements, preferences, expectations and experiences of existing systems. Generally done in a one-to-one setting and assumes some level of knowledge or experience with the topic being surveyed. Has the benefit that quantitative data could be extracted if designed into the survey but came with a level of background knowledge expectation for the group that was unrealistic.

- **Focus group** - closer to our needs and the knowledge level of the group but more generally focused on broader topics, and less about specific examples or experiences. Ideal for gathering opinions, needs, and existing issues or challenges and carried out in a group setting with opportunity for discussion between the participants and cross-fertilization of ideas.

- **One on one interviews** - Like user surveys (and with the same concerns), but with the ability to potentially explore the field more generally if unstructured techniques are used, e.g., Cognitive Task Analysis (CTA).

- **Usability testing** - This was not appropriate at this stage in the research since we were seeking specific requirements, issues and opportunities to inform the most valuable research to undertake. However, a formal experiment with aspects of usability testing was carried out later in the research, to validate the ability for inexperienced human users to carry out OSINT analysis. This experiment and results are reported in Section 6.4.

Enterprise DT was chosen as a potential sweet spot for efficient extraction of relevant ideas for a broad technology area without trying to be too prescriptive of the specific functions or applications that should be considered. The user-centred methodology of the DT approach [81] also enables more conceptual and less functional ideas to be shared and can indicate areas for further exploration even if the actual solution cannot yet be articulated. Such workshops can be designed to accommodate stakeholders from different disciplines and with different levels of expertise. The level of experience required for active participation from experts, and the creation of valuable insights, is low. This therefore was an ideal approach for this early stage of the research and could lead to more conceptual or foundational ideas and provide a firm underpinning for any subsequent investigation.

In all the options listed above, and for the enterprise DT techniques used in the workshop, it is important to note that the pool of participants was small and drew heavily on their own relevant domain experiences. We recognise that the results of this exercise may not be representative of the wider population, but the desire for outputs that target more broadly applicable issues or themes was important regardless of breadth of representation. The findings arising from the workshop can be generalised to wider insights that can be addressed more fundamentally than for the specific use cases and personas investigated within this exercise.

3.2.2 Scope

The overall scope for the workshop was to consider a typical military environment. For example, working with local agencies or populations for security, or dealing with the aftermath of a natural disaster such as a flood or tornado. The teams were asked to consider how this is done today, and how it could potentially be achieved in the future (up to 20 years from the date of the workshop, so approximately in the 2030-2050 timeframe) with estimated technology advances, and specifically focused on AI technologies and how any specific XAI features or capabilities might be needed or used in this setting. This workshop is described in [22] with some of the more relevant summary material included in Appendix C of this thesis.

Whilst the participants were selected for their highly specialised knowledge of military systems and processes, the purpose of the workshop was to try to unearth more generic requirements or issues preventing possible adoption of AI capabilities in the medium to long-term. This extraction of more general findings was carried out through post-hoc analysis of the artefacts created during the session and was completed as a stand-alone exercise after the workshop itself had completed.

We ran a compressed set of DT exercises, focused only on a limited subset

Figure 3.1: A participant contributing to the big ideas exercise

of techniques at the earlier stages of the enterprise DT process. The aim was to ensure a rich set of generated ideas arising from development of the personas and consideration of their existing typical working environments. The group was split into three teams, and each carried out these techniques in sequence:

1. **Empathy mapping** - to develop a stereotypical user persona with a specific name and key attributes [50]. This persona is then used throughout the rest of the exercises and the empathy mapping considers what that persona *thinks*, *feels*, *says* and *does* whilst performing their role, creating useful hooks for the subsequent exercises. It was important for the military experts to pick a persona relevant to them and ideally to gain a range of personas across the three teams (but we did not give any kind of guidance or advice as to specific personas, ranks or roles to the teams as it was most important that they chose one which collectively resonated with them).

2. **As-is scenario development** - to consider what that persona does in their role today. This exercise defines the typical high-level *steps* and what the persona *does* to complete those steps and how they *think* and *feel* about it.

3. **Pain point analysis** - reflecting on the as-is scenario, this compressed version of pain point analysis helped to ensure that any negative or painful

aspects of the relevant task were captured, since these can often be the strongest indicators of areas for potential interventions or improvements.

4. **Big idea generation** - by considering the pain points (and any other relevant ideas, perhaps not arising as pain points) from the earlier exercises this is an ideation (idea generation) exercise that generates potential solutions to these. Hypothetical solutions can be identified by improving existing processes or by a more fundamental change that perhaps removes the pain point altogether by taking a different approach.

5. **Prioritisation grid** - the final exercise takes the big ideas generated previously and plots them on a simple two-dimensional space according to their feasibility (from expensive to cheap) and importance (from low impact to high impact) based on a consensus view from the team. The ideas that are both feasible and impactful are usually the top candidates to be taken forward into the later exercises, but for our workshop this was the final exercise with this weighting being useful to us to inform our later triage of the ideas for potentially useful insights and opportunities.

3.2.3 Outputs

These five exercises were carried out in sequence for three separate teams (A, B, C) with playback sessions from each team to the whole group after each exercise. Each exercise was based on an open discussion with all team members using each of the techniques, discussing ideas and contributing specific comments via post-it notes onto a shared team board. The workshop ran for the whole day with good engagement from all the participants throughout the exercises with over 600 post-it notes created across the five exercises for the three teams.

The teams were left to define their own personas based on their individual experiences and ranks. Each team therefore defined a different persona resulting in a range of roles spanning ranks as shown in Figure 3.2. These ranged from a junior ranked equipment maintainer (Team A), through a staff officer tasked with delivering the outputs of a team of analysts (Team B) and a commander running a joint operation with considerable influence (Team C). The diversity of rank, role and responsibility helped the teams to generate a broad range of relevant ideas in the later exercises. Full descriptions of the personas can be found in Appendix C, Section C.1.

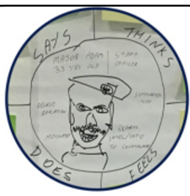
Team A	Team B	Team C
		
Corporal Palmer	Major Adam	Commander Brian
First-line maintainer	Staff Officer	Joint Ops, 1 star
26 year old woman	33 year old man	Man (no age specified)

Figure 3.2: Empathy-map personas created by each team

3.2.4 Findings

A post-hoc manual analysis of the materials created during the workshop was carried out during the weeks after the workshop. The focus of the analysis was the outputs from the final *big ideas* exercise, and the ideas ranked most feasible and impactful via the *prioritisation grid*. This analysis identified four topical clusters within the results as well as a general category for some of the ideas that were out of scope for the workshop but relevant to the participants. The out-of-scope category is not shown here. The four on-topic clusters were:

- **Machine agent** - AI and XAI was the focus of the workshop as advertised to the participants, so it is not surprising to see that some of the suggested ideas aligned directly to the capabilities of future machine agents within their environment, usually with some form of AI capability. The main sub-categories for this cluster were the ability of a machine agent to *learn* from data or experience, and the delivery of specific AI *functions* such as prediction or classification.

- **Human user** - this covers ideas relating to human issues or opportunities either with the task itself or with a potential future solution involving AI assistance. Sub-categories include *trust* (humans trusting the data or results from analysis of it), *impact* (humans being impacted positively or negatively by AI assistance), and *knowledge* or *understanding* (specifically in terms of gaining or sharing that knowledge with AI agents or learning new things to interact with AI agents).

- **User experience** - specifically, some kind of interface capability to enable interaction with AI agents, usually by chat, voice or custom visualisation such as diagrams, charts or map overlays.

- **HAT** - typically, a new or improved capability or process/technique that becomes potentially feasible because of human users working with new AI agent capabilities.

The analysis of the distribution of ideas across these four main categories is shown in Figure 3.3 and it is interesting to note that the most senior ranked persona (Team C) has the highest relative distribution of ideas considering HATs. The more junior ranked personas have slightly more ideas relating to the human user when using such systems and care more about the user experience specifically. Given the focus of the workshop it is not surprising to see that all teams created a large relative proportion of their ideas around specific machine agent capabilities that could be offered in the future.

Several of the suggestions from the highly ranked shortlist of ideas produced during this workshop directly informed the potential scope and subsequent research for HAKF, especially when considering the conceptual basis for exchang-

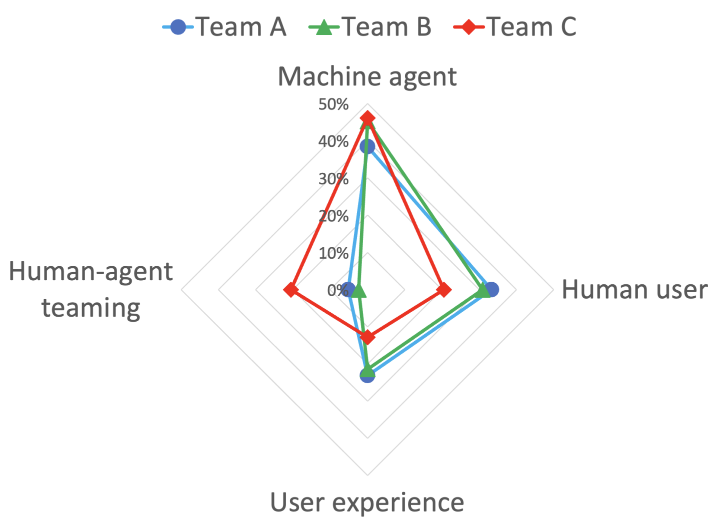


Figure 3.3: Category analysis of all big idea suggestions per team

ing[1] information between human users and machine agents in a scalable and flexible manner. A full list of the highly ranked ideas can be found in Appendix C (Section C.2) with the most relevant summarised below:

- The need to *ease the cognitive load of the human user*. For this the AI must help rather than hinder, and this could be an app or chatbot etc.

- *Easy access to data*, for prior examples, lessons learned, historical precedence for relevant issues, etc.

- Generic AI assistance to *help with reassurance, calming of the human*. Not domain specific or especially complex.

- Improved *confidence in data or decisions* through communication of provenance, analysis of completeness, use of a taxonomy of confidence etc.

- AI to *generate alternative ideas*, red teaming etc. Potentially a competing set of alternative AI agents each with different goals that can all attempt

[1]At this stage in the research we still referred to *information exchange* in this workshop, rather than *knowledge co-construction* which was a later highly valuable insight that arose when considering the conceptual basis for knowledge representation in HAKF.

to find weaknesses or attack the suggested courses of action.

- Super-powered AI agents: Able to *access material at a higher classification than the human* (without revealing/sharing), able to act as a *digital conscience*.
- The concept of a *life-long learning pair of human and machine agent* that train together and stay together throughout their career, learning about each other and honing their collaboration.

These items represent a broad set of future potential augmentation of human processes with AI-based machine agents. They span considerations of the human user in terms of cognitive workload and trust, and consideration of how some valuable machine agents could be completely generic and nothing to do with the human user's functional task but can still provide valuable contributions to the human through relevant interventions to provide reassurance or other generic assistance.

The suggestions recognise the need for additional (meta) information in addition to whatever the task-relevant information is, for example: provenance and confidence level, which can better inform the human user, perhaps leading to greater trust and confidence. Finally, and perhaps most importantly the suggestions recognise a very human-like set of potential characteristics for highly capable machine agents in the future. These agents may have access to more information that the human user is allowed to see, and they may be able to embed a wide range of additional considerations into their behaviour to help guide the human user about second-order or higher-level effects that might result from their actions and are perhaps unanticipated by the human. For example, they may act as a digital conscience or a devil's advocate. The idea that a human may train and serve with a lifelong machine agent that learns and grows with them is also a very tantalising idea, as is the sensible conclusion that a fall-back method is needed for all these capabilities in case the machine agent fails, and the human user has become overly dependent but must now function unaided.

It was interesting that most of the *big ideas* were for general machine agents and the ability to easily communicate contextually relevant information and corresponding meta-data between human and machine agents, rather than for specific functional needs that could be fulfilled in the future. Building any knowledge-based environment to support this would require substantial investment in general mechanisms for human-agent interaction, sharing of knowledge and the ability to rapidly build, or re-purpose existing machine agents into new settings or operational contexts as the situation evolves. This explicitly captures the recognition that this must be a dynamic activity, rather than having time or exact foresight to build such specific capabilities in advance. Reflection on these findings from the workshop led directly to the formalisation of several of the required capabilities necessary for the realisation of HAKF to support these kinds of future needs. From the relevant factors presented earlier in Figure 1.1 this DT workshop confirmed directly with expert human users that at least the following topics are of importance to any credibly useful solution involving machine agents in this space:

- *Trust* - and the ability to rapidly form it with machine agents operating in a particular setting.
- *Operational tempo* - some tasks are time-pressured, whereas others have more time for precision (planning etc), and the tempo can vary within a task over time.
- *Specificity* - the three different personas had very different problems to solve, and certainly for Team C (joint commander) the problems were less well-defined than for the lower ranks.
- *Knowledge sharing* - specifically, to achieve tasks but recognising that it was generally not just that one specific task that needed to be undertaken. With the *life-long learning* and *digital conscience* concepts it was clear that a powerful solution for handling contextual knowledge would be required.

Interestingly there was little focus on XAI capabilities specifically, but *explanations* were regularly mentioned in the context of communication, trust building, and HAT activities generally. The ability to enable or achieve explanation seemed important to this community, with an expectation that the explanation itself would always be possible, even though the technical SMEs in the workshop knew that explanations are not always possible for AI solutions without additional effort.

3.3 A conceptual basis for HAKF

Whilst the DT workshop was focused specifically on a military setting, with SME users related to that field, the goal for HAKF is to support a broader approach to enable more general capabilities. These are reflected in the set of *relevant factors* reported in Section 1.1, and will be subsequently developed into required capabilities for HAKF. This section describes this conceptual basis for HAKF and collates input from multiple sources including the DT workshop.

One important premise is that machine agents using AI or ML can augment human performance on a wide variety of challenging tasks, but to be effective they must be understandable and usable by non-expert human users. These users must often integrate different kinds of data from a variety of sources in an operational tempo, to attempt to make sense of potentially fast-moving dynamic situations, for example when performing SU or intelligence analysis [32, 115]. In earlier work [21] we have identified the need for broad and open solutions to support exploratory processes with both human and machine agents fulfilling distinct roles within the hybrid team.

HAKF is motivated by questions such as: whether it possible to define information systems that can be the backbone for rapidly integrated capabilities from both human users and machine agents in evolving situations where there is no predefined application or solution to that specific problem? Can the respective power of human cognition be tapped into, alongside machine agent processing capabilities? How much additional value can be achieved through co-construction into a shared knowledge graph to support this rapid but low-cost integration approach to collaboration? Is it possible to unlock the ability to combine generic machine agents into situation-specific applications with minimal technical effort by the human users who define and build such systems? The tempo of this is also extremely important as it is not credible to predefine an exact solution ahead of time, since there would be a vast permutation of machine agents and processes to cover all possible situations that may arise.

Given these questions and challenges, the highest-level conceptualisation of a system to support fluid and extensible interactions between human users and machine agents must be focused on the ability to exchange or share task-relevant information, or more specifically to collectively define information through co-construction. If information can be collectively defined through co-construction, then the ability to exchange it can arise from this *knowledge fusion* environment with little additional effort. This should support a range of *specificities*, including low granularity of information as needed, with small increments being possible by human users or machine agents to improve the structure or relevance.

Unlike existing solutions, particular information exchange mechanisms (such as conversation) are not the goal, but instead a broad knowledge fusion environment to support co-construction of information in a collaborative and dynamic setting [79]. This must support both *human users* and *machine agents* but does not need to always include both at all stages of the process. The environment must support human users collaborating (without machines) and vice versa, as well as single human user or machine agents using the environment individually, without collaboration. The ability to add or remove collaborators dynamically is also important and this can be easily achieved by simply managing access to the environment and ensuring that it can support multiple users and record the information or knowledge that is contributed by each. Access is important too, but explicit dissemination of knowledge, or specific sharing techniques are not a major research focus for this thesis. However, any knowledge fusion solution should support fine-grained access capabilities to assure knowledge creators that they can have full control over the assets they create and their dissemination.

The ability to define and share information exists at a low level today, for example distributed file systems or even centralised file systems with multi-user access, or simple document databases for the storage of structured data. These can be harnessed for collaborative use-cases, but specific interfaces are needed to foster collaboration between human users and machine agents.

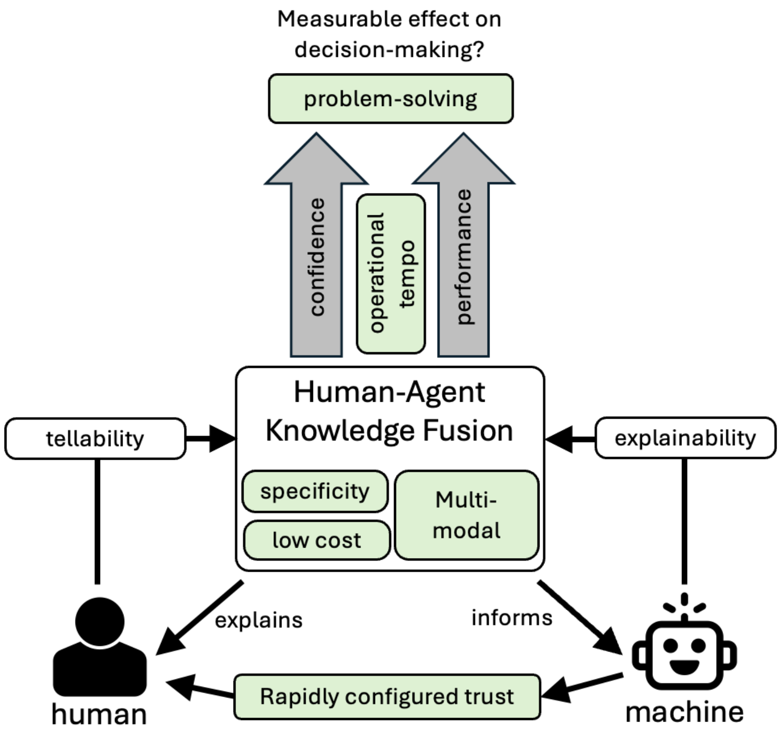


Figure 3.4: Human-Agent Knowledge Fusion (HAKF) - an expanded view.

Figure 3.4 shows the conceptual proposal for HAKF that builds on the high-level HAKF concept presented earlier in Figure 1.2. In this expanded version, the relevant factors for HAKF (See Figure 1.1) are also included, showing how each of these aligns. At the core of HAKF is the concept of *knowledge fusion*, supporting *human users* and *machine agents* in their collaborative creation of fluid information and knowledge through co-construction. This responds to the need for the agile integration of human users and machine agents from multiple sources into dynamic and responsive teams. HAKF is designed to support this deep interaction, comprising bi-directional knowledge and information flows to support meaningful communication between machine agents and human users [26].

Within the central component of *knowledge fusion*, the term *knowledge* is used to specifically capture the fact that both data, information and knowledge can be shared within the environment, and the exact category is often contextual. Therefore, as mentioned previously the *K* in HAKF signifies that *knowledge* is possible and explicitly supported, but HAKF does not exclude information and data as equally valid forms[2]. The creation of knowledge from information (or information from data) can be achieved incrementally (and potentially collaboratively) through the co-construction approach, with different agents providing different aspects of the context or meaning, enabling knowledge to be dynamically created, and that behaviour and capability is a key goal for HAKF. Also, the same information can be consumed differently by different human users or machine agents based on their own conceptual models or views of the world.

The basic premise behind HAKF is that such an environment must enable collaboration between human users and machine agents, specifically through two main affordances [39], represented as high-level communication flows: *tellability* and *explainability*.

Such a HAT environment must support a two-way flow of knowledge or information to allow all agents to contribute information (through tellability) and enable improved consumption of information (through explainability) to enable

[2]In the context of HAKF: *data* are raw facts and figures without any context, *information* is data that has been processed or structured in a particular context to provide meaning, and *knowledge* is actionable information according to the interpreter [83].

them to modify their internal models of the domain, as well as each other, and potentially their external behaviour as a result. These flows must be broad and encompassing and enable incremental refinement of information and knowledge as the emerging story iterates towards an improved form, rather than seeking to find a single universal truth in a particular setting [156].

HAKF systems with explainability aim to increase human user confidence (through transparency), and systems with tellability can increase machine agent performance (through rapid and explicit customisation or configuration within the same environment).

The ability to improve machine agent performance can include more ambient modes of configuration where the machine agents can modify their behaviour based on the relevant contents of the knowledge graph as they are created, rather than requiring explicit or direct reconfiguration. If such a system could be created within which multiple agents can rapidly share contextually relevant information, and individually react to the presence of information in that environment, it is possible that a measurable effect on decision-making for the team could be observed. This could be in terms of accuracy, speed, fidelity or some combination of these, with the potential to measure and quantify the improvement.

Creation of a hybrid environment that can drive improved confidence alongside increased generic machine agent performance or agility could provide a powerful tool for many knowledge-intensive applications such as sensemaking and SU. Especially when considering the *operational tempo*, such as in rapidly evolving congested and contested settings. Also, the desire to support *low-cost* capabilities, such as edge-of-network settings where it is infeasible to prebuild high-quality custom solutions to support the decision-maker in predictable ways with predefined machine agents with known APIs. Whilst HAKF is a necessarily high-level concept, the twin flows of tellability and explainability capture the basic mechanisms to underpin the general goal of *knowledge fusion*: the incremental co-construction of knowledge and information between human users and machine agents.

3.3.1 Tellability

This flow is for the case where new knowledge or information is conveyed from one of the agents to the system, often to impart useful and task-relevant information that could, if known, improve the performance of the system overall. Depending on the role of the agent within the system this information or knowledge could have an effect at any level and could be acted on by any other human user or machine agent within the system. Typically, tellability covers the creation of new information or the refinement of existing information, with consideration of provenance, certainty and confidence being key to helping establish *trust*. Both human users and machine agents can create knowledge or information via the tellability flow, but for simplicity in HAKF, as shown in Figure 3.4, this flow is shown as typically originating from a human user, and is most directly aligned to RQ1. In addition to providing new or refined knowledge or information via the tellability flow they also interact with information in the environment created either by themselves or others.

A key focus for HAKF is on supporting these agents through configuration of the system to rapidly apply it (or refocus it) on a particular situation without needing to build a whole new application each time. It is unlikely that machine agents (such as ML systems) can be retrained in the short time frame for the agile operations that are represented by the consideration of *operational tempo*, but through tellability the agent may be able to connect the lower-level generic classification (or other) outputs of machine agents to higher-level concepts that are driven more dynamically by the human users as the situation unfolds. It is this conflict between the cost and time-taken to retrain custom models, versus the generic capabilities provided by existing pre-trained models that drives the HAKF approach and the requirement for flexibility in the co-constructed information[3].

[3] As mentioned previously, the work reported in this thesis predates the advent of LLMs and popular applications such as ChatGPT, and has not attempted to integrate these more versatile and reusable models, but it is recognised that more general capabilities such as these

Clearly a superior technical solution would involve custom trained models and pre-built machine agents specifically for the exact situation and data feeds, but the operational tempo of the unfolding situation and the inability to predict the exact context in advance means that the luxury of these custom models and machine agents cannot be assumed. Instead, we must look at options and architectures for harnessing more generic pre-trained models and generic machine agents instead and harnessing them in-context quickly. In other words, one important capability that is enabled through tellability is the ability to rapidly add relevant contextual information, and this can then be immediately used by machine agents to modify their behaviour based on the context, assuming they were built with this flexibility in advance.

3.3.2 Explainability

Conversely, explainability provides a greater level of transparency into a conclusion or output from either a human user or machine agent within the HAKF environment. Typically, explanations come from machine agents, so for simplicity the explainability flow within HAKF is shown originating from the machine agent (and is most directly aligned to RQ2), but human users can also provide explanatory information. Amongst humans this is a familiar concept and is often invoked through *why?* questions and appropriate responses. Explanations can be well served through interactive discourse (verbally, textually, or through other means such as interactive visualisations) and for the machine agents this might be through traditional XAI techniques [6] such as attention or saliency highlighting, or description of configuration, highlighting relevant operating rules or constraints, or training data. They also have the potential to provide certainty information about any of these. For example, how much the input data aligned

may mean that custom capabilities could be created more quickly in rapidly emerging situations. However, they also provide potential additional exciting capabilities for integration within the HAKF environment, and these are outlined and briefly discussed in Chapter 7, Section 7.2.3.

with training data for the classifier [33], and other similar information that might better illustrate the context for that information or result.

The proposed HAKF approach supports dynamic explanations where a human user or machine agent can dynamically seek explanation for existing knowledge or information within the system by invoking specific explanation APIs for machine agents or contacting human users to ask for more information. Also supported is *pre-emptive explanation* where any agent can additionally contribute knowledge to provide contextual information in advance so that it is already present for review or analysis by other agents in the future.

These pre-emptive explanations and other similar explainability capabilities can be the basis for an interactive discourse between agents within the HAKF environment. This kind of conversational interaction involves both explainability and tellability flows and is explicitly supported in HAKF as an interaction overlay to the core co-construction basis for the knowledge graph. i.e., the basic information created in the HAKF environment can be consumed, by either human users or machine agents, in raw form as part of generally exploring the knowledge or it can be accessed via a different interaction modality such as a text chat conversation that refers to information in the knowledge graph to provide answers.

HAKF explanations can serve many purposes, with their goal in this multi-agent knowledge co-construction environment being informed by human explanations both in terms of form and intent (as surveyed and summarised by Miller in [95]). In some situations, the purpose of the explanation may be in the relationship forming phase, where the team members are assessing the capabilities of others. For machine agents in particular the purpose of an explanation requested by human users may be to assess the credibility of the activities performed by that agent and takes the form of *trust calibration* [150]. In other cases, the explanation may serve a broader purpose and form part of the ongoing development of the body of knowledge — knowledge expansion — thereby becoming new information in the knowledge graph and able to be further refined, linked or commented on by other agents within the system as part of the ongoing co-construction between the collaborating human and/or machine agents.

Intelligence analysis or analytic techniques embodied as machine agents can also be used to filter, fuse, and learn from data, extracting task-relevant knowledge from the co-constructed knowledge graph to assist the human users [15]. To enable the explainability flow and all the benefits that it can bring, machine agents should use AI methods that are explainable, to ensure that the combined human user and machine agent performance can improve through increased confidence from the explanations when compared to the human or AI machine agent working alone [10, 150].

HAKF supports these multiple intents for explanations whilst also recognising that the form of an explanation is important to the other agents and the operational context, assuming there is enough information in the knowledge graph. Flexibility for an initial explanation to be selected from that information is important. For example, a contrastive explanation as a starting position based on relevance and brevity, with the ability for agents to seek further details or alternative explanations being supported. This can be achieved through co-construction where new explanation information can be provided if possible and relevant. Since the act of explaining is a social process [95], it is represented in HAKF as a set of additional knowledge or information added to the knowledge graph which can be expanded and explored as needed, and the specific interaction format of the explanation is an implementation decision that can be considered an operational layer built as an extension to HAKF.

3.3.3 Towards measurable performance improvement

The HAKF concept can be used as the basis for tasks such as sensemaking (e.g., as identified in RQ3). A flexible approach like HAKF is most valuable in an operational tempo with rapidly evolving situations where low-cost solutions are necessary. In these settings it is infeasible to build in advance high-quality custom solutions to support the decision-maker in predictable and predefined ways. If the latter is possible then improved confidence and performance can be designed in, and specific tests for decision-making performance could be carried out before deploying and using such systems, but that is not the operational setting that has motivated the need for HAKF-based systems.

The potential for measurable performance improvement is an important aspect of HAKF and some explicit measurement of this is undertaken, as part of the quantitative analysis of user behaviour, mapped to aspects of a sensemaking process, as reported in Section 6.4.3. Refer to that section for relevant details on this analysis and the results. There are further experiments and tests that could be defined to more explicitly measure (and potentially quantify) any performance improvements, but these are not undertaken in this thesis.

3.3.4 Roles: For human users and machine agents

An important consideration for HAKF is the role of the human user(s) and therefore what they are trying to achieve within the system, including the types of information or explanation they require. Machine agents are also relevant for this consideration, both directly in terms of the ‘needs’ for different kinds of machine agents to enable them to interact with the system, but also from the perspective of the human users who will be working alongside them. These roles include direct users of any HAKF system as well as other stakeholders who may be more distant but still have defined requirements that can affect the behaviour or implementation of the system. Refer to [149, 16] for a deeper analysis of the different roles of the user and their specific needs, explained through a series of worked examples.

It is useful to understand and separate the different roles that may be required for HAKF systems, and to identify specific requirements for each. This builds directly on our earlier research to define roles to help clarify and articulate general explainability and interpretability requirements for collaborative HAT systems [149]. For human users it is also important to note that sometimes more than one of these roles can be fulfilled by a single individual, but it is still useful to separate their requirements based on role. In some cases, there may be associated accountability or audit related needs which need to be carefully considered if, for example, the same individual is fulfilling the role of operator and executor.

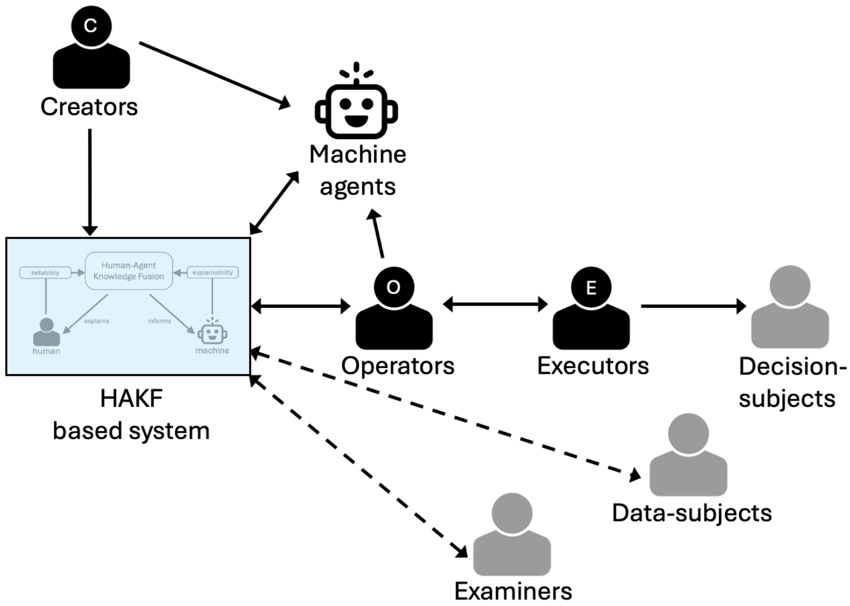


Figure 3.5: Roles for human users and machine agents using HAKF

Figure 3.5 is a HAKF-specific derivative extended from the original in [149] which was aimed at more general ML systems without the co-construction and collaboration capabilities of HAKF or the tellability and explainability flows. This figure shows the key roles for human users and machine agents in a typical HAKF system, with the main roles shown with black figures and the secondary roles shown in grey. These roles are briefly described below, extended from our earlier work in [149]:

- **Creators:** Create the HAKF-based system. Several teams of creators may work on different aspects of the same system e.g., architecture, design, implementation, training, documentation, deployment, and maintenance.

- **Operators**: Interact directly with the HAKF system, provide the system with inputs (*tellability*), and directly receive the system's outputs (*explainability*). In some cases, they may be able to interact directly with the *creators*.
- **Executors**: Make decisions that are informed by the HAKF-based system and receive information from *operators*.
- **Machine agents**: Any machine processes that can interact with the HAKF-based system, working alongside or for the human operator users.
- **Decision-subjects**: Anyone affected by decision(s) made by the *executor*(s).
- **Data-subjects**: Anyone whose personal data has been used to train any of the models used by machine agents in the HAKF-based system.
- **Examiners**: agents auditing or investigating the HAKF-based system. Depending on the system, they may interact with one or more of the other roles and the system itself. Usually this only occurs when the system is being audited/inspected.

We return to the topic of roles later in Section 4.3.1 where a further refinement for some of these roles is explored in more detail, within the specific context of Cogni-sketch as a particular HAKF-based system.

3.4 HAKF required capabilities

This section outlines a short list of *required capabilities* that would be necessary for any implementation of HAKF to fulfil the *relevant factors* identified earlier (See Figure 1.1) for human users and machine agents to work together. These required capabilities typically aggregate a set of the relevant factors, and in some cases some of the relevant factors span the required capabilities. HAKF is explicitly defined as a high-level concept that can provide a strong set of principles for development of future HAT systems that are flexible and extensible, but providing the required capabilities listed here adds a further layer of specificity to aid anyone considering an implementation based on HAKF. The recognition of these required capabilities can be thought of as a step towards a more implementation-oriented perspective for the necessarily high-level and general HAKF concept. These are listed in order of centrality to HAKF, and it is recognised that there could be more of these required capabilities added, or existing required capabilities could be split or further refined. This flexibility is inherent to the HAKF concept and an important principle to enable future extensions, and applicability in a broad variety of domains and applications.

The mapping of the required capabilities to the previously defined relevant factors is shown in Figure 3.6, with each of the visualised required capabilities (a-e) described in more detail in the subsections below. There is also a generic and non-functional required capability (f) that is listed but deliberately not shown on the diagram as it applies universally.

Since the human user and machine agent feature in many of the requirements they are left separate in the diagram but are mentioned where relevant in the text descriptions below. As are the two flows of tellability and explainability which are almost always both present in any implementation of HAKF.

The HAKF required capabilities are listed below:

3.4.1 Rich knowledge representation

HAKF requires *knowledge fusion*: a flexible and extensible information co-construction mechanism to support collaboration between human users and machine agents. An ideal mechanism for implementation is as a knowledge graph that can contain nodes and links, with labels and any number of properties on both. Nodes on

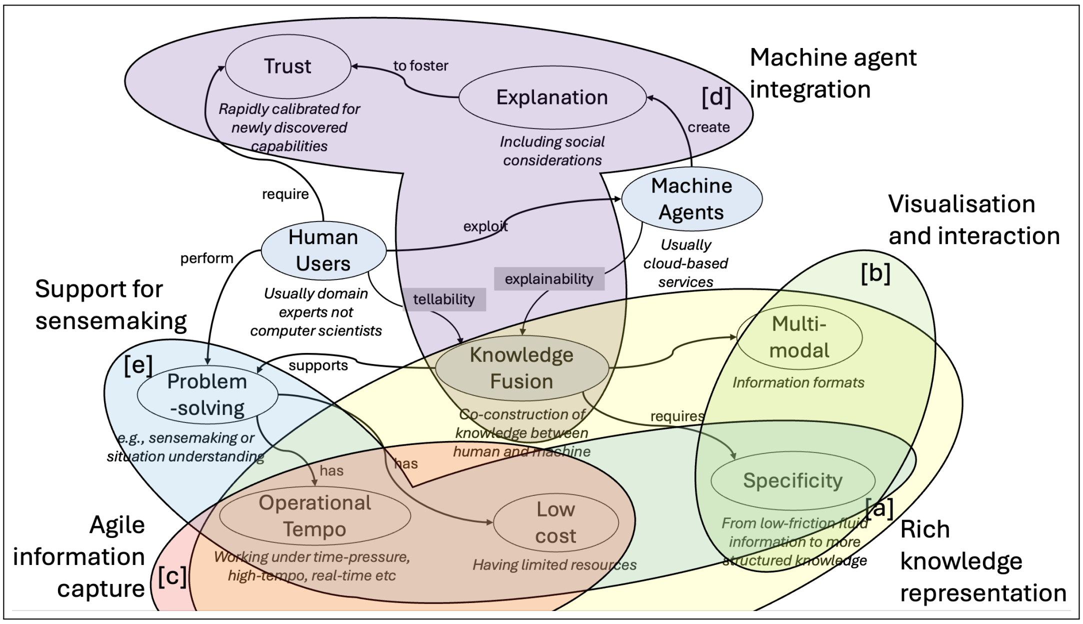


Figure 3.6: Mapping required capabilities to relevant factors for HAKF.

the graph will have *types* that can contain optional semantics, with the types being dynamically extensible during usage to enable freedom of expression by any human user or machine agent. These semantic types can correspond to an ontology or any similar structured model resource, and the types can be drawn from existing common models where possible, or specifically created for a unique task when needed. The starting point should be simple, with the minimum set of types to distinguish between concepts being captured in the graph. Whilst it is possible to add more advanced semantics it is important to recall the human user (typically a domain specialist and not a computer scientist or ontology expert), and their ability to understand advanced semantics may be limited and this should not be overlooked as it could lead to cognitive overload for the human users. There will likely be an ideal level of semantic expressivity for any given task and set of human users and machine agents, and it is important that this level of semantics can be dynamically set within the system in each case.

The core knowledge graph enables human users and machine agents to create task-relevant information and knowledge, but the raw form of the knowledge may be insufficient to achieve easy understanding for the human users, or efficient processing for the machine agents. Therefore, different modes of interaction with the knowledge graph will be needed, and this rich knowledge representation required capability must support such interactions (with other requirements taking advantage of the extensibility potential to enable specific capabilities). This can be supported through the development of *plugins* and those can be defined within the specific implementation. Typical modes of interaction will be needed, and these will likely be different between human users (e.g., graphical layout, textual summary, geographical layout, temporal logic, conversational interaction etc) and machine agents (e.g., logical inference rules to infer new nodes, links or properties, data fusion and information association etc). There should be no inherent limitation in the complexity of these, with the core knowledge graph providing the simple but powerful mechanism to capture the knowledge alongside a flexible mechanism to create these extensions as needed.

Collectively these capabilities can be considered as fulfilling the requirement for rich knowledge representation, and that is predominantly underpinned by the *knowledge fusion* core of HAKF. Providing an extensible and shareable semantic knowledge graph as the basis on which all agents (human and machine) interact, with the ability to both read and write knowledge from or to the graph. Also relevant to rich knowledge representation are the factors of *multi-modality*, *operational tempo*, *low-cost* and *specificity*, and the ability to create knowledge and information that account for these factors is essential. They are described in detail in the other requirements below, but the need to account for them in the rich knowledge representation required capability is noted here.

The rich knowledge representation required capability is driven by *RQ2*, and the need to interact with it in human-friendly ways, support an appropriate level of semantic precision, handle larger volumes of data etc are informed by *RQ1*.

3.4.2 Visualisation and interaction

Storing the knowledge and information in the form of a knowledge graph is necessary, but it is also important that the human users and machine agents can interact with that information in meaningful and relevant ways, especially at scales where the volume of raw information may be overwhelming for human users. This requires a minimum set of mechanisms for visualising the information in the knowledge graph to human users and must recognise that as the information grows larger so the ability to focus on specific subsets of the information is important, as is the ability to search or filter. For machine agents this will take the form of APIs to access the data in the knowledge graph programmatically, and these should be designed to be efficient and extensible to enable specific application relevant variants as needed in the future. For human users this may include the ability to define larger knowledge constructs and therefore capture and render underlying information at different levels of *specificity* according to the task they are undertaking.

For example, HAKF can be used as the baseline to allow specific extensions for higher-level interaction capabilities such as the support of conversational explanations [29]. In this example the focus is on the explainability flow to enable the human users of the system to interact with the knowledge graph data in the form of a conversation, enabling machine agents to provide explanations of results arising from ML classifications directly into the knowledge graph, but then surfacing them via conversational interaction. Other examples could include timelines for visualisations of temporal data within the knowledge graph, or geospatial data rendered dynamically on maps, as well as contextually relevant information rendering such as social media data (e.g., tweets) being rendered into the environment in the familiar visual format used within Twitter itself).

This focus on useful visualisation formats to better support the human users also extends to the common information modalities, so *multi-modal* support is needed for easier consumption by human users in the relevant native modality (e.g., for image data, text, videos, graphs, etc). Also, both human users and machine agents must be able to easily create or edit the *multi-modal* information within the knowledge graph. For the human users this may take a variety of forms depending on the type of data and any applications being used, but for example, text data is best supported through familiar UIs such as existing text editors, whereas video data can be played through a video player, and image information rendered directly as pictures to the human user.

The support for human-friendly visualisations and specifically the support for layout choices, colours and other specific formatting decisions is driven directly by *RQ1*.

3.4.3 Agile information capture

The previous two required capabilities have defined the knowledge graph, the ability to store a wide variety of types of information, and the need to interact with it using techniques that may aid human understanding and support efficient machine processing. The need for agility, and specifically the ability to seamlessly capture information is equally important to these (and often overlooked).

This requirement directly acknowledges both the *low cost* and *operational tempo* factors and recognises that, for human users in particular, they must be able to capture raw data into the environment very quickly, and if relevant they can spend time later creating additional information related to that or placing it into a more complex knowledge structure. The agile capture of data must also account for relevant contextual information such as provenance data relating to the user who created or modified it, when they did so, where the original data came from etc. Whilst aimed mainly at human users, and specifically to enable them to operate at a fast rate when processing simpler data, the same considerations also apply to machine agents, especially the need to capture relevant provenance information and operate at a tempo relevant to the task and accounting for the human users.

This ability for agile information capture in some format that is relevant to the meaning of the data is especially driven by the need for human-friendly representations (*RQ1*). The previously noted ability for machine agents to easily create and access contextually relevant information is also important here too, but it is deemed to be a secondary need.

3.4.4 Machine agent integration

There may be multiple modes in which machine agents can interact with the HAKF system, and all these will be facilitated through knowledge fusion, and more specifically through the reading or writing of task-relevant knowledge into the graph. Machine agents may be *autonomous* (able to operate undirected and able to respond to information when created), or they may be *directed*, and triggered by human users in a particular context. There may be any number of machine agents in any given system, and they may each bring their own special capabilities to the task. Access to the knowledge graph will be provided via APIs as described in the rich knowledge representation required capability.

Trust is an important consideration for the human users of the system and given the wide range of potential machine agents and their capabilities, it is important that only an explicitly approved set of machine agents are granted access to the environment and the data within it. This can include a predetermined subset of the data, e.g., based on fine-grained access rules. These permissions for individual machine agents must be explicitly granted for each environment and have their interaction mechanisms defined within the system. Directed machine agents will be available as tools to be triggered manually by the human user when needed, whereas independent autonomous machine agents can 'watch' the knowledge graph for relevant information and interject when they can make relevant contributions to the environment through their processing. Any such interjections are made as new knowledge added to the graph. *Trust* will also be built by the human users based on the behaviour and contribution of the machine agents within the system.

By providing a HAKF-based system for machine agent interaction it does not rule out direct machine-to-machine communication via other mechanisms (e.g., [91]) in the same way that it does not exclude human users from interacting directly with each other outside of the system. The ability to connect multiple machine agents and enable those agents to efficiently contribute their results, for example as additional data back into the knowledge graph, is also important. Such contributions may include *explanations*, and these may take multiple forms including additional contextual data added to the knowledge graph, or specific results of XAI processing. Explanations can help the human users better understand the meaning and relevance of data contributed by machine agents and thereby potentially increase their *trust* in those agents.

The ability for machine agents to participate in the environment, but in a carefully controlled and managed way, arises from *RQ2*.

3.4.5 Support for sensemaking

Whilst HAKF can be used for any knowledge or information co-construction problem-solving task, within this thesis the scope is deliberately limited to that of sensemaking, SU and OSINT analysis. Any system built on the principles of HAKF will be able to support a much broader set of problem domains, but these are not covered here. This required capability is therefore focused on sensemaking and is a good example of the kind of specific required capability that may be needed for particular applications and builds on the more generic required capabilities listed previously.

Simplistically, the ability to perform sensemaking for human users draws heavily on the previous *agile information capture* required capability in terms of specific factors that are needed to enable it. However, since sensemaking requires the processing of information from multiple sources to make inferences about states of the world, it may also draw on the *machine agent integration* required capability if there is the potential for automated support in the process itself, or in processing any of the data sources etc.

When such a system is hybrid and comprises both human users and machine agents, this means that the system must, at least at some level, operate in terms of human-understandable concepts and relationships as described for the rich knowledge representation required capability. These can provide the basis for a potentially powerful deep integration between the machine agent and human user elements of the wider system. For example: to quickly and efficiently harness the potential power of advanced machine processing capabilities without requiring that the human users are deep specialists in the specific technical fields. Instead, the human users see directed machine agents available and can discover their purpose and use them in context without needing to configure them in detail (but the option to do so may also be available if needed).

The ability to progress a problem-solving task such as sensemaking arises from *RQ3*, and for simplicity only the human user aspects are highlighted in Figure 3.6 since the machine agent integration aspects only apply to sensemaking cases when machine agents are used and map entirely to the separate machine agent integration required capability.

3.4.6 Novelty, feasibility and open access

This final required capability is generic and applies to all aspects of HAKF and provides non-functional as well as functional requirements for any implementation. It is therefore not explicitly drawn in Figure 3.6.

For all the previous required capabilities it should be noted that when taken individually they are not necessarily novel and can often be found in other systems or approaches. It is the unification of them, and the basic approach of HAKF as a building block for designing specific applications that brings the novelty and powerful flexibility. The core HAKF concept can enable human users and machine agents to collaborate through co-construction of knowledge and information and the required capabilities described previously enable substantial extensibility. If any of these are missing then HAKF will be prevented in some way, and the potential flexibility and extensibility may be harmed. The ability to create a system that supports all the above required capabilities in order to enable a HAKF solution is believed to be unique amongst existing open-source software capabilities even though some of the individual principles can be found in some existing solutions today.

An important consideration for HAKF is therefore extensibility and the ability for a set of models and machine agents to be produced and reused by different communities to further accelerate the efficiency and speed-of-implementation for HAKF-based solutions in the future. Typical open-source software contributions are an excellent way to create and share such components, and in the following chpater the open-source Cogni-sketch platform is introduced as an exemplar of a HAKF-based system.

3.5 Chapter Summary

In this chapter the concept of HAKF has been defined, starting with the basic high-level concept, and then aligning the *relevant factors* identified previously, culminating in a set of *required capabilities* that draw these together into distinct but inter-related groups. The results from a DT workshop with military stakeholders were reported, with an analysis of the findings presented. These findings helped to identify the relevant factors and inform the required capabilities and clarify what the scope and role of HAKF should be in the context of supporting the design and implementation of more rapid and efficient applications for HATs.

The three research questions (RQs) identified in Chapter 1 have been mapped to the HAKF concept in this chapter, specifically against required capabilities as shown in Figure 3.6 and summarised below:

- The rich knowledge representation required capability is driven by *RQ2*, and the need to interact with it in human-friendly ways, support an appropriate level of semantic precision, handle larger volumes of data etc are informed by *RQ1*.

- The support for human-friendly visualisations and specifically the support for layout choices, colours and other specific formatting decisions is driven directly by *RQ1*.

- The ability for agile information capture in some format that is relevant to the meaning of the data is especially driven by the need for human-friendly representations (*RQ1*). The ability for machine agents to easily create and access contextually relevant information is also important here too (*RQ2*), but it is deemed to be a secondary need.

- The ability for machine agents to participate in the environment, but in a carefully controlled and managed way, arises from *RQ2*.

- The ability to progress a problem-solving task such as sensemaking arises from *RQ3*.

Further details regarding the exact answers to these research questions (RQs) can be found in Chapter 7 along with a summary of the research contributions (RCs), including a mapping between RCs and RQs as shown in Figure 7.2.

Chapter 4

Cogni-sketch: an experimental instantiation of HAKF

4.1 Introduction

This chapter introduces the experimental instantiation of HAKF known as Cogni-sketch which demonstrates both the feasibility of a HAKF-based solution and investigates the novelty of such a capability within the crowded space of existing solutions for different aspects of the required capabilities. The Cogni-sketch platform is described, along with the solution features that have been implemented to support the required capabilities of HAKF. The open and extensible architecture is explained, along with details of the various extension points (panes, windows, functions and palettes) that have been designed to support extension of the core platform.

The HAKF roles defined in the previous chapter are revisited in the context of Cogni-sketch, with specialisations for some of the user roles defined. Cogni-sketch has been implemented as described in this chapter and released as open-source software with a few commonly used plugins also released. It has been developed to a point of maturity where substantial evaluations and experiments can be supported as reported in Chapters 5 and 6. Novelty is asserted based on the earlier literature review in Chapter 2 combined with a thorough assessment of relevant tools and techniques in Section 4.2.

Additional information relative to Cogni-sketch can be found in Appendix A.

Specifically, details relating to the Cogni-sketch environment, including links to videos, copies of data, other examples of usage and other related material.

4.2 Analysis of existing capabilities

Given the goals of HAKF: to support human users and machine agents in their sensemaking activities through knowledge co-construction via the explainability and tellability flows, it is important to consider the state-of-the-art for existing tools and techniques in this space. This is an important exercise to ensure novelty and emphasise that it is the amalgamation of all relevant capabilities for HAKF that presents the opportunity for achieving these goals, rather than just an existing or convenient subset.

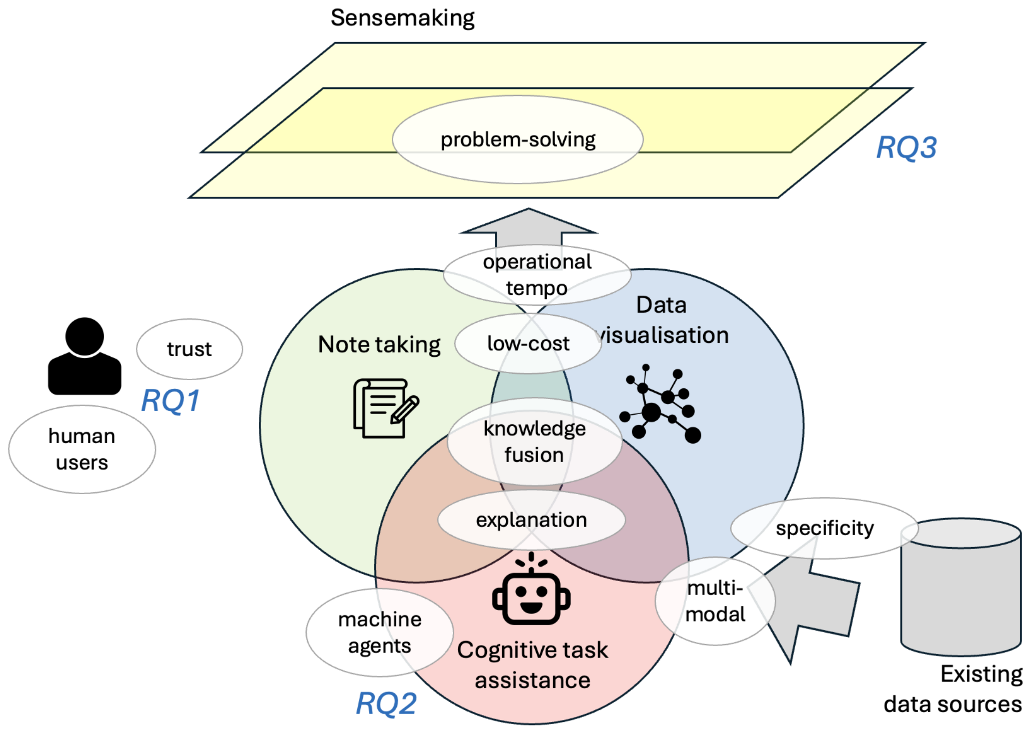


Figure 4.1: Intersection of related sensemaking needs

While Chapter 2 surveyed the relevant literature for the HAKF relevant factors, our focus here is on functional tools and other implementations of existing capabilities (e.g., software libraries or practical techniques).

Figure 4.1 shows the typical operating context for a system that may attempt to support or manage sensemaking activities. It is annotated to show the research questions (RQs) and relevant factors (See Section 1.1), and how they relate to the different parts of the diagram. Given the necessarily general nature of HAKF and the desire to support a broader set of use-cases, any specific existing systems built purely for sensemaking or SU, are not considered in this exercise. Instead, systems or approaches that represent the interconnected abilities to capture and visualise information, provide facilities for note-taking and information capture from either human or machine users, and integration with cognitive task assistance for analytics or other relevant tasks are covered here. Also, the recognition that much of the relevant data for the task will likely already exist in other sources and will need to be linked to where relevant.

It is the aggregation of these capabilities and the benefits that arise from their combination that is the area in which HAKF, when tailored for sensemaking, can specifically add value. A unified solution based on all these capabilities with clear extension and plugin points is asserted to be a better solution than a collection of specific or rigid tools with individual integrations to achieve the same outcome. The latter is often the case today, so an ideal solution would support a flexible and extensible core that can support this sensemaking focus, but also many others.

It is also important to consider that maximum extensibility will be enhanced by any such environment being available as open-source software to allow the community not only to contribute additional plugins and extensions, but to also extend the core as needed and to maintain the software as relevant standards, components and techniques evolve over time.

So, in summary the intersection of these three related areas of data visualisation, note-taking and cognitive task assistance, are where the value of a sensemaking application of HAKF is perceived to be.

Each of the intersecting sensemaking needs in Figure 4.1 are briefly discussed in the sections below:

4.2.1 Data visualisation

The ability to store networks of data and subsequently visualise it is a common capability found in many systems, frameworks and tools, both open and closed source. There are too many to list here, but many are simple visualisations of existing data as a static or interactive graph, and do not allow that graph data to be edited or extended as part of the visualisation. Those that do allow modification or extension of the data often don't provide much visual customisation for the human user, for example by extending the visualisation for whole classes of data or configuring the rendering for individual nodes in the knowledge graph. Since there are many libraries and tools for building and rendering knowledge graphs, any which don't allow any level of user control over visualisation are discounted.

A good example of graph editor with support for customisation and extension is Neo4J Bloom[1] which: "...allows users to visually explore and manipulate data stored in a Neo4j database[2], using a drag-and-drop interface to create and edit nodes and relationships. The Graph Editor[3] also provides features such as auto-layout, filtering, and search, which enable users to easily navigate and explore large datasets[4]".

Other examples of graph visualisation tools or environments include: Cytoscape[5] (for biometrics data), Tableau[6] (a powerful visualisation-only tool), Microsoft Power-BI[7], Google data studio[8] and more. For developers who wish to write specific visualisations using a low-level library, there are a range of options

[1] See `https://neo4j.com/product/bloom/`.

[2] See `https://neo4j.com/product/neo4j-graph-database/`.

[3] See `https://neo4j.com/docs/bloom-user-guide/current/bloom-tutorial/edit-graph-data/`.

[4] From: `https://neo4j.com/developer/tools-graph-visualization/`.

[5] See `https://cytoscape.org/`.

[6] See `https://www.tableau.com/`.

[7] See `https://powerbi.microsoft.com/`.

[8] See `https://datastudio.google.com/`.

including D3[9] (which is used as the graphical baseline for the Cogni-sketch environment), as well as a range of specialised graph visualisation packages from Cambridge Intelligence such as Keylines[10], ReGraph[11], and Kronograph[12]. The breadth provided from just Cambridge Intelligence shows the proliferation of approaches for different environments and different types of graph data. An ideal solution to instantiate HAKF would be one core extensible library that can be customised as needed for each use-case and then be re-used in any example that required the same type of visualisation. This minimises the effort for both creators and operators alike but is not available in any of these data visualisation components with the exception of the low-level libraries which are too time consuming to customise from scratch in each engagement.

A recent example that creates a consumable and accessible web-based visualisation of data in a traditional database is AirTable[13]. The database schema is readily extensible to handle new requirements, and custom interfaces to interact with or create/edit the data are easy to define. The main limitation with AirTable is that the data is inherently structured in the form of relational tables with columns and rows rather than a full knowledge graph implementation which is much more flexible but harder to easily and consistently visualise that the simpler table-based structure. AirTable provides an excellent user experience but is limited in the kind of semantic expressivity that is required to maximise the potential for machine agents to effectively participate in the environment. Similarly, LucidApp and LucidChart[14] are well placed for collaborative construction of drawings and workflows, but by human users only and with limited ability to define underlying semantic models for the data.

[9] See `https://d3js.org/`.

[10] See `https://cambridge-intelligence.com/keylines-javascript-graph-visualization/`.

[11] See `https://cambridge-intelligence.com/regraph/`.

[12] See `https://cambridge-intelligence.com/kronograph/`.

[13] See `https://www.airtable.com/`.

[14] See `https://www.lucidchart.com/`.

4.2.2 Note-taking

There are a wide range of note-taking applications available ranging from highly configurable and popular products such as MS-Word and powerful but basic platforms that use the simple 'markdown' format [106] such as Notion[15] and Obsidian[16]. Many of these tools are available locally or via online services (sometimes both). Some provide substantial advantage in terms of typesetting and layout, while others provide powerful but simple mechanisms to not only record data but to easily create links between data via the markdown format. Many of the markdown-based tools provide rich capabilities to explore the links in the data and to visualise them, and through simple conventions the users of these tools can create large graphs of nodes and links by simply typing text in the markdown format.

There are no known note-taking tools of this kind that allow more structurally detailed meta-data to be provided for the links, or to visually edit the nodes or the links, or edit their position in a visualisation, and any attempt to support a palette or ontology of types is achieved by convention within the format rather than explicitly being able to be defined. It has been observed that there is much more potential value to be had from improved capabilities for complex linking of data [9] but the techniques to achieve this are not straightforward when using the current text-based markup approaches.

In the example context of sensemaking it is very common for analysts today to use existing products such as MS-Word or MS-PowerPoint to capture their data as they collect it. This has the benefit of being easy to do and very flexible, but accessing the data later in a structured way, or providing access to machine agents to provide cognitive task assistance is not possible without substantial effort. For a note-taking tool to be compatible with the principles of HAKF it must provide easy access to read and write data for both human users and machine agents, all

[15] See `https://www.notion.so/`.

[16] See `https://obsidian.md/`.

of whom can contribute semantically relevant information with optional named properties that conform to a task-relevant and extensible schema or ontology for the data.

4.2.3 Cognitive task assistance

The ability to apply cognitive task assistance to data is easy for developers, and for end users it can be done with predefined cases that are supported in the tools that they are using, either directly or by exporting their data. For example, a user may wish to run a machine learning algorithm on tabular data that they have in a Microsoft Excel spreadsheet. If they are a developer, they can write python code to do so, either directly using the spreadsheet file, or after extracting it.

Some software products embed machine processing capabilities directly within them, for example to identify objects or people in photographs, or to clean up noisy audio. However, in these cases the machine processing has been defined in advance and is designed to work with that type of data. In some cases, tools provide plugin points where new agents can be dropped in as long as they conform with the Software Development Kit (SDK), and in the most extreme cases there are environments designed to build workflows that can fuse code and/or machine agents with data. In short, the ability to connect a wide range of machine agents to data and then successfully read and write that data in a manner that is compatible with concurrent human users is usually complicated and requires specific technical software development skills not usually found in typical operator users.

HAKF cannot remove the need for software developer skills completely, but it can separate the need for them between creator and operator roles and minimise the amount of effort required, and provide standard interfaces for integration of machine agents, as well as a marketplace/library approach for re-use and/or extension of existing agents that may already implement many of the required capabilities.

4.2.4 Sensemaking and shared understanding

As explained in the introduction to this section there is no major focus on dedicated proprietary tools for sensemaking and SU because the focus of core HAKF stands alone from the sensemaking and SU use-case. However, there are a small number of capabilities and approaches that can be used for sensemaking and SU and can also be more generally applied, so these are listed briefly below as they do meet the relevance criteria as a result of this broader applicability.

For sensemaking specifically there is a deeper dive into the Pirolli and Card sensemaking loops [113] in Section 6.2.3 as well as the broader material from the literature review in this area as reported in Section 2.4.

i2 Analyst Notebook

i2[17] Analyst Notebook[18] represents the closest overall identified match to HAKF when applied specifically to sensemaking and SU. It is a graph data visualisation desktop tool designed for data analysis and visualisation with some abilities for data collection. It supports a flexible data model that can be extended with additional features. The notebook supports multiple views of data and linkages between them, enabling different perspectives onto the same underlying graph along with the creation and sharing of visual briefing charts. It has the significant advantage that it has a large user base and has been extensively tested and used at scale for real operations by many organisations including police forces and intelligence agencies. It is however limited in terms of extensibility and is not open-source software. The advantage of maturity and stability therefore is offset by limited extensibility and accessibility.

[17]Previously owned by IBM. I mention this for transparency as IBM is my employer.
[18]See `https://i2group.com/i2-analysts-notebook`.

Cynefin

The Cynefin framework [80] is different to the previously mentioned items because it is a technique rather than a tool, but it is mentioned here specifically because it is close to the spirit of the open and extensible activities that can be supported by HAKF.

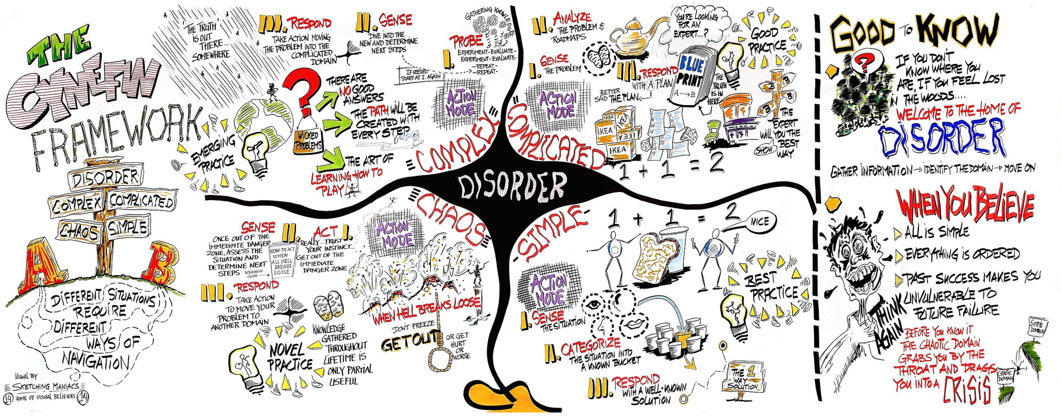


Figure 4.2: Sketch of the Cynefin framework[19]

A drawn sketch of the Cynefin framework can been seen in Figure 4.2 and the intentionally chaotic and informal style of the sketch is a useful reminder that HAKF tools should support human creativity and the creation of a sketch like this, but with structured data behind it too.

The Cynefin domains [143] of *clear*, *complicated*, *complex*, *chaotic* and *confusion* (previously named *disorder*, and shown as such in the figure) have some correlation with the phases of the foraging and sensemaking loops in Pirolli and Card [113] as well as corresponding to the matrix of "knowns and unknowns" ([127]). These Cynefin domains would also likely benefit from different palettes, plugins and machine agents in a HAKF environment, and could be investigated in the future with the existing open HAKF architecture providing multiple inte-

[19]Sketch of the Cynefin framework: (`https://en.wikipedia.org/wiki/Cynefin_framework#/media/File:Cynefin+framework_by_Edwin_Stoop.jpg`) from Edwin Stoop, licensed under CC BY-SA 4.0 (`https://creativecommons.org/licenses/by-sa/4.0/`).

gration points to support the Cynefin domains via plugins.

4.2.5 Summary of existing tooling

In summary, there are a variety of tools and techniques that can be applied to the collaborative collection, processing and visualisation of information and we have considered the inter-related categories of note-taking, data visualisation and cognitive task assistance. There are also some tools that are specifically aimed at sensemaking and SU, but only those few with a broader applicability have been reviewed. In aggregate these existing tools at least partially cover all the relevant factors and required capabilities for HAKF, but no single tool or environment has been identified that provides all these capabilities:

- The ability for human users to creatively capture their information directly into a machine processable format whilst retaining artistic freedom to layout or construct their knowledge graph in a style that appeals to their human desire and ability to consume detailed information quickly (*RQ1*).
  Agile information capture, Visualisation and interaction, Rich knowledge representation.

- Direct access for machine agents to both act on the existing data and contribute or modify that data based on their processing, as proposed in HAKF (*RQ2*), requires a streamlined solution that is not widely supported. Deep technical skills tend to be needed to achieve integration with machine agents, and if these agents cannot be directly embedded into the environment, then export of existing data is needed, with corresponding control and version tracking, as well as for the subsequent import of any machine generated information, creating a substantial amount of additional friction to the end user.
  Machine agent integration, Rich knowledge representation.

- Specific support for sensemaking (*RQ3*) but only as a specific example that is built on the open and extensible HAKF base. The tools identified that specifically apply to sensemaking do not have this open and extensible HAKF base and are not open-sourced, but they do provide highly relevant capabilities for sensemaking and SU.

Support for sensemaking.

HAKF enables all of this to be achieved in a simple core environment with multiple plugin points to enable a wide variety of different extensions and machine agents to be contributed by the community for sharing, modifying and remixing.

The ability to visualise data is widely supported, as is the ability to easily take notes, and the ability to run machine processing on data. But the ability to do all three of these things in the same environment is not. Even with just visualisation: the ability to interactively create and edit data via the visualisation is far less common, and the ability to easily create a task-relevant palette of semantically meaningful types alongside this and have full control over the visual rendering of the graph whilst making the data available in real-time to machine agents who can also contribute edits to the graph is not possible in any available software at the time of writing. Therefore, the novelty of the proposed HAKF approach across all these dimensions is high since there is no identified open-source implementation that currently supports these requirements.

Having explored the relevant existing tools and techniques that correspond to HAKF, and identified the need for a unifying implementation, in the next section the open-source Cogni-sketch implementation to address these gaps is defined.

4.3 Bridging the gap: Cogni-sketch

HAKF, as described in the previous section, is deliberately high-level to allow freedom of implementation whilst providing a novel but feasible solution for the three research questions that are the focus of this thesis. In this section we outline the Cogni-sketch experimental implementation that has been created as part of this research. It is a substantial development effort that comprises the core library with numerous plugin points for user or community defined extensions. Most of the development effort was invested in building the core platform with a well-defined rich set of extension points and then creating a small set of example extensions to demonstrate flexibility and support evaluations and experiments. The core and these various extensions underpin the research reported in the remainder of this thesis. Many more extensions could be made in the future by anyone in the community, to provide increasingly rich sets of capabilities. The currently released open-source implementation represents a Minimum Viable Product (MVP) to demonstrate the potential and enable the experimental aspects of this thesis to be recreated as needed.

Cogni-sketch was released as publicly available open-source software in March 2022, licensed with a permissive MIT license to facilitate broad reuse and extension[20]. Cogni-sketch is an experimental instantiation of HAKF but with a stable set of core capabilities that have been well tested in a variety of example uses[21].

As required by HAKF, Cogni-sketch enables contextually relevant information to be shared between multiple human users and machine agents in the form of a simple but extensible knowledge graph. To ensure flexibility and broad applicability to different tasks Cogni-sketch supports extensions by the community in a variety of forms. The simplest of these is the ability to define task-relevant models and typically includes support for multi-modal domain-relevant information, through the creation or reuse of domain concepts and relationships with well-defined semantics (in the form of ontologies). However, these ontologies are not exposed directly to the user but are instead made available as a set of concepts within one or more *palettes* that appear within the Cogni-sketch environment

[20]The core Cogni-sketch platform is available at `https://github.com/dais-ita/cogni-sketch` and a small subset of the stable plugins are also available, at `https://github.com/dais-ita/cogni-sketch-plugins` with plans to release more plugins in the future.

[21]For a video demonstration of the capabilities of Cogni-sketch applied in a variety of settings please see video V12 in Appendix A.3.

and are aimed at less technical users. These contain *palette items* that can be used by human users or machine agents to represent and link different types of information relevant to the problem-solving task or the wider domain and can be contributed by any agent at any time.

The simple semantics (based on first-order predicate logic [31]) is expressed in the form of inheritance within the concept hierarchy in the palette. This is alongside the definition of named relationships between concepts, and the ability to have unnamed concepts and relationships to support customisation and extension during operations. All of this is optional to ensure a simplified fast start when the need to capture information quickly overrides the ability to create a model or ontology in advance[22].

Cogni-sketch is a platform that offers a middle ground between human users and machine agents, enabling the simple representation of information and knowledge, to support machine inference or human intuition. Machine agents are available in a number of forms but the most common are the directed agents that are typically located as *functions* which can be invoked by human users to fulfil different tasks.

The Cogni-sketch platform is built as a set of web APIs and a browser-based UI that provides straightforward access to the knowledge graph. It is designed to be easily to install and run either locally or on a simple hosted web server. It supports encrypted communication (https) and has simple role-based access with user ID and encrypted password authentication to manage user access.

4.3.1 Revisiting user roles

The concept of different human user and machine agent roles for a HAKF system was introduced in Section 3.3.4. We now return to that topic, but specifically in the context of Cogni-sketch systems and accounting for the features offered by this environment. These user roles are shown in Figure 4.3 where the previously listed

[22] For a video summary of semantic capabilities please see video V3 in Appendix A.3.

secondary roles have been removed for clarity (Decision subject, Data subject and Examiner) - refer to Section 3.3.4 for a brief description of these. The figure has also been expanded in two key areas: the human creator users, and the machine agents. These expansions are listed below along with a recap of the other roles that were defined earlier:

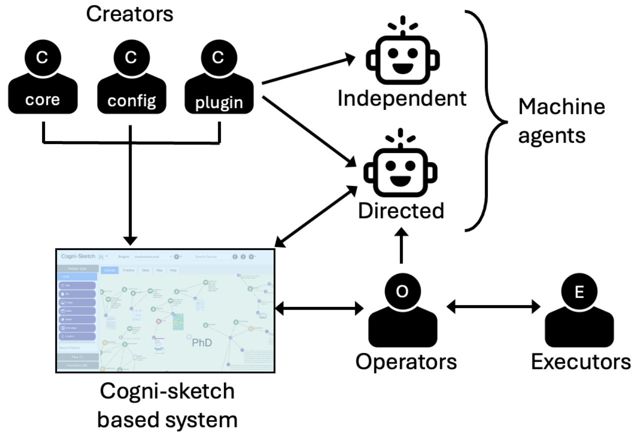


Figure 4.3: Specific Cogni-sketch roles for human users and machine agents.

- **Creators:** Create the HAKF-based system. Several teams of creators may work on different aspects of the same system e.g., architecture, design, implementation, training, documentation, deployment, and maintenance.
  - **Core creator:** This creator user builds the core Cogni-sketch system, defining the extension points for plugins, designing the APIs and implementing the main elements of the solution. They must ensure that the required capabilities of HAKF are respected as well as keeping the core implementation generic and reusable by task-specific applications. The core creator is a design-time role, with changes being made during operation being unlikely.
  - **Plugin creator:** This creator user builds one or more plugins within the constraints of the plugin points offered by the core Cogni-sketch

environment. These plugin points are defined by APIs and SDKs and typically plugins are created to customise the Cogni-sketch environment to more task-specific settings. For example, by introducing support for sensemaking. Currently plugins can be panes, functions (directed machine agents), windows, or custom palette items. The plugin creator is also a design-time role, with changes being made during operation being unlikely, but the tempo for plugin creators may be better aligned with a given operation, allowing plugins to be developed and deployed during the time frame of an operation.

  - **Config creator:** The config creator can modify their palette to define new task-relevant palette items in the form of semantic concepts. They are also able to customise their environment with plugins. This is a run-time role, with the intention being that the config creator user can make their changes in real-time as the operation progresses. Often the config creator user is also the operator user (using the system) but for larger operations the role of config creator may be more centralised with shared palettes being issued to teams of operators.

- **Operators**: Interact directly with the HAKF system, provide the system with inputs (tellability), and directly receive the system's outputs (explainability). In some cases, they may be able to interact directly with the *creators*.

- **Executors**: Make decisions that are informed by the HAKF-based system and receive information from *operators*.

- **Machine agents**: Any machine processes that can interact with the HAKF-based system, working alongside or for the human users.

  - **Directed:** Directed machine agents perform specific focused functions or activities and are invoked in the context of some specific request, typically from a human operator user. General directed functions (e.g., Natural Language Processing (NLP) or Named Entity Recognition (NER) services) are located in the function palette below the main palette and files section and are typically invoked by dropping them onto a node or the canvas itself. Other directed machine agents may be more specific and embedded behind buttons or links within custom panes or windows that have been developed as plugin extensions. These may be used to search data sources, invoke external APIs or perform other well-defined machine processing services.

– **Independent:** Independent machine agents are typically located outside the Cogni-sketch environment as remote agents that can observe and interact with the knowledge graph via APIs. They can consume changes as they are made, as well as contributing their own changes in the form of *proposals* back into the knowledge graph. These are a form of autonomous agent but still constrained within their operations to only consume or create knowledge and information through co-construction in the Cogni-sketch environment. The term independent is used rather than autonomous to capture this constraint, as well as the distinction between directed machine agents since independent machine agents operate independently from human user activities.

4.3.2 User experience

The annotated diagram in Figure 4.4 shows the basic Cogni-sketch UI elements that are made available to the user and are briefly described in the following list.

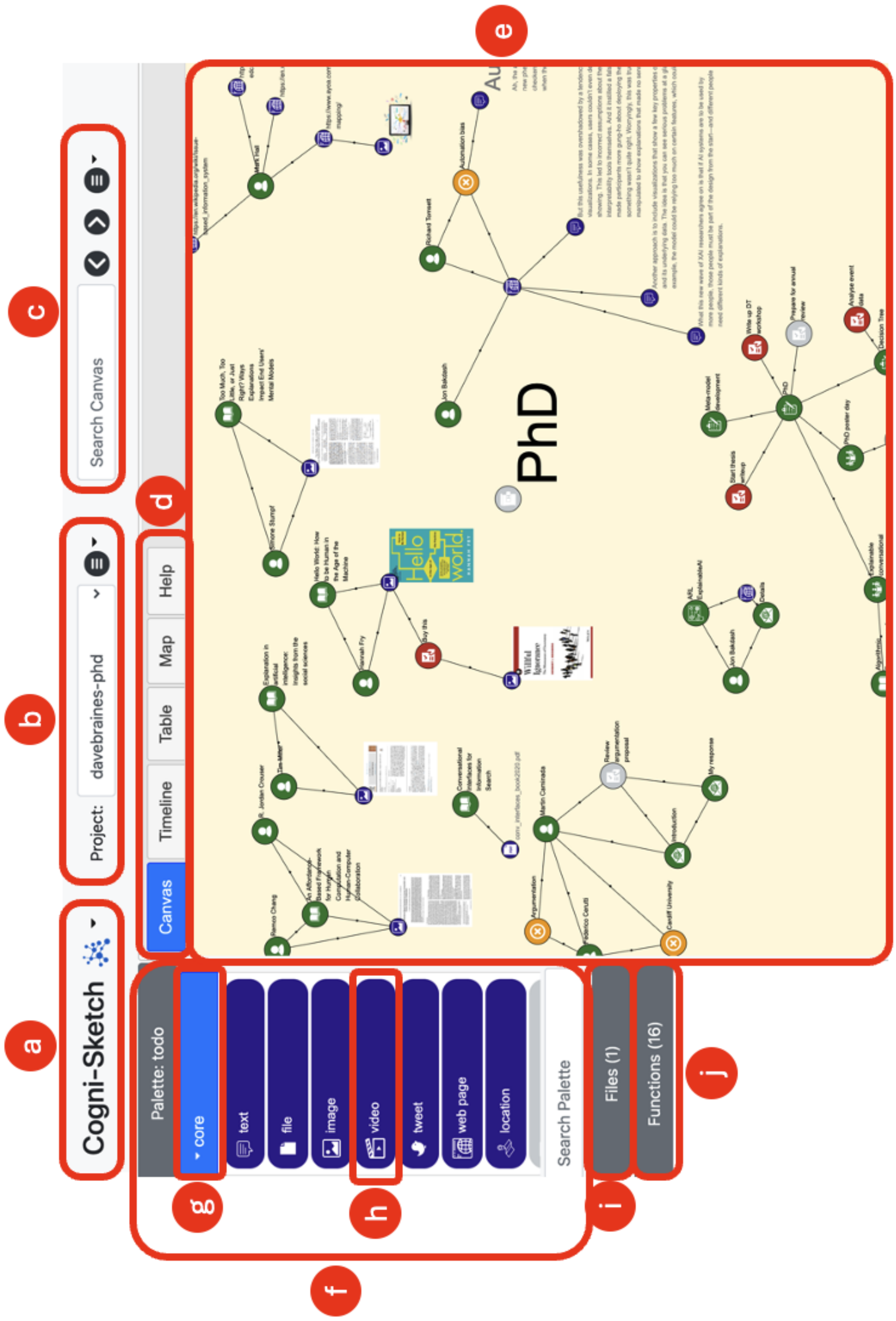


Figure 4.4: Elements of the Cogni-sketch UI

(a) **Cogni-sketch menu**

Shows the version of Cogni-sketch that is running plus enables debug mode to be switched on or off. Also enables the user to change the palette that is being used for the loaded project as well as creating, renaming or deleting palettes.

(b) **Project menu**

The project menu enables the user to change to a different project (by changing the selected project in the drop-down list) or, by expanding the menu, a project can be saved, deleted, renamed or exported. This menu also allows a user to access playback mode to see their knowledge graph redrawn in a time sequence animation, based on the list of actions that human users and machine agents have carried out to get it to the current state (when enabled).

(c) **Search**

This is a simple search function for the current project. The drop-down menu enables case-sensitivity to be specified as well as whether the search looks only at node and link labels or also searches inside the properties of nodes and links and looks at hidden items.

(d) **Pane list**

This is a simple list of the panes available to the user. All panes can show any information in the knowledge graph but often in different styles or formats. There can be any number of these panes depending on what plugins the user has loaded. The default panes are:

- Canvas - described in (e).
- Table - a simple tabular representation of all the nodes and links on the canvas. Useful for certain activities that involve detailed reading, analysis or checking that every node and link has been reviewed (system 2). There are hyperlinks to get from a node or link on the table view to the same node or link on the canvas.

- Timeline - any nodes that have any timestamp property are rendered onto the timeline pane.
- Map - any nodes that have geospatial data (such as a latitude/longitude pair of properties, or a location name within a location node) are rendered onto the map pane.
- Help - describes how to use the Cogni-sketch environment and provides links to various videos and downloadable plugins as well as a summary of the versions for all the loaded plugins (and the core Cogni-sketch version).

(e) **Canvas**

The canvas is the default pane and shows the nodes and links for the specified project. The user can position the nodes within the canvas and the links will be drawn as they are attached to the nodes. The user can create new nodes on the canvas via drag and drop from any external sources such as local files, Uniform Resource Locator (URL)s or fragments of text or images. Also, by using copy and paste or by dragging palette items from the palette to create new nodes. They can zoom and pan and select or deselect nodes individually or by drawing a selection rectangle, and they can choose the types and visual attributes of the nodes as they see fit. It may be important to human users to layout their graph on the canvas in a visually meaningful way to assist with their understanding or thinking/reasoning. The nodes on the canvas can be cycled through three different visual modes, with the ability to add new modes when needed:

- Blank: where only the node and optional label is shown.
- Normal: where whatever default rendering is used (e.g., text is shown for text nodes, image is drawn for image nodes, video player is drawn for video nodes etc).

- Table: a simple table of name/value pairs for all properties defined on the node.

(f) **Palette**

The palette is comprised of collapsible sections and palette items within those sections. The items can be used to create nodes on the graph and those nodes will be rendered according to the visualisation specified for the corresponding palette item. Users can extend the palette by adding new palette items, editing or deleting items and/or sections. They can choose the colour and icon for palette items, and palettes can be exported and imported as needed. For any project the user can switch between palettes by changing the palette in the drop-down list, with the nodes on the canvas being rendered according to whatever palette is chosen.

(g) **Palette section** - A simple collapsible section that contains palette items.

(h) **Palette item** - A palette item is a node type with associated simple semantics (such as parent/child palette item and any properties or named links that are defined for that palette item).

(i) **Files** - If the user has pasted or dragged any files onto the canvas these will be shown as a file node on the canvas but also listed separately as files in this collapsible file list. These files can be sourced from local folders or from network sources and a copy is taken and stored directly within the project for all files that are saved. This is especially useful for online sources where both the original URL and a link to the copy of the file is retained.

(j) **Functions** - This is where any user-invocable directed machine agents are located. They can be defined within sections if needed and can be triggered in different ways. Commonly they can be dragged and dropped onto either the canvas or directly onto nodes or links. The drop event onto the canvas may be different to that of a drop event onto a specific node and handled differently by the agent. For example, a 'language translation' agent within the function list may translate text from English to French and if dropped on a node will create a new linked node with the French translation and provenance data about who invoked the agent, and when it was done as well as the translation text. If dropped onto the canvas it may generate the entire graph of all nodes and links and create a summary report with the French translation. Agents can access any data in the knowledge graph and the exact actions of an agent within the functions list are defined by the developer of the agent (a plugin creator role). Agents are installed via plugins and can carry out any processing both locally with code or by invoking remote APIs. Examples of these directed machine agents that are invoked as needed by the human users include: Image analytics (e.g., reverse image search, entity detection), text translation/summarisation/conversion, entity identification, search, integration with other platforms, conversion to other formats etc.

4.3.3 Solution features

To fulfil the UI elements described above there are a number of technical capabilities that are provided by Cogni-sketch, and these are referred to as *solution features*. These relate to the HAKF required capabilities defined earlier in Section 3.4 with each main mapping called out for each solution feature listed below[23]:

- **Knowledge graph co-construction**

  A project within the Cogni-sketch environment is a knowledge graph com-

[23] The 'Support for sensemaking' required capability (and corresponding RQ3) is not mapped to any of these as they are solution features for the core Cogni-sketch platform whereas sensemaking is a particular use case with explicit support provided by plugins for each such use case.

posed of nodes and links. The nodes have types which are defined within the corresponding palette in the form of palette items, and these types define the semantics for the nodes. Links are between two nodes and may be named or unnamed, and if they correspond to names defined in the palette, they will also have semantic meaning. Specifically, the knowledge graph is implemented as a Directed Cyclic Graph and is stored in JavaScript Object Notation (JSON) with references between the nodes and links via unique ids. The minimum schema for the knowledge graph is deliberately simple but flexible and easily extended. As the knowledge graph expands it may become hard to see all the data. Different areas of the canvas can be used for different parts of the graph, or knowledge graphs can be split into different projects, each containing a sub-graph. Alternatively, the user can select any number of nodes and links and choose to hide them. They disappear from the canvas but remain available within the knowledge graph. Machine agents are still able to access these nodes and links, but a visibility flag is set to false, so they know they are hidden.
Related required capabilities and RQs: *Rich knowledge representation*, *RQ1*, *RQ2*.

- **Human-friendly semantic models**
  The palette is extensible and shareable. It contains a small set of predefined core palette items that broadly correspond to the common media data types that are always supported within the Cogni-sketch platform (e.g., text, image, video, webpage etc). The palette can be extended with any number of additional palette items, and these can either be new fundamental data types provided by plugins (e.g., tweet or geolocation), or they can be domain-relevant node types for the activity that Cogni-sketch is being used for (e.g., person, vehicle, document etc). Palettes can be exported and shared, and a single palette can be used for any number of Cogni-sketch projects. Palette items can be arranged into sections for convenience and

a semantically meaningful parent/child hierarchy of palette items can be specified. The colour and icon type for palette items can be defined along with the default formatting for each palette item when it is rendered on the canvas. Palette items can also define any number of named predefined properties that will automatically be created for any new nodes of that palette item type that are added to the knowledge graph.
Related required capabilities and RQs: *Rich knowledge representation, Visualisation and interaction*, *RQ1*.

- **Support for meta-data**
  Nodes and links can be created on the canvas by any agent (human or machine). By default, they have properties created for whatever palette item they are deemed to be. For example, a 'text' node gets a 'text' property that corresponds to a rich text field by default. In addition to any default properties that are automatically created, any agent can also create any number of specific properties for a node or link and specify their contents. The values of properties for nodes and links can be seen by the human users when they double click on the node or link to see the contents, are listed on the table pane, or they can be used when rendering the node on the canvas. Properties are typically used to store data or meta-data that relates to any node or link within the knowledge graph.
  Related required capabilities and RQs: *Rich knowledge representation, Machine agent integration*, *RQ2*.

- **Change history**
  Every action that is taken within the Cogni-sketch environment is optionally recorded as a specific event in a log. This can be disabled via the configuration file if needed. These event logs are primarily used for audit tracking of changes to the environment, but they also serve multiple additional purposes and are available for future enhancements also. The first usage is

a simple undo/redo capability which allows the user to quickly undo their changes action-by-action when needed. The second is the support of a playback function within the Cogni-sketch environment which allows the user to see their knowledge graph recreated dynamically based on the content of this change history. This playback function can be useful when explaining an area of the graph to other users or when revisiting progress so far. The event history is available via the APIs so it can be accessed by machine agents or used in any higher-level interactions with the human users, such as a conversation about the content of the knowledge graph. Event analysis of user behaviour is also supported through change history events.
Related required capabilities and RQs: *Agile information capture*, *RQ1*, *RQ2*.

- **Support for collaboration**
  Each node and link that is created on the canvas is attributed with the user that created (or modified/deleted) it, as are the events in the change history log. This provides an open and flexible basis for supporting many types of collaboration, however there are important usability considerations to be taken into account, so there are only two specific collaboration modes between human users that are supported currently in Cogni-sketch:

  - The granting of read-only access to a project so that any number of other users may remotely observe the creation and development of a knowledge graph in real-time as it is developed by the owner.
  - The generation of *proposals* which are sent from one user to another. These take the form of a sub-graph of new nodes and links that are entirely stand-alone, or which relate to nodes/links already in the receiving knowledge graph. Proposals can also include edits to existing nodes or links as well as added/edited/deleted properties. These proposals can be accepted or rejected by the receiving user.

Free-for-all open collaboration where any user can make any modification to any node, link or property of any project is deliberately not supported at this stage. This is simple to implement but is not expected to be a useful mode of collaboration for the human users as the scope of change and lack of control by the owning user may be unhelpful, especially in an intelligence analysis or sensemaking context where confidence in the content of the graph will be of high importance to the owner.

In addition to these two explicitly supported collaboration modes any user can easily share any palette or project file with any other user via import and export, or they can copy/paste collections of nodes and links between knowledge graphs. Many more collaboration modes can be added using the open and extensible APIs, but they are not the focus of the research at this stage.
Related required capabilities and RQs: *Rich knowledge representation, Visualisation and interaction*, *RQ1*, *RQ2*.

- **Multi-user**
  The Cogni-sketch server can host any number of users concurrently, assuming the hardware and network are scaled to support the necessary load. The users are managed via a single encrypted definition file but integration with external identity providers is also supported. The bulk of the processing required for Cogni-sketch is carried out in the local knowledge graph for example, via the canvas pane, and therefore within the browser in the users own environment, with the server only being responsible for saving knowledge graph updates, or palette changes and storing any files or images saved by the users, all of which are low cost operations.
  Related required capabilities and RQs: *Visualisation and interaction*, *RQ1*.

- **Multi-project**
  Any user can create any number of projects and quickly switch between

them via the project drop-down list. Projects can be shared with other users and exported as needed. The core project data is held in a simple JSON file that contains the nodes and links. This can be easily integrated into a document database or similar but for ease of consumption and installation a simple file-based solution is used for this version, and it supports all the examples needed so far with minimal development or support effort. Any additional files are stored within the project folder on the server as well as being represented as a file node in the graph. Projects are easily integrated into GitHub for version tracking and if exported they can be easily imported by other users into different Cogni-sketch environments.
Related required capabilities and RQs: *Rich knowledge representation*, *RQ1*.

- **Machine agent integration**
  Machine agents are easily integrated into the environment. There are two main techniques:

  - Directed machine agents, defined as plugins directly within the environment.
  - Independent machine agents, defined as remote agents that observe and interact with the knowledge graph via APIs and can consume changes as they are made, as well as contributing their own changes in the form of *proposals* as described previously.

  Related required capabilities and RQs: *Machine agent integration*, *RQ2*.

- **Extensible visualisations**
  Any node on the canvas can have a custom visualisation defined. For example, an 'image' node will show the image data as a picture on the canvas, and this can be sourced locally from image data on the node directly (e.g., because it was pasted from the clipboard) or retrieved from a remote source such as a URL. Any custom palette items that are created by users can

have their own visualisations defined. This could be by taking property values on the node and rendering them in a specific format such as a table or using the value in a property to inform a dynamic contextual rendering of the node. This is fully configurable by the users at any point during the construction of their knowledge graph or editing of their palette.
Related required capabilities and RQs: *Visualisation and interaction*, *RQ1*.

- **Hidden data storage**
  In some cases the user may wish to store data but not as nodes or links in the knowledge graph. There are three places that additional data (in the form of name-value pairs) can be stored within the environment, and these are currently available only via APIs rather than the browser-based UI.

  - The project - name value pairs are stored in the project and available when that project is loaded. They are included in project exports.
  - The palette - name value pairs are stored in the palette and therefore available to agents running in any project that uses that palette. They are included in palette exports.
  - Globally - name value pairs are stored in the Cogni-sketch server and can be accessed by agents in any project for any user on that server. They cannot be exported without writing custom agent code to do so or by the server administrator accessing them on the server.

  This hidden data can easily be used by machine agents when needed without the information being added to the knowledge graph. This can prevent clutter of the knowledge graph with config-related information, but if the information in question is more generally relevant to the task, then it can still be located within the knowledge graph instead, even if it is also used to configure machine agent behaviour.
  Related required capabilities and RQs: *Machine agent integration*, *RQ2*.

4.3.4 Extension points

Cogni-sketch is available as open-source software and any component within it can be replaced as needed in the future, however there are also aspects that have been designed as run-time extension points. This enables the community to create new capabilities as plugins that can be added to the environment to provide additional capabilities or could replace existing default capabilities with better alternatives. The number and variety of these that can be created is only limited by the imagination of the contributing community and the constraints of the plugin points. If needed these plugin points can be expanded, with future modifications to the core (although none are currently planned).

Refer to Appendix A for a list of available plugins, some of which offer palette items, while other offer panes or functions (and some offer combinations).

The currently supported plugin-in extension points are:

- **Palettes**
  Any config creator user can create new palettes or extend existing ones. They can export their palette and share with any number of users. In addition to the creation of simple palette items it is possible for plugin creators to define more advanced ones that require some code to be run as part of the creation or rendering of nodes of that palette item. A good example of this is the tweet palette item which uses a Twitter URL to request the tweet content is created in the exact style of a tweet (by a remote Twitter service) before being rendered on the canvas.

- **Panes**
  Custom panes can be created as plugins by plugin creators and shared within the community. They tend to render the knowledge graph content in a particular style and can include any number of special functions to modify or create nodes on the canvas too. An example of a pane plugin is the story pane that was created to support the sensemaking pilot and

subsequent experiment described in Section 6.3.4.

- **Functions**

  These are the most common form of machine agent integration and are also created by plugin creator users. They represent directed machine agents and are simply code which can be invoked as needed by the operator users of the environment. Typically, they involve some kind of analysis or processing of data within the knowledge graph and often result in new nodes being created, or modifications to the existing nodes, links or properties. An example of a function is language translation: dropping that function onto a node can call a remote web service to translate from one human language to another and put the result as a new node onto the canvas with a link to the original node and any generated provenance information.

- **Windows**

  Like panes but these appear as embedded or popup windows, usually to provide extra contextual information, to guide the user through the creation of more complex information in a more compact form than defining the nodes and links on the graph manually. They can also be created as plugins and shared within the community. Examples of windows can be found in the science library example in Section A.1.1.

The implementation of Cogni-sketch is sufficient to serve as an experimental basis for the evaluation of capabilities reported in the remainder of this thesis. There are a small number of areas that have been implemented with minimal functionality, but which could be further enhanced in future iterations as required. The main examples of this are reported in Section 7.2.1.

4.4 Chapter Summary

Cogni-sketch has been created after extensive investigation into existing tools with capabilities that fulfil similar needs found that no single existing environment can provide all elements to fulfil the HAKF required capabilities. The Cogni-sketch platform offers a common ground between human users and machine agents, enabling simple representation of information and knowledge, defined by human user activities or machine agent processing. Using the definition of simple semantics — concepts, inheritance, relations and inference rules — through an extensible palette, both human users and machine agents can contribute knowledge to the canvas using these palette items which should be more amenable to typical domain expert operator users who don't have deep experience with ontologies or schemas. This knowledge is contributed through nodes, links and properties in the knowledge graph. Directed machine agents are available as functions which can be invoked by human operator users to fulfil different tasks, and the plugin nature of the environment means that any number of new machine agents can be created by plugin creator users, or existing agents can be evolved into more complex forms or tailored to specific processes.

Chapter 5

Integrating explainable machine agents

5.1 Introduction

As defined in Chapter 3, HAKF has two main flows that can occur between human users and/or machine agents working together in a collaborative setting. This chapter is mainly focused on the explainability flow, and specifically how machine agents contribute task-relevant information into the knowledge graph through co-construction, some of which is in the form and style of explanations. This content from the machine agents can be both explanation-related nodes and links as well as first-hand machine generated content such as imagery, classifications or other information. The tellability flow is also relevant, and identified where appropriate, typically with human users providing task-relevant information to configure the behaviour of machine agents. This chapter is split into three sections, each relating to examples of human-machine collaboration using the Cogni-sketch environment that have been previously published.

This chapter first describes a pilot exercise in Section 5.2 in which a series of related but independent machine agents were integrated into a common scenario using the Cogni-sketch environment as the knowledge graph co-construction basis for consuming and creating task-relevant information. Each of the services that contribute to the scenario were active research components from collaborators, so the pilot exercise was to integrate them into the Cogni-sketch environment. The goal was primarily to exercise the Cogni-sketch APIs, test the ability to rapidly create customised palettes for the required semantic expressivity, and enable the wrappers for each of these separate machine agents to consume and create knowledge, according to the palette items, using the Cogni-sketch knowledge graph. A high-level description of the machine agents, human users and motivating scenario is given, with the full details reported in [21].

Next, in Section 5.3, is a summary of a more formal evaluation of these machine agent co-construction and explanation capabilities, specifically reporting on the rapid implementation of a subset of capabilities which run live in the environment generating and responding to locally detected events. The evaluation was performed by the author of this thesis who was also the developer of the Cogni-sketch environment and built the required plugins for the evaluation (i.e., fulfilled the roles of core creator, plugin creator and operator). This evaluation does not include any formal assessment of behaviour or results, or feedback from participating users. Instead, the work was built in conjunction with feedback from the information fusion community and published alongside related work also relevant to that community [27].

The methods for the pilot and the later evaluation are described (in Sections 5.2.2 and 5.3.2 respectively). Both were post-hoc evaluations of the ability for machine agents to be embedded into the Cogni-sketch platform. There was no direct or formal analysis of data relating to the exercises, or any subjective assessment of the capability, and the exercises were undertaken by the author of this thesis. The findings from the pilot led to the improvement of the Cogni-sketch platform to better support such integrations, and this was then tested in a real-time setting as reported for the subsequent evaluation (in Section 5.2). The findings from the pilot and evaluation are summarised and compared in Sections 5.2.3 and 5.3.3 respectively.

Finally, in Section 5.4, there is a summary of relevant material related to conversational forms for explanations from machine agents providing explanations based on multi-modal sensor analytics [26, 29]. This research predated the development of the Cogni-sketch environment but informed the required capabilities for HAKF alongside other activities such as the DT workshop with military, government and industry stakeholders as reported in Section 3.2.

These pilot, evaluation and conversational investigation activities all contribute to the definition and subsequent exploration of required capabilities for machine agent integration within HAKF. Both in terms of how agents like these can operate within such an environment once defined, as well as confirming the Cogni-sketch API and SDK mechanisms needed to enable agile integration and evolution of such machine agents. The plugin creator and operator users are the most relevant roles for these exercises, with a focus on co-construction and consumption of explanation-related material, and interaction with the system in task-relevant settings at an operational tempo relevant to the problem-solving task.

5.2 Pilot: Explanations through co-construction

For the pilot exercise described in this section Cogni-sketch is used to define various independent machine agents and have their outputs dynamically generated as task-relevant information into the environment, with contributions from human users when needed. The material generated by these machine agents is rendered in the usual interactive mind map knowledge representation format, with different nodes and links used for different defined purposes based on the palette.

Our scenario features a Situation Understanding (SU) example based on the collaboration between human operator users and independent machine agents. The purpose of the Cogni-sketch environment in this setting is a virtual workspace where these operator users and machine agents can rapidly co-construct knowledge through the two key HAKF flows:

- Explainability: providing explanation relevant information and associated certainty information and going deeper into rationales for why information exists or has been added.

- Tellability: injecting new knowledge and information (e.g., task-relevant rules and facts).

5.2.1 Pilot objectives

There are two specific questions that drove the scope and approach for this pilot exercise, arising directly from the broader *RQ2* (machine assistance) for this research thesis, introduced in Section 1.3. In addition to this primary focus there were also contributions from human users to direct machine agent behaviour, and the overall goal of seeking and maintaining SU, but these are secondary to the main focus on machine agents and their ability to provide explanation-related content.

Q1 Can existing services be rapidly integrated as machine agents into the environment using the plugin mechanism at an operational tempo relevant to the task?

Q2 Can machine agents be configured or re-tasked by human users contributing relevant information into the environment?

These two questions framed the objective for this pilot alongside the development of relevant services to support this effort. The pilot therefore aims to demonstrate the ability for multiple machine agents to be rapidly integrated into the environment based on a set of existing services that were under active construction by colleagues within the DAIS ITA research programme. In addition to exercising the core capabilities of HAKF as defined within Cogni-sketch, and demonstrating the role of machine agents and human users the following relevant factors are investigated:

- *Operational tempo*: the pilot scenario mimics a real-time operation, with the need for operator users to establish and main SU in a rapidly evolving situation. In addition, the pilot directly tests whether it is possible for plugin creator users to create independent machine agents in a similarly rapid time frame, to enable the integration of new sensors and services as plugins to the environment.

- *Multi-modal*: the sensors and associated services deal with a variety of modalities and expose these as created items in the knowledge graph as needed. Fragments of the overall situation picture can be found across these modalities, just like it would be in a real operational setting.

- *Trust*: whilst not formally assessed or validated, each of the services declares a certainty value for the results that are generated, and this plus the nationality of each service is a mechanism by which additional contextual information is communicated and could be used as the basis for trust forming between human users and machine agents [150].

- *Explanation*: each of the examples within this pilot generated an explanation node, with the relevant sensed data and associated explanation linked to this. Depending on the modality the explanation format will differ, with a range of examples given.

The pilot aims to prove that the HAKF approach and the Cogni-sketch implementation can achieve integration of these services, with machine agents creating (and consuming) task-relevant information via the knowledge graph. The machine agents must be created within the constraints of the existing plugin capability, with all types of information being communicated according to items defined in the palette.

In this pilot exercise the component services were created separately to demonstrate different contributions to the situation, but a flexible and agile environment was needed in which they could be loosely integrated, with opportunities for contribution of relevant information from human users. The Cogni-sketch system had already been designed for independent machine agents such as these, with appropriate APIs to allow consumption and creation of information within the knowledge graph, but the ability to do so with these specific services had never been attempted.

Human operator users can use the tellability flow to inject new task-relevant knowledge, or hypotheses, about patterns of activity rapidly through co-construction, e.g., by addition of new rules that can be consumed by generic machine agents operating on this task. This means patterns can be recognised in situations where there is insufficient time or data to train a new ML model for that specific context.

5.2.2 Pilot method

In order to demonstrate the integration of multiple machine agents, in the context of monitoring a rapidly evolving situation in an urban setting, we define a set of four services and use the fictional NATO *Anglova* urban setting [145]. These machine agents operate independently and are typically processing sensor data of different modalities, to enable active management of SU. We envision a situation where events indicate growing threats to, and attacks on, the fictional *Capulet* community which we created for the purposes of this exercise.

This pilot represents an example of dynamic SU based on pseudo-real-time[1] information coming from multiple machine agents processing sensed or inferred data in an area of interest. In this pilot most of the sensed data was coming from the machine agents, with the human operator users tasked with managing the overall SU picture through their sensemaking of the incoming data, and their occasional configuration of machine agents through definition of task-relevant

[1]By this we mean that the information was created in advance to support the testing of each of the component services but is presented into the Cogni-sketch environment in a real-time setting, with each of the services being triggered with the predefined data as if it were live.

additional information.

Each of these four services were under active research and development by research collaborators at the time of the pilot. They are not defined in detail in this thesis as the focus here is how they were integrated via independent machine agents, and the kinds of task-relevant information and explanations they contributed. Full details of the implementations and other relevant factors can be found in [21].

Figure 5.1 shows an overview of the Cogni-sketch environment showing various outputs from independent machine agents. Explanations are provided as additional nodes and links, and in some cases additional explanations are provided within the media assets, e.g., through overlay saliency map highlighting.

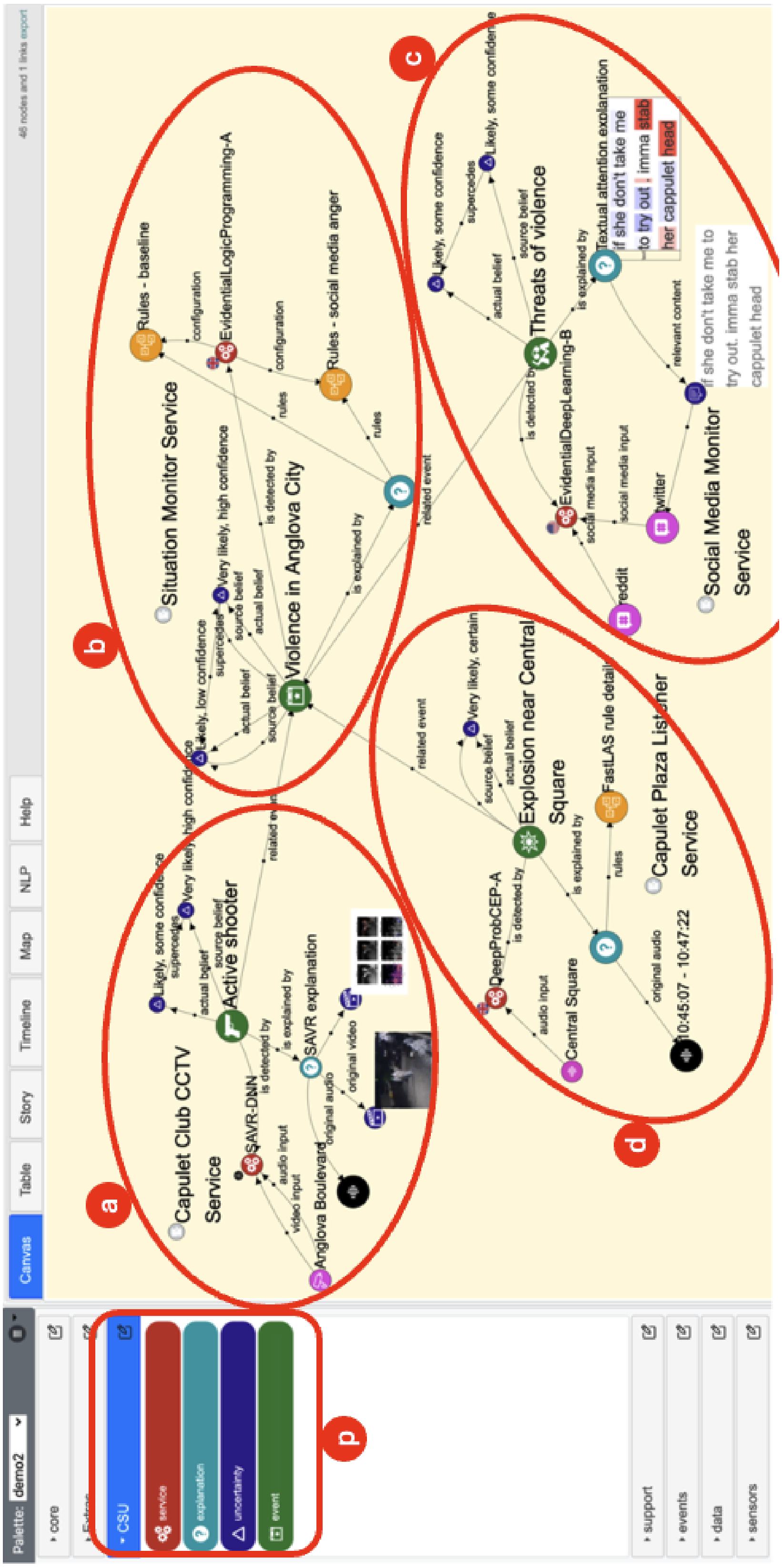


Figure 5.1: Detected events and corresponding explanations

The four Cogni-sketch independent machine agents representing these services and the portion of the knowledge graph that they contribute is highlighted in Figure 5.1 with each service identified from a to d and described below:

a. **Multi-modal (audio-visual) event detection**: Based on a simulated Closed-Circuit Television (CCTV) feed with associated audio, running an edge-based event detection and explanation service.

b. **Rules-based situation monitoring**: Based on evidential logic programming and using logical inference rules to infer larger scale events from individual reports.

c. **Social media sentiment analysis**: Natural Language Processing (NLP) and sentiment analysis on social media, trained to classify/detect threatening language.

d. **Audio event detection**: A static audio sensor with an onboard (edge-based) event definition algorithm, able to detect defined event types.

There are three links on the graph that span the different service boundaries, and each of these are a *related event* link that is created between the individual detected event and the compound inferred event of *Violence in Anglova city*.

The palette on the left-hand side of Figure 5.1 (p) provides a set of capabilities and services that can be used, customized or extended, and shared as needed. For example, the figure shows the specific palette items designed to support SU for any set of machine agents. These are in addition to the standard palette items and other specialised palette items that support this pilot exercise which are located in the other palette sections that are collapsed and hidden in Figure 5.1. The palette items shown are:

- **Service** (): The individual machine agent instances that have been deployed into the environment. Typically, these services are running in a distributed setting on the remote sensors themselves, but they could also be run remotely in cloud infrastructure.

- **Explanation** ( ): Additional information from the machine agent which provides some form of explanation for any events that are detected.
- **Uncertainty** ( ): All observations and inferences from these machine agents come with uncertainty information which is essential for downstream machine agent inferences or human assessment.
- **Event** ( ): The events that can be detected by machine agents in this environment. These are specialized into specific types, based on the capabilities of the services.

This simplified palette supports modifications and specialisations as needed by the config creator user who, in this scenario, is likely to also be the operator user, seeking to gain or maintain SU assisted by these machine agents. The other palette sections are collapsed but contain other relevant items such as the events that can be detected, the sensor and service instances and other related items.

Four independent machine agents were able to be successfully integrated into the Cogni-sketch environment through fairly simple wrapping techniques to enable them to be invoked when needed based on the contents of the knowledge graph. For the purposes of the pilot the first agent was triggered independently by running a script to start the process.

Each machine agent was able to generate task-relevant nodes containing raw sensor data in various modalities, with different forms of explanations, certainty information and links between nodes to show the relationships. The machine agents acted independently and used the simple blackboard architecture implemented in Cogni-sketch to listen for relevant events, and then contribute new knowledge and information via the *proposals* mechanism for externally contributed knowledge graph fragments. For this exercise these proposals were set to be auto accepted since it enabled the chain of services to be triggered automatically based on the creation of relevant content. In a real setting it may be that the human operator user wishes to review the proposals coming in and decide which to accept. They can also exercise the usual human creative considerations of layout and location (within project and on canvas) as needed. This may be especially useful to support the operator in explaining a more complex situation to a decision-maker, rather than for each individual event as it occurs. The nodes created by the machine agents in this example were simplistically laid out based on baseline example layouts defined by the plugin creator user for each machine agent. In a real system with a wider variety of outputs careful consideration would be needed for a more dynamic layout solution.

The services that were successfully integrated as independent machine agents are now described, with a brief discussion of the results for each case[2].

Multi-modal event detection

Figure 5.2 shows the result on the Cogni-sketch canvas of the multi-modal event detector service detecting an active shooter situation. This corresponds to section (a) in the overview shown in Figure 5.1.

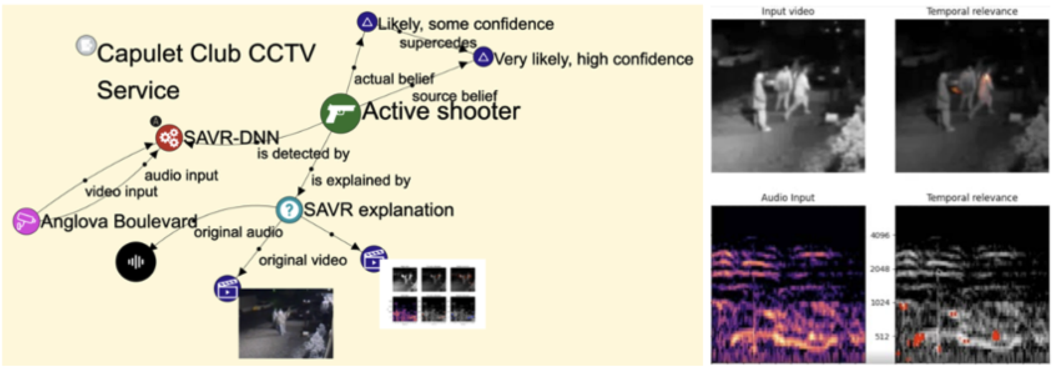


Figure 5.2: Explainable multi-modal event detection

The explanation for the event contains links to the relevant raw input from the

[2]For a video demonstration of this pilot exercise please see video V13 in Appendix A.3.

sensor, which in this case is the CCTV video and audio feed (shown bottom-left in the figure as a directly embedded playable video on the canvas, and in enlarged form on the right). This can be seen on the canvas as an attention-highlight video and associated audio feed, with saliency map highlights showing the audio and visual relevance of the scene.

This was implemented as an external service by colleagues, and employs selective relevance [146], a post-processing step that can be applied to generate an explanation such as the example shown. Within the context of this example, selective relevance is seen to highlight the shooter's arm (upper right quadrant of the far top-right image in Figure 5.2). Similarly, temporal elements of the audio track are highlighted on the spectrogram in the lower right image (corresponding to gunshot sounds). The details of the implementation will be unknown to the human operator, but the video and audio highlight provide an explanation of what in the raw feed, rightly or wrongly, led to the active shooter classification. Human user agreement with the saliency highlighting can lead to improved trust and confidence in the machine agent over time (or vice-versa).

The active shooter event has been assessed to be *Likely, some confidence* based on the underlying numeric details for the uncertainty that are recorded as properties on the related uncertainty node and can be easily accessed by both human users and machine agents as shown in Figure 5.3. This is a natural language representation of a subjective logic opinion [69] that distinguishes the amount of belief, disbelief, and epistemic uncertainty in the truth of a given proposition.

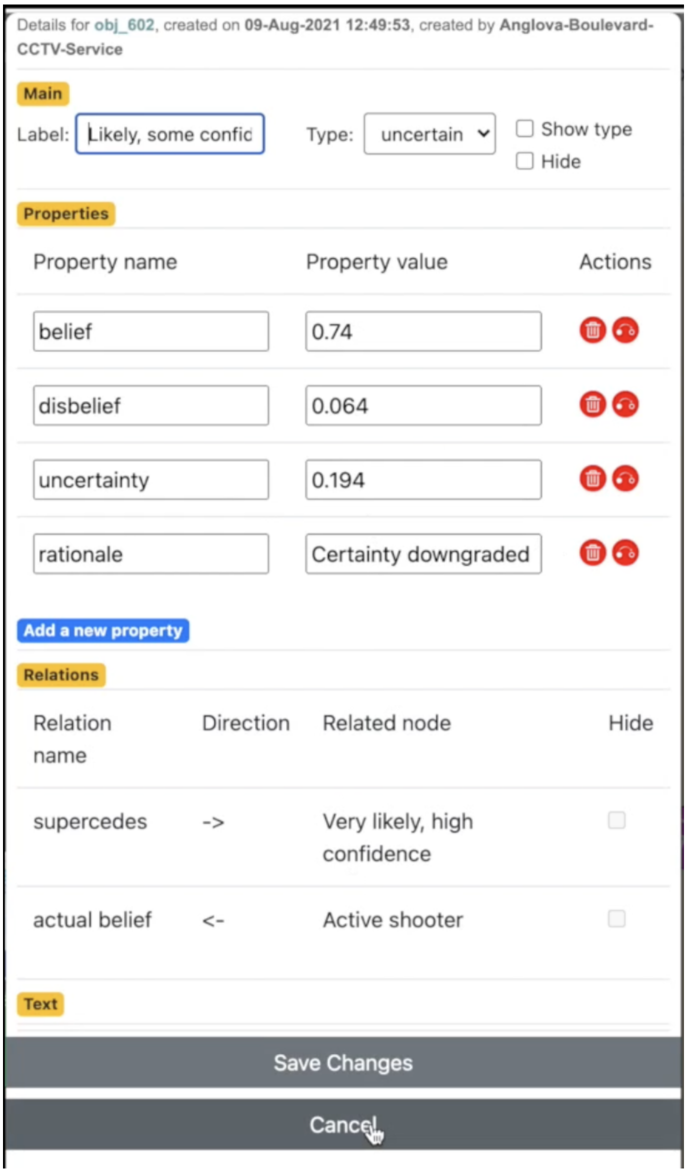


Figure 5.3: Property details for an uncertainty node

Each event comes with a *source uncertainty* assessment that is created by the machine agent that generated the information but can be modified upon ingestion into Cogni-sketch platform based on user defined rules. The source uncertainty, here, has been discounted to 80% because the sensor and machine agent are run by the local Anglovan authorities whereas the user maintaining SU and running the Cogni-sketch canvas is from the U.K. This uniform discounting to 80% is represented as the difference between the *source belief* and the *actual belief* links to the active shooter node on the canvas and is a generically applied discounting rule based on the nationality difference. This serves a two-fold explanation: the original confidence from the machine agent (*source belief*), and the modified confidence after ingestion into the environment (*actual belief*), providing a clear breakdown of the two. Additionally, a property on the actual belief node captures the textual *rationale*[3] for the reduction in confidence, and a new link is created to note that the *actual belief supersedes* the *source belief*. This enables any subsequent machine agent processing to consume the correct uncertainty information and process the associated provenance as well as supporting human user reading of the graph. Miller [95] notes that "*Probabilities (probably) don't matter*" but the information is available in the knowledge graph for use when needed and may be downgraded to secondary information in any higher-level UI that may be built to operate on top of the knowledge graph.

This information is all available in the knowledge graph, but nodes can be hidden (or deleted) as needed to help the human users focus on relevant information, as well as being able to navigate to any relevant information when needed. It is also possible to set up different projects to contain different parts of the situation if needed, enabling easy switching between different views.

Situation monitoring enhanced through tellability

The scenario now progresses to a second machine agent service. The analyst has seen the shooting event and has formed a hypothesis of escalating violence. There is no time or data to re-train a service to learn this hypothesis directly, but the rules can be readily modified by the human operator user who has seen this for themselves. They therefore decide to exercise the tellability flow to create new task-relevant knowledge into the environment, specifically to define a rule to support this new hypothesis. This is then able to be consumed by the rules-based

[3]As shown in Figure 5.3. The text in the figure is truncated and is: 'Certainty downgraded because partner trusted at 80%'.

situation monitoring service in real-time. This corresponds to section (b) in the overview shown in Figure 5.1.

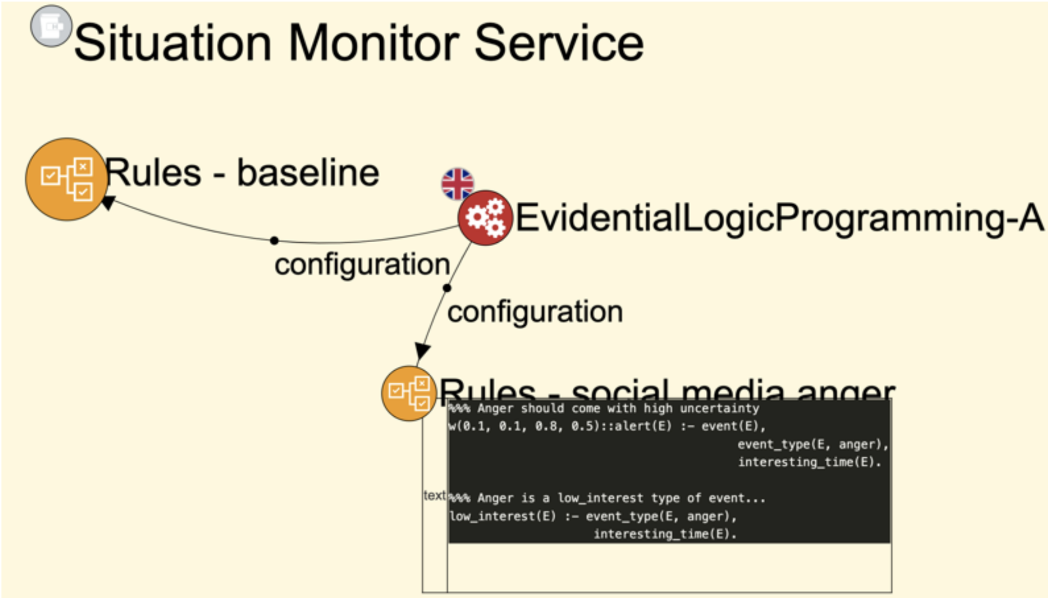


Figure 5.4: Modifying agent behaviour with a new situation-relevant rule.

This is a technical update (containing formal rules or source code) and would likely be made by an operator who has responsibility for maintaining the services and therefore understands how to configure them. The HAKF approach can also support the input of higher-level task-relevant information from which more technical code such as this can be generated, but that extra capability was not created for this pilot.

After reviewing the current configuration, the operator user decides to extend the rules and link this into the configuration, as shown in Figure 5.4. This could either be done by extending the existing configuration rule, or by creating a new node with the additional rule, as shown here. The rule is now live and has updated the running service accordingly, showing the tellability flow enabling dynamic reconfiguration of an existing service. For simplicity the full python code is not shown, and the baseline version is hidden. Hiding component information is one of the optional forms of display for any node (as described in the canvas description in Section 4.3.2) and can help avoid information overload on the canvas.

Social media sentiment analysis

As time passes, the social media sentiment analysis service, processing social media data, signals a credible, *Likely, some confidence*, threat. This corresponds to section (c) in the overview shown in Figure 5.1.

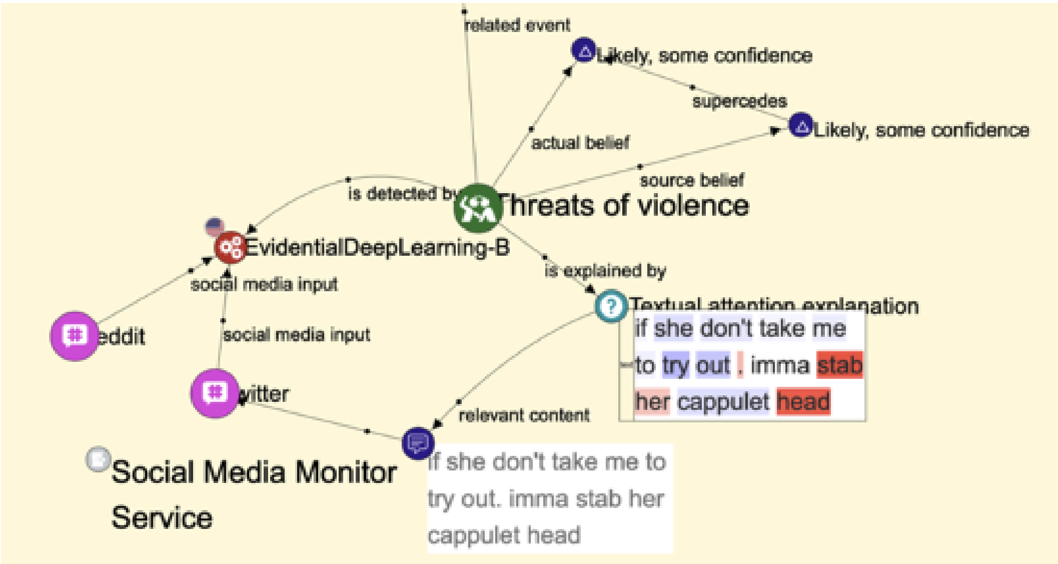


Figure 5.5: Attention-based textual highlight explanation

The triggering social media message processed by this service is shown at the bottom of Figure 5.5, along with an attention-based explanation that highlights (in red) the parts of the message that caused it to be classified as threatening. The uncertainty is *Likely, some confidence* and because the services is run by a trusted partner there is no difference between the source and actual belief, but both are generated as nodes onto the canvas for clarity, and the rationale property explicitly states this fact.

Due to the earlier rule change, the situation monitoring service (b) is in turn triggered by the detection of this new *threats of violence* event, creating a new situation named *Violence in Anglova City*, based on both the new event and the previous active shooter event, using *related event* links between the relevant green event nodes in Figure 5.5). Initially the corresponding uncertainty is *Likely, with low confidence* with the rationale noting that this is computed from related events. In this case the source and actual belief links are to the same uncertainty node because the same partner runs this machine agent and is running the Cognisketch canvas, so no modification to certainty is needed. The rules continue to run, and the situation will be updated as any new related events are detected, with the certainty being revised accordingly.

Audio event detection

Finally, an explosion event is detected by the audio event detection service. This corresponds to section (*d*) in the overview shown in Figure 5.1.

As with all detected events, a link to the explanation is shown along with the original audio, should the analyst wish to assess the audio evidence for the event themselves. This is shown in Figure 5.6. This probabilistic AI service [161] is very confident, generating a *Very likely, certain* classification for this event, and since this listener service is run by the same nationality there is no modification to this certainty just like the previous example. The new explosion event is also linked to the unfolding situation, due to the spatio-temporal proximity, with the certainty for the *Violence in Anglova City* situation updated accordingly (not shown in Figure 5.6), with increased confidence as more related events are added.

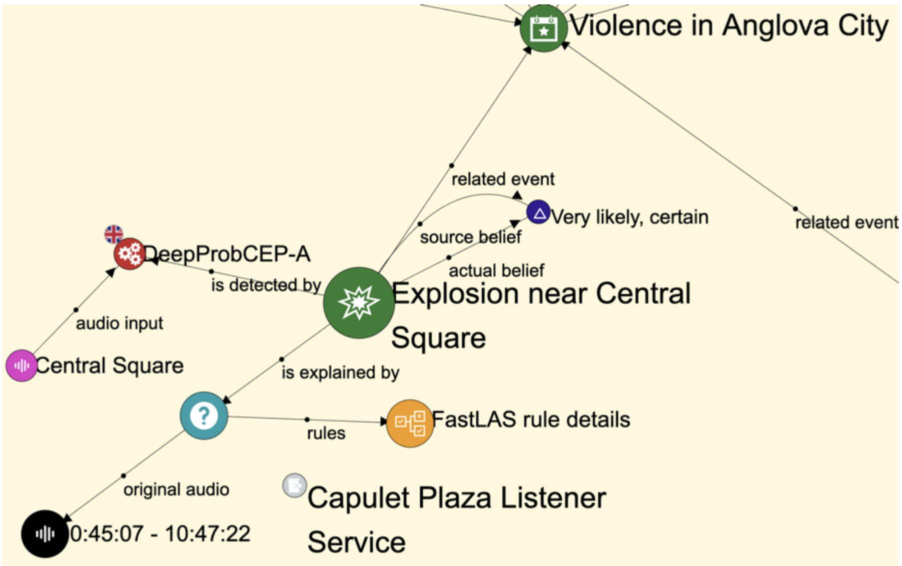


Figure 5.6: Additional event contributes evidence

5.2.3 Pilot findings and discussion

There was no analysis of any quantitative results undertaken during this pilot exercise since no such data was generated. Instead a post-hoc review was carried out of the agents that were successfully created and integrated, and the material that those agents were able to contribute into the knowledge graph. The ability to achieve an integration of disparate agents contributing information into a shared situation was proven, with some extensions to the framework implemented during the pilot exercise.

As a result, this pilot exercise enabled the integration of multiple existing services into a unified scenario to demonstrate the ability for machine agents to contribute task-relevant information and explanations into the environment, as well as a human operator user providing configuration information for one of the machine agents based on their understanding of the unfolding situation.

The pilot scenario is manufactured to specifically enable each of these threads of separate research to be integrated into a single demonstration, however it was a useful pilot test case for the Cogni-sketch environment and the ability to rapidly integrate services like these as machine agents.

The pilot scenario was designed to investigate the two questions in Section 5.2.1 and found the following:

- Machine agents can generate task-relevant information and related explanations into the knowledge graph, and can be re-purposed more rapidly. The machine agents must be designed with this in mind, but if they are they can use the Cogni-sketch environment to source relevant data either specifically contributed for that purpose or for other purposes.
- Humans can use tellability to inject new knowledge about patterns of activity they observe. In this example through addition of new rules in real-time.

The plugin creator user was able to successfully wrap each of these services into a machine agent and have each of the machine agents (except (a) which started the flow) watch the knowledge graph for relevant information and be triggered when it was created. This was achieved using the APIs to Cogni-sketch to access the relevant project information in real-time. The plugin creator was also able to dynamically create new nodes and links onto the canvas from each of the machine agents, using the SDK and the proposals mechanism for such contributions. The implementation for all these machine agents was rapid, taking only a few days, but the agents were simplistic and designed only to work on a predefined set of test data.

Three specific findings arose from this pilot exercise which substantially validated the understanding of how to achieve integration with machine agents in a manner that could be useful for human users. All were known before the pilot, but new significance or detail was realised for each:

- **The proposals mechanism is essential**
  A human user managing a knowledge graph will wish to have control over their content, and in a dynamic environment where multiple machine agents

can contribute new knowledge at any time it is important that the user has control over the acceptance of this into their knowledge graph. For the pilot exercise the ability to have proposals accepted by default was introduced as a simple mechanism to enable rapid iteration and easy demonstration, but the ability to support more advanced forms of collaboration was clear, especially accounting for human user cognitive needs such as layout and level of detail. This topic is revisited in Section 7.2.2.

- **Acceptance of content may require modification**
  Figure 5.7 shows a simple three-layer model (a, b, c) for the integration of a service into Cogni-sketch. At the outset of this pilot layers a and b were well understood. Layer (a) represents services that exist already, or will be specifically created, with layer (b) representing them when integrated as machine agents. Layer (c) had been hinted at in the XAI literature in the difference between the terminology of *explanations* produced by models or agents, and *interpretations* as formed by the receiving agent. The proposals mechanism should allow modification of information into the knowledge graph when needed, but there may be other simpler cases that can be automatically applied or proposed. In this pilot this was shown with the consumption of certainty information, and the creation of a small set of simple rules to modify *source* certainty information into *actual* certainty information, for example by discounting at a uniform rate for material generated by a service from a particular partner. The ability to define ingestion filters or similar based on the types of palette items being generated and other contextual factors is likely to be a useful capability that is aligned to the core *knowledge fusion* relevant factor.

- **Layout is important for communication**
  The initial attempt to create a simple generic layout algorithm for the generated nodes was not successful. The machine agents created dense (or

dispersed) graphs and were hard to visually parse. Further focus on this would likely yield better results based on the node types and links, and a simple but extensible semantic model of the visual/layout rules to be used could be easily added into the environment. However, for the purposes of this pilot, and given the fixed nature of the input and output data it was feasible to simply manually layout the nodes and links and then use these relative positions for generation of the graphs by the machine agents. This enabled communication of the results in a clean and meaningful style, with acceptance that additional development effort would be needed to achieve something equivalent in a more dynamic setting.

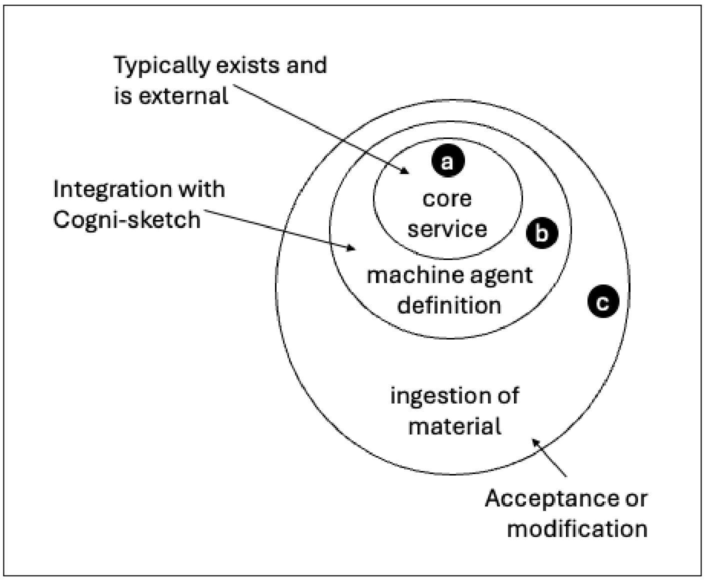


Figure 5.7: Ingestion of content from machine agents

This pilot scenario highlights a loose coupling of hybrid types of machine agents in the Cogni-sketch environment. Moreover, this exercise has demonstrated AI services operating on a variety of data modalities, to provide awareness of SU to human operator users, and the ability to compute certainty based on unfolding events and accounting for affiliation of machine agents and human users.

The ability to use the Cogni-sketch environment as a flexible and extensible environment to support sharing of explanation information was further extended in [153] which focused on a machine agent tasked with detecting content violations in social media platforms and the explanation of the perceived violations within the Cogni-sketch environment, as a direct follow on from the social media monitoring aspect of the pilot scenario reported here.

5.3 Evaluation: Real-time event detection and explanation

The previous section described the pilot for integration of machine agents with human users in the Cogni-sketch environment. That pilot confirmed that the information these services produced could be converted to the correct format and structure for co-construction onto the Cogni-sketch canvas through the available APIs and plugin extension points. This required the creation of a generic palette that could be rapidly extended to support specific details from the various services, along with small wrappers for the services to define them as independent machine agents and allow them to read and write relevant knowledge and information from Cogni-sketch. Having proven that this task was possible during the pilot we moved to an evaluation of this in a real-time setting rather than using predefined events whose structure and content was known in advance.

This evaluation aimed to more formally test whether such machine agent integrations could be undertaken in conditions aligned with the goals of HAKF and Cogni-sketch. Specifically, whether the environment can be rapidly extended and configured at a tempo relevant to the operation, otherwise a slower but more traditional software development process could be used in such cases, with associated delays and constraints in flexibility. This evaluation took place during Covid-19 restrictions so the scenario needed to be one that could easily be achieved in a typical home setting[4].

5.3.1 Objectives

This evaluation focuses on event-recognition in support of SU and, like the previous pilot exercise, draws on capabilities provided by machine agents as well as human users.

Building on the findings from the earlier pilot exercise, there are two further questions that informed the scope for this evaluation, again motivated by the same broad *RQ2* (machine assistance) for this research thesis, introduced in Section 1.3.

Q1 Can entirely new machine agents be built using local sensor feeds at an appropriate operational tempo?

Q2 Can a human user configure and connect multiple relevant machine agents to guide their interaction towards a specific goal?

The evaluation aims to investigate whether the kinds of machine agent capabilities shown to be possible in the pilot can be more directly configured and controlled by human users. In addition, the desire to show real-time event detection from a live sensor source leading to event inference. The example chosen serves as a simple proxy for more complex remote machine agents that can perform similar but more advanced variants. From an audience perspective this work was aligned with the Information Fusion community, leading to an alignment of terminology and validation that the HAKF approach and Cogni-sketch implementation were of interest to that community.

5.3.2 Method

One example of Cogni-sketch usage, relevant to SU and sensemaking, is that a human operator user can easily define target events that correspond to entities

[4]See video V10 (`https://www.youtube.com/watch?v=QiCOYQywjs8`) for a video demonstration and associated description of the research for this evaluation, as presented in [27].

that can be observed by a machine agent with vision processing. Event definition is possible because the Cogni-sketch palette is extended to contain a *detectable object* palette item, instances of which can be generated by the machine agent. These are created as new palette items, as specific sub-types of the detectable object palette item. For example: using a webcam in a typical home setting, detectable object may include people, animals, household items etc. The human users (or other machine agents) are then able to take these new palette items and use them in their models for subsequent processing, thereby enabling communication between the human users and machine agents based on a simple semantic definition within the palette.

With regards to the terminology from the Information Fusion community: This example highlights the use of low-level information fusion techniques in the form of an assessment (e.g., object tracking and identification) and higher-level information fusion (e.g., situation assessment and user refinement). It also highlights involvement of the human users, for example in supporting SU, as well as process refinement of the machine agents by those human users[5]. Since Cogni-sketch is based on HAT, all agents (human and machine) are recognised as both producers and consumers of information, and supported within HAKF through the tellability and explainability flows respectively, and the general co-construction approach.

5.3.3 Findings and discussion

As a simple illustrative example for this evaluation, a *breakfast* event is defined, based on the detection of a person, an item of fruit, and a cup within a 20-second period as shown in Figure 5.8. This innocuous example was chosen as a proxy for

[5]Consistent with the Data Fusion Information Group (DFIG) multi-layer model of information fusion (also referred to as the Joint-Directors of Lab (JDL) model) [17], the user is called out as both a consumer and producer of information within such information fusion systems, consistent with the definition of "Level 5 - User Refinement" within that model.

more sensitive equivalents that might be based on the same kinds of sensors and agents, with all the detection, definition and extension steps being equally valid in any problem domain. For example, the more advanced remote sensing services described in the previous pilot example, seeking SU relating to violent events in the community.

To detect such breakfast events a local webcam is used[6], but any video source to the system could achieve the same result. The *webcam* is represented as an item in the palette and is dropped onto the canvas by the user. A separate *object detection and tracking* function is also available on the palette, which is a simple wrapper to an existing widely used video object detection algorithm (CocoSSD [85]). Finally, there is a simple *event processing* algorithm that is wrapped as a separate machine agent, embedding an approximation of [161] and is based on the listener service from the earlier pilot exercise. The plugin creator user built both the webcam palette item and event processing machine agent for this evaluation.

[6] For a video demonstration of this evaluation please see video V5 in Appendix A.3.

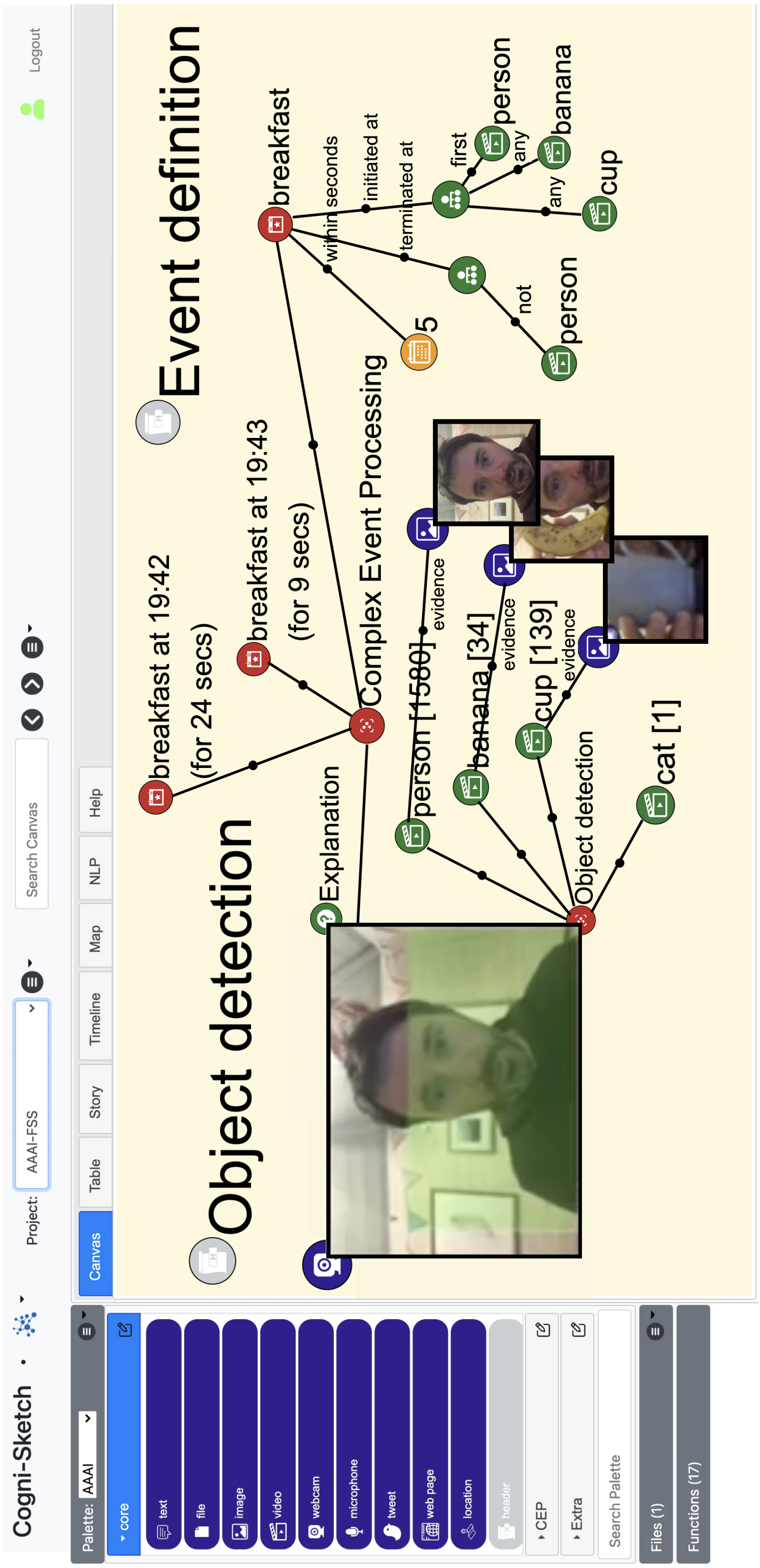


Figure 5.8: Cogni-sketch used for object detection and event definition

The config creator user creates palette items in the palette to represent the objects of interest in the *breakfast* example and draws instances of them on the canvas according to a simple model for defining events from constituent observations. This is a form of tellability that creates configuration data for the generic event processing agent based on what the user has created on the canvas. This can be thought of as the human operator user defining a rule for the event processing agent which can then immediately be used. Any objects that can be reliably detected within the webcam image can be defined as *detectable object* palette items and therefore become available to be added to the canvas and can be created on detection. The rule in this example is shown in Figure 5.8 and defines a *breakfast* event as something initiated at the appearance of a person followed by any item of fruit and any cup within a 5 second period. The breakfast event is terminated when the person is no longer detected.

Once drawn, the model is checked in real-time against objects detected from the webcam, thereby integrating live object detection to event processing, with detected objects and events also being written to the canvas via *proposals* from the event processing machine agent using the same mechanism described in the pilot. The ability to access local webcam and microphone sources, plus the integration of the entity detector function and the event processing machine agent were simple to implement for this evaluation by a plugin creator user, using the existing plugin architecture for Cogni-sketch. The integration was carried out by the author of this thesis, only using the available plugin extension points defined in the SDK.

Figure 5.8 shows the embedded entity detector agent generating different events and capturing these as nodes on the canvas with the bounded entity image from the live feed video providing a form of provenance for the detection. These detected events are then consumed by the event processing agent which in turn can infer a *breakfast* event based on the defined rule and detect the end of the breakfast event (when the person is no longer detected) and compute the duration of the event. All the information relevant to the detection is created on the canvas directly by the event processing machine agent. The layout of nodes was achieved using simple relative spatial layout patterns, and the human operator user can manually reposition these as needed. In this evaluation there was no attempt to mitigate large numbers of detectable object nodes being generated, but in any real system techniques would be needed to handle this to avoid information overload for the human user(s).

The evaluation was simple but successful. With these new services available, Cogni-sketch enables an operator user to access their webcam to detect entities and therefore infer tasks that the user performs, supporting rapid experimentation with different models and services. The evaluation specifically aimed to test whether an example of a flexible service like this could be created and deployed into the Cogni-sketch environment, and the machine agents were able to communicate and co-construct information as well as consuming task-relevant knowledge from human users to configure their behaviour. This simple evaluation incorporates the ability for an operator user to label objects, track objects and detect events of particular interest to an separate *executor* user who may sit outside the system; for example, *late breakfast* events that occur outside a predefined time window may be of interest to an executor, whereas breakfast events are standard and can be ignored by the operator. The human operator could configure the environment to detect these late breakfast events and receive alerts, enabling them to review the situation and prepare a summary before contacting the executor (assuming human intervention delay made sense in a real setting).

No formal assessment was undertaken for this evaluation since it was simply testing whether the extensibility aspects of the Cogni-sketch environment could be used quickly and effectively, rather than any broader evaluation involving additional users or more types of activity. The work was demonstrated live to the Information Fusion community during the virtual Information Fusion conference in 2020 and is reported in [27].

Specific findings from this evaluation were:

- **Integration with a local sensor feed was simple**
  The webcam and microphone capabilities were implemented directly on the operator user's browser and were therefore traditional palette items rather than machine agents that are located in the functions palette. These were simple to implement and gave a direct video and audio feed within the browser environment. Creating these custom palette items was simple and aligned with the existing SDK provided for custom palette items (previously used mainly for rendering, such as for embedded videos or tweets). Making the live feed available to a more traditional machine agent service was also straightforward and achieved by simply creating a link between the webcam (or audio) feed and the machine agent. This then provided easy programmatic access for the plugin creator user to access the feed via the existing SDK. This flexibility between existing extension points with new capabilities was encouraging.

- **Beginnings of an orchestration flow**
  This was the first example in which the operator user was configuring a workflow from sensor to processing directly within their environment. There may be better solutions for doing this at scale, or in a production setting, but for experimentation and exploration this is a powerful adjacent capability that complements the more traditional co-construction use cases for Cognisketch. Specifically in this example the operator user defined the rules to be used for detection of events, using named entities that could be detected in the video feed, and aligned with simple semantics for event definition. They then connected this rule to the event processing machine agent, and the webcam feed to the entity detector, and in doing so they built their end-to-end flow, with events being generated based on the detection of entities and the execution of the rule. This was not a focus area for this research but the ability to achieve this in the same knowledge co-construction setting was an important demonstration of the potential for other use cases.

- **Layout is important for communication**
  This is identical to the finding from the pilot, but in this case, it was impossible to pre-define the layouts to be human-friendly like was done in the pilot. Specifically for the generated provenance events for the detected entities (where the bounded area of the video that corresponded to the detected entity was copied onto the canvas as an event node and linked to the object detection node). In this evaluation a simple layout style was used, and the events were laid out in a radial slice structure, meaning that only a certain number of events could be generated before overlaps would occur and the canvas would get messy. In a real system it is likely that these provenance events would be hidden by default and only needed when validating an event, but just like the pilot this issue of layout for human consumption is also important.

5.4 Using conversation for explanation

In this section we report on earlier research that followed the initial informal definition of HAKF but predated (and informed) the final HAKF design, especially the required capabilities relating to explanations. A conversational mechanism was the most useful form for the interaction style, with a variety of XAI considerations, as well as different types of explanations considered. This also predated the build of the Cogni-sketch environment and drove the requirement for knowledge co-construction as a broader and more useful base rather than a simpler conversational interface that would be more limited.

This exercise was more informal than the previously reported pilot and evaluation exercises, again based on *RQ2*, but was shaped by a single research question which was an important focus for the research at this stage of our activities:

Q Can explanations be generated and shared, accounting for contextual factors and choosing between explanation types as needed?

This question was informed mainly by the variety of techniques for explanations and emerging techniques for providing post-hoc interpretability methods, as well as the recognition that a broad variety of contextual factors need to be accounted for in our particular operational setting. Later work from Miller [95] was also highly relevant, investigating research from the social sciences that relates to the topic of XAI. Whilst our conversational explanation work reported in this section predated the publication of Miller, it was encouraging to see good alignment with a number of relevant social interaction factors as identified in Chapter 2.3. For the work reported here the conversations themselves are all simple examples, with the explanation details coming within a single response from the machine agent following a *why?* question from the human user.

Here we investigate multi-modal explanation types and outline the development of an initial conceptual model to support the provision of explanations and related information. This conceptual model was statically defined here but informed the need for dynamic palettes in Cogni-sketch. A simple scenario is defined, describing a publicly available dataset with multi-modal derivatives that are useful resources for this work, and three specific services that are the subject of the explanation examples.

This conversational explanation example uses still images and video related to traffic congestion, with full details provided in [102] and [61].

5.4.1 Scenario

Three simple services are defined that can be used against the traffic-related video and imagery data. These services are listed below and shown in Figure 5.9:

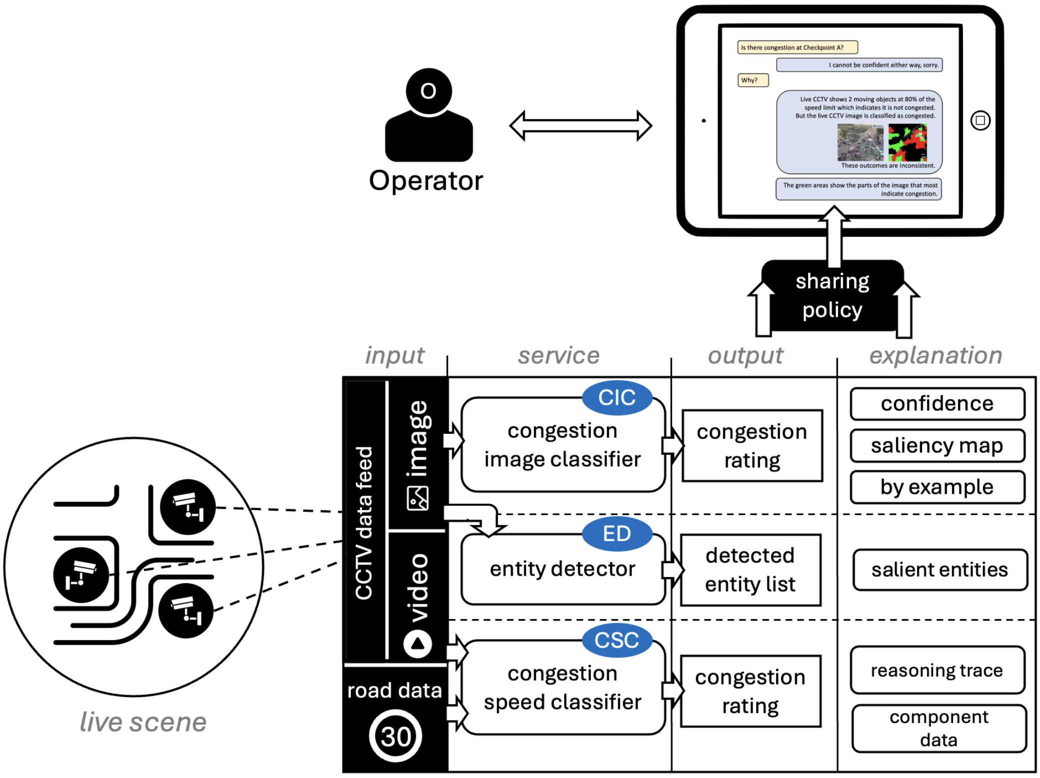


Figure 5.9: Explanation-oriented services and data sources

- Congestion Image Classifier (CIC)

  A trained ML model that classifies images as *congested* or *not congested*. It is a black-box algorithm and requires post-hoc explanation via saliency mapping to show the relevant parts of the image that led to the classification. The service is trained on images from multiple scenes in different conditions (day, night etc).

- Entity Detector (ED)

  A simple ML function that will detect entities within an image. It can detect *salient* entities such as cars or buses, as well as *other* objects like bushes or lamp posts.

- Congestion Speed Classifier (CSC)

  A video-based service that computes the speed of objects based on their transit between frames and has knowledge of the speed limit for the road being observed.

Each of these three services can contribute to the declaration of a congestion status indicating whether the sensor data shows a congested road. The two classifier services (CIC and CSC) can directly declare a congested/not-congested status from the same data source but using different techniques, and through the generation of higher-level derived data in the case of the CSC. The entity detector service (ED) cannot directly declare a congested/not congested status but can be used to provide further insight or evidence to support either of these classifications from the other services as part of an explanation through the identification of entities, and in some cases salient entities (those deemed relevant to traffic congestion such as cars or buses).

5.4.2 Conversational explanation examples

The examples that follow take an abstract form of text-based messages, with support for additional modalities such as embedded imagery within the textual response. In addition to the use of CCTV imagery by the machine agents there is the concept of nationality and geolocation. In the examples you will see cases where a machine agent from a different nationality may not be able to share all information, and the human user asks for congestion information for different locations (checkpoints) as they are seeking SA about a possible route through the city. Full details can be found in [26] along with a corresponding conceptual model for the sensors and other features, but the abbreviated description included here is sufficient to understand the examples that follow.

Case 1: Fully transparent explanation

In this example the human user has asked whether there is congestion at checkpoint A. The response is definitive: The system is confident that there is no congestion. Upon asking for an explanation the human user is provided with a transparent explanation based on usage of the Congestion Speed Classifier (CSC) service. Numerous moving objects were detected, and they are moving at 80% of the speed limit.

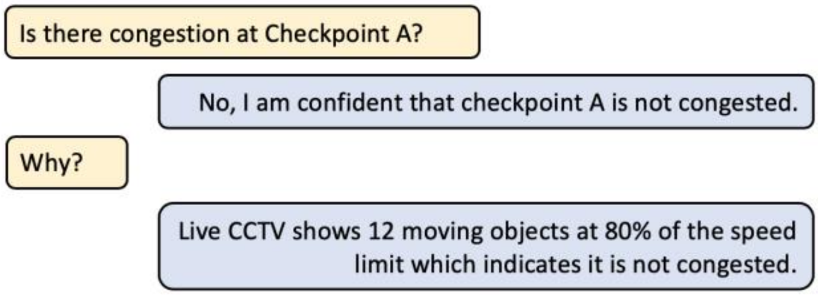


Figure 5.10: Fully transparent explanation example

This information is relayed directly to the user as shown in Figure 5.10 and shows the inner workings of the rule-based system (in the form of a reasoning trace), with a small number of rules and is therefore a transparent explanation from a rule-based system. Note however that the components that yielded the input to that rule are not explained. In a Cogni-sketch solution all this information would be created into the knowledge graph by the machine agent, enabling a deeper exploration of the explanation(s), additional annotations from the human user, and could also be surfaced as a conversational interaction through the chat pane plugin (See Section A.4).

Case 2: Post-hoc explanations

There are two variants within this example, both of which result in a post-hoc explanation, with the system able to determine which is the correct technique to use in each case, based on the user profile and their affiliation.

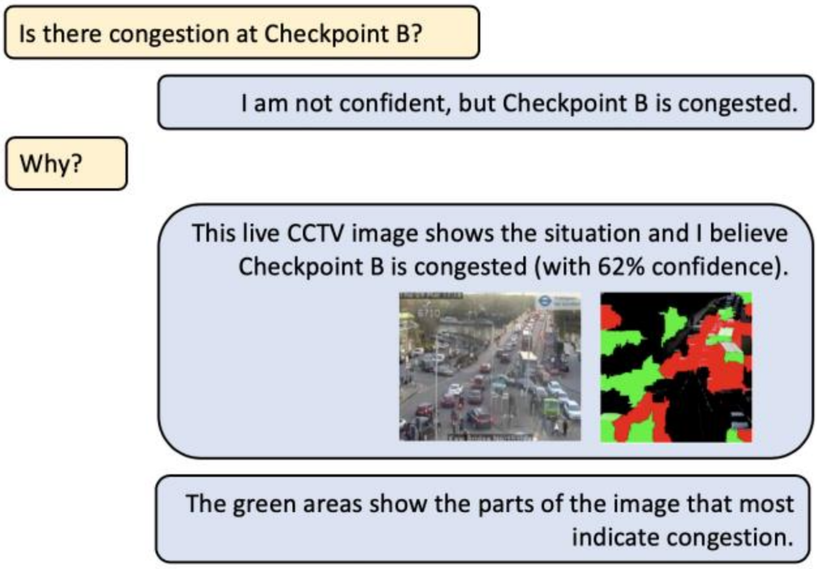


Figure 5.11: Post-hoc explanation via saliency mapping

In the first post-hoc explanation case (Figure 5.11) the user is told that the system has low confidence that the checkpoint is congested. Upon asking why, the system responds by showing the user a saliency map (highlighting the areas of the image that were most relevant to the *congested* classification by the CIC service), using simple human language to convey the confidence of that classification (low). The user concludes that the system has correctly identified congestion when they are shown the image, but it is likely the raw image rather than the saliency map explanation that convinces the user, based on their human understanding of what the image means, regardless of the saliency map which is mainly providing insight into which aspects of the image were used by the machine agent to achieve the classification.

This is an example of post-hoc explanation via saliency mapping, with the saliency map being generated by the LIME technique [124]. Both images are provided to enable the human user to decide what is relevant and important, and it may be a precursor component rather than the final result; in this case the raw image rather than the saliency map. With a HAKF solution such as Cogni-sketch the user will have access to all relevant information in the knowledge graph, assuming they have suitable permissions, and can therefore decide for themselves what is relevant rather than a chat system designer needing to make that decision in advance.

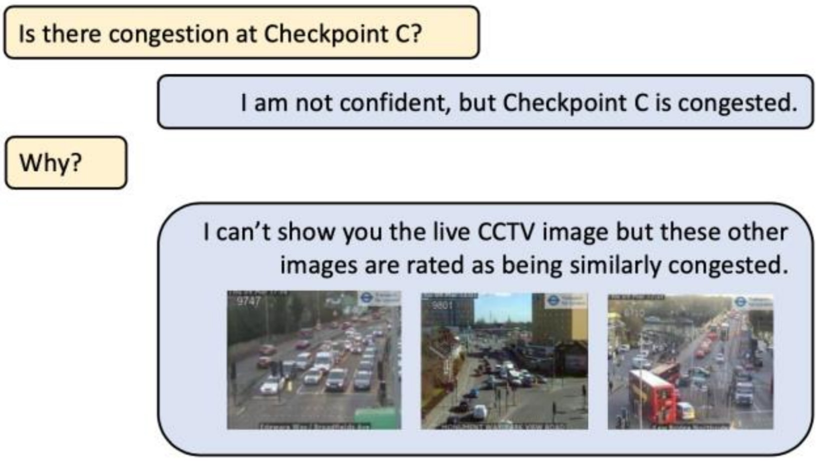


Figure 5.12: Post-hoc explanation by example

In the second post-hoc explanation case (Figure 5.12) the user gets the same response but upon asking for an explanation the system concludes (through the role and affiliation of the user) that they are not authorised to see the original image for security reasons. The system is therefore not authorised to show the image, or anything derived from it, to explain the classification to the user.

Instead, the system chooses to show a series of images that are similarly congested to the classified source image. This is achieved by using the scalar value for the degree of congestion detected in the image, showing other images that the user does have the authority to view (e.g., from the training data) which have a similar level of detected congestion. An alternative to this would be to show the textual results of the Entity Detector (ED) service if relevant.

The user finds it hard to conclude whether the system has correctly identified congestion but sees that the example images served are indeed congested. This is a form of *post-hoc explanation by example* and may provide confidence in the general model quality, or the training data used, but maybe not the specific classification since no actual information about that can be shared.

Case 3: Disagreement within services

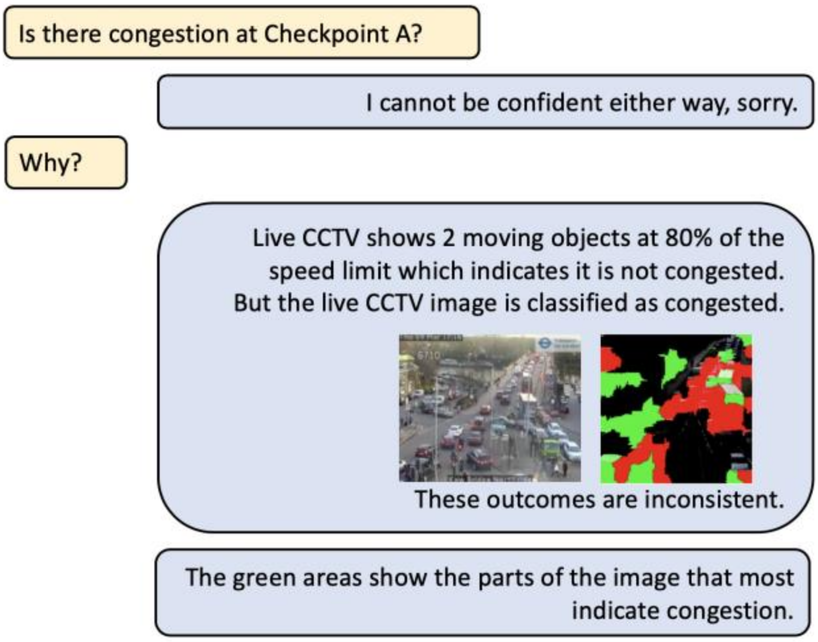


Figure 5.13: Combined explanation arising from inconsistency

In this final example, as shown in Figure 5.13, the user is told that the congestion status is unknown due to inconsistent information. For an explanation the user is advised that the CSC has concluded *no congestion*, based on the speed of objects within the video, whereas the CIC has concluded *congestion* based on the ML model trained on imagery. The saliency map explanation is again provided since the user is authorised to see the image, and in this case the user is forced to make a judgement for themselves.

This is a combination of *transparent explanation* from a rule-based system, and *post-hoc explanation* via saliency mapping. The ability to detect inconsistencies across services is useful for alerting to possible cases of misclassification and the ability to do so will increase as the number of data sources and relevant services increases within the overall system. Delegating the final decision to the human user in this way is the simplest solution and motivates the need for a HAKF environment to better enable the human user to explore relevant information in the graph, rather than building increasingly complicated machine agent solutions to attempt to present a final answer to the human user.

5.4.3 Lessons learned for HAKF

This early research into conversational XAI techniques and their ability to be integrated into a HAT setting highlighted several issues that would inform the required capabilities for HAKF and ultimately a solution like Cogni-sketch to mitigate such issues:

- **Significant amount of specific code was needed**
  To achieve these integrations specific code had to be written by technical people; in our case the researchers carrying out the work. This included writing the core services, but more importantly the interaction templates for the chat messages and the decision of whether to embed images etc. The overall desire is for a system in which services can be made available and consumed by less technical human operator users as described in Section 4.3.1, although we recognise that the composition of such services into a usable configuration may still require more technical plugin creator users even in an environment with improved ease of integration.

- **Information relevancy chosen during design**
  By presenting the overall outcome during the conversation there are cases where important information may be missed or overlooked. For example, in the first conversation example (in Section 5.4) we state that "After asking for an explanation the human user is provided with a transparent explanation based on usage of the Congestion Speed Classifier (CSC) service: numerous moving objects were detected, and they are moving at 80% of the speed limit". This is correct and accurate however it overlooks that the

CSC service is comprised of two steps. The second step is the one reported in the explanation and is correct. The logical inference rule provides a fully transparent explanation and that is what is reported. However, the first step is the computation of the speed of the vehicles from the video image feed (by comparing frames to detect distance travelled for individual vehicles and then to compute the speed). There may be issues in the processing of this first service, but for the case we reported this was assumed to be accurate and not requiring investigation.

By using HAKF the existence of that sub-service and the outputs from it (and their associated confidence) becomes more information in the knowledge graph and therefore available to the human users if needed (See the earlier evaluation in Section 5.3 for examples). The ability to carry out a higher-level conversation against all that data would need to be delivered and integrated as a set of bindings between the graph and the conversation outputs but by using HAKF and Cogni-sketch it is possible to achieve a more thorough and higher fidelity solution without the need to write as much code.

- **Security of information and sources is important**

  The concerns of the service providers around ensuring obfuscation of their capabilities can be provided in a secure manner just as was the case in this example. Depending on the specific security requirements this additional information can either be suppressed within the machine agent, or it can be available to Cogni-sketch but withheld from users without appropriate permissions.

So, whilst the examples reported here predate HAKF and Cogni-sketch this exercise directly informed both. Specifically in terms of motivating the need for a much more flexible environment to support information co-construction between human users and machine agents at a lower level of technical fidelity but with less code needing to be written by creator users, and the key items within a conceptual model being available to the environment and able to be extended when needed (as the palette).

5.5 Chapter Summary

In this chapter a series of worked examples have been used to show how the Cogni-sketch environment based on HAKF can be used to rapidly align data feeds and analytic explainable services. These convey information from those services to other human users and machine agents as co-constructed information within the knowledge graph. The examples show a set of capabilities starting with a simple pilot exercise to integrate existing services operating on predefined sources, and a more formal evaluation based on real-time event detection and integration of a simple workflow via the knowledge graph. Also included are a small set of early examples based on conversational interactions for explanations that predates and informs HAKF and the eventual Cogni-sketch implementation.

The focus in all these examples has been the human operator user as the consumer and creator of information within the Cogni-sketch environment, alongside relevant machine agents. Also important are the human roles of config creator and plugin creator, especially for their activities to define new machine agents and configure the environment at a tempo appropriate to the operation.

These examples show how both plugin creator and operator users are able to use HAKF and the Cogni-sketch environment to rapidly define or configure analytic services and specify events for later use in a SU context with the executor human users likely sitting outside of the core Cogni-sketch system and receiving alerts from Cogni-sketch or embedded machine agents or being informed by human operator users of the system.

Table 5.1 summarises the main findings from each of the three exercises reported in this chapter, and suggests the closest matches to both: *relevant factors* that inform the requirements and design for HAKF, and *required capabilities* that aggregate these into higher-level collections to better inform implementations such as Cogni-sketch. For the relationship between relevant factors and required capabilities refer to Figure 3.6, and the associated textual descriptions for their definitions in Sections 1.1 and 3.4 respectively. In Table 5.1 the different exercises are referred to as *scopes* due to their relative focus to machine agent integration as defined in *RQ2*.

The coverage of relevant factors and required capabilities shown in Table 5.1 is encouraging, with all factors relevant to *machine agent integration* featured across the various scopes of the exercises. A number of the relevant factors related to *rich knowledge representation* are also seen to be relevant, as are a smaller number for human users both in terms of their ability to configure and direct machine agents, as well as their cognitive needs for consumption of machine-generated information as seen in the multiple occurrences of *visualisation and interaction*.

Whilst there has been no formal evaluation of the ability to create machine agents within the environment, or an assessment of the materials that they create, it is clear that a wide variety of typical capabilities can be implemented using the existing APIs and SDK. The target user for building such machine agents is a technical/developer user in the role of plugin creator, but they must also be aware of the impact on the operator user for the material that is created by their machine agents. In both the pilot and the evaluation it was clear that the proposals mechanism for machine agent contributions is important, leaving the operator user, as owner of the knowledge graph, in control of what information ends up in their environment. Also relevant is the need to consider the spatial layout of generated information and how it is presented to human users to appeal to their visual preferences. In the pilot and evaluation only simple solutions to this potentially complex problem were considered.

Finally, the earlier work involving conversational explanations informed a number of important considerations into the early design of HAKF, and the subsequent implementation of Cogni-sketch. Any future implementation of a higher-

Scope	Finding	Relevant factors	Required capabilities
Pilot	Proposals mechanism is essential	• Machine agents • Explanation • Knowledge fusion • Operational tempo	• Machine agent integration • Rich knowledge representation • Agile information capture
Pilot	Acceptance of content may require modification	• Human users • Trust • Knowledge fusion	• Machine agent integration • Rich knowledge representation
Eval	Integration with local sensor feed was simple	• Machine agents • Explanation • Multi-modal	• Machine agent integration • Rich knowledge representation
Eval	Beginnings of an orchestration flow	• Human users • Machine agents	• Machine agent integration • Rich knowledge representation
Pilot & Eval	Layout is important for communication	• Machine agents • Explanation • Multi-modal	• Visualisation and interaction • Machine agent integration
Conv	Significant amount of specific code was needed	• Machine agents • Operational tempo	• Machine agent integration
Conv	Information relevancy chosen during design	• Machine agents • Explanation • Knowledge fusion	• Visualisation and interaction • Rich knowledge representation
Conv	Security of information and sources is important	• Machine agents • Trust • Knowledge fusion	• Rich knowledge representation

Table 5.1: Mapping findings to relevant factors and required capabilities.

level interaction mechanism built on top of the core knowledge graph (including conversational interaction, but there may be additional graphical/navigable forms as well) will be able to take advantage of these insights, many of which align to the work of Miller [95], and which will be the source of more in the future.

As the platform matures and the variety and capabilities of machine agents grow it is likely that specific experiments could be designed to measure human reaction to different formats or interaction methods for machine agents. There could be many such experiments, and in the next chapter a firm baseline for the measurement of human activity against a sensemaking task is presented, with results analysed. This could form the basis for a series of machine agent extensions with the potential for a comparative analysis when machine agents are mature enough to make meaningful contributions into that problem-solving context.

Chapter 6

Co-constructing knowledge graphs for sensemaking

6.1 Introduction

This chapter focuses on co-construction of task-relevant knowledge via the tellability flow within HAKF, mainly the ability for human operator users of the Cogni-sketch system to contribute meaningful task-relevant information with directed machine agents performing an important supporting role. This is supported via the required capabilities of *agile information capture* to enable those users to collect and structure their information as they progress, with *visualisation and interaction* enabling the user to arrange their information according to their taste and to better support their cognitive needs. These capabilities are underpinned by the *knowledge fusion* capability implemented as the knowledge graph within Cogni-sketch. The human users are supported where needed by dependent machine agents to perform specific task-relevant activities that are appropriate, such as searching or querying large volumes of data.

Both the pilot and subsequent evaluation were more formally structured than the earlier experimental exploration of machine agent integration capabilities presented in the previous chapter. Details for the pilot and evaluation methods in this chapter can be found in Sections 6.3.2 and 6.4.2 respectively. The pilot involved assisting a long-running intelligence analysis exercise with an OSINT analyst, having only limited ability within the platform for instrumentation to support subsequent analysis. The Cogni-sketch platform was also under active development to support the needs of the analyst during the pilot exercise. Discussion of the findings from the analysis of created artefacts over time during the pilot can be found in Section 6.3.3, with full details of the identification of a better mechanism for instrumentation of the platform and a mapping of events to common sensemaking capabilities as described in Section 6.3.4. The findings from the pilot led to modifications to support a substantial improvement in analytic capabilities for the later evaluation with human participants. This evaluation also included a System Usability Scale (SUS) survey as well as a qualitative analysis of the artefacts created during the experiment. Details of the method can be found in Section 6.4.2, with results reported and analysed in Section 6.4.3.

Through a pilot and subsequent formal experiment, we demonstrate that users are able to capture their information and add different levels of semantic detail as appropriate, but can start with no semantic information at all, using generic text or media nodes from the palette to capture whatever they are able to forage in a sensemaking context. The focus of this chapter is on sensemaking specifically as the use case, but as mentioned previously, HAKF and the Cogni-sketch implementation are not limited to operating in only the sensemaking domain.

6.2 Supporting sensemaking

As observed in Chapter 2, there are relatively few specific calls in the literature to improve the sensemaking process for human analysts from a technology system perspective. During this earlier literature review a small number of these were identified and the most relevant are revisited in this section, providing details of support within HAKF and Cogni-sketch or, in some cases, the minimum updates that would be required to do so. In addition, there is a brief discussion on the Pirolli and Card sensemaking model [113] that has been chosen as basis for sensemaking in the pilot and experiment.

Before getting to these topics, we start with a brief overview of the relevant human user and machine agent roles for these sensemaking activities when undertaken in the Cogni-sketch environment.

6.2.1 Roles for sensemaking

It is important to briefly revisit the roles of any agents that may be involved in such sensemaking systems, since this subset of roles will be mentioned throughout this chapter. The full set of roles for a Cogni-sketch environment were shown earlier in Figure 4.3. Generally, the activities of the *creator* role (specifically the *core creator* and *plugin creator*) all occur prior to the usage of the system which is undertaken by the *operator* users. It is likely that these operator users will also carry out *config creator* activities to extend their environment as they use it, for example by customising their palette.

The focus of the pilot exercise and formal experiment reported in this chapter are on the activities and behaviour of the operator users, with any config creator activities they perform being called out where relevant. During the long-running pilot exercise there was an ongoing feedback loop from the intelligence analyst (operator) and the Cogni-sketch creator (the author of this thesis), with extensions to the core platform and plugins being undertaken to improve the experience for the operator throughout the pilot. *Executor* users are implied but not explicitly covered in the pilot or the experiment since they sit outside the system, but the artefacts created by the operator users would typically be used to brief the executors in a real setting. The focus of this chapter in on these human user roles, but *machine agents* (*directed* and *independent*) are mentioned when relevant.

6.2.2 Supporting sensemaking principles

In this section the 9 principles from Attfield et al [8] are listed, along with a brief description of how HAKF, and specifically the Cogni-sketch environment directly support each of these principles. These principles were introduced previously in

Section 2.4.1 where each of them has a summary description provided which is not needed here but may provide useful additional context to the list below. (For the corresponding list of possible improvements to further support them please refer to Section 6.5.2).

#	**Support in Cogni-sketch**
PR1	*Provide sufficient cues for sufficient sensemaking*: The ability to rapidly build knowledge graphs from relevant fragments of multi-modal data is core to the Cogni-sketch environment. As discussed in Section 4.3.3 the ability to create palette items to serve as markers or placeholders for further investigation, or to task other investigators is very relevant to *PR1*, along with labelling hunches and ideas within the graph, directly alongside more explicit information.
PR2	*Support low-cost information workflows*: The ability to invoke directed machine agents on demand and in context, and to rapidly integrate new machine agents or repurpose existing ones for new settings (e.g., to perform targeted NLP, information extraction, visual processing, etc). Also, the ability to define appropriate semantic definitions within the palette to support machine inference whilst not over-burdening the human users with these semantics from the start.
PR3	*Represent information quality and provenance*: Good support for capture and communication of information quality and provenance meta-data, with machine agents able to provide uncertainty and provenance information automatically as properties against the nodes or links that they create (e.g., see Section 5.2). Corresponding ability for the human user to also do so, and can be encouraged/enforced through plugins, when appropriate.

PR4	*Promote expertise and domain knowledge*: (1) The ability to define (and share) palettes comprising palette items with defined properties against each. The ability to add simple semantics for inheritance and domain/range restrictions, enabling external machine agents to reason over the expertise captured in the knowledge graphs created with these attributes. (2) Easy embedding of custom directed machine agents which can operationalise any repeatable and non-creative procedure that the core or plugin creator may wish to encode. These can consider the context of the knowledge graph at the time of invocation.
PR5	*Allow time to acquire data/information to build an evidence-based and coordinated situation picture*: Different panes (such as table and timeline) allow the user to interact with the knowledge graph in different ways and can support a more methodical (system 2) approach. Also, the ability for multiple users to concurrently access the same information and, where appropriate, observe the real-time construction of it. The incorporation of core material from the knowledge graph into higher-level narratives (such as stories, described in Section 6.3.3), and the ability to extend and share task-relevant models in the form of palettes.

PR6	*Use strategies for negotiation of sense*: The ability to easily share knowledge graph information either entirely, or as relevant subgraphs, and that each element of each graph is annotated with the creation time and the originating user/agent as well as any explicit meta-data. This could support a bounty system for rewarding human users based on their input, or for identifying high-value gaps to incentivise collection activities, either through machine agent quantification or human expert user annotation. Therefore Cogni-sketch already supports both the *advertising* and *tracking* of this incentive data by the creation of specific node types (as palette items in the palette).
PR7	*Where appropriate, use strategies for frame enumeration and elimination*: The ability to dynamically fit a subset of the knowledge graph against different frames is not yet well supported. An experimental simplistic mapping of twelve structured analytic techniques [56] was used in the definition of stories to support diagnostic, contrarian and imaginative thinking sensemaking activities and can be seen in Figure 6.4.
PR8	*Provide explanatory context for actions, orders and requests*: It is already possible to define palette items to capture future requirements such as 'to do' actions, markers for further information gathering, or simply requests for others to do something. These nodes can be linked to other nodes to provide relevant context or use agreed properties on the node(s) and/or link(s) as required by the tasking human user or machine agent. The human user (or machine agent) responding to the request can contribute any resulting information using the same techniques.

PR9	*Minimise the costs of achieving and maintaining common ground*: Cogni-sketch is designed explicitly with this goal in mind, and simply by providing access to the knowledge graph (and corresponding palette) a level of common ground can be achieved. It may be that not all relevant information is able to be practically stored in the graph, so additional techniques to convey important contextual information such as traditional intelligence briefings or discussions between human users can also help.

Table 6.1: Support within Cogni-sketch for Attfield et al's 9 principles

In summary, these nine principles are generally well supported by both HAKF and the Cogni-sketch implementation, and in cases where there is not yet support there is usually a clear path describing how to achieve the desired goal (See Section 7.2 for specific examples of this). In many cases the support is necessarily generic since the goal for HAKF and Cogni-sketch is to support a wide variety of problem-solving use cases and not just sensemaking. However, as needed each of these requirements could be more explicitly supported through the creation of specific plugins or through modification of existing ones. The flexibility of the basic approach plus the reusability of certain components that were designed for an adjacent purpose is encouraging. Whilst Cogni-sketch was not developed with these 9 specific principles in mind it is good to see the degree of natural alignment and coverage that is already present.

6.2.3 Pirolli and Card as a model for sensemaking

Several sensemaking techniques were introduced in Section 2.4, and the work of Pirolli and Card [113] has been selected as the model to be used. This therefore forms the basis for supporting sensemaking within Cogni-sketch for the pilot and experiment, as shown in Figure 6.1. For other methods and a summary of the differences between them refer to Section 2.4.

As a brief recap: the Pirolli and Card sensemaking process comprises a series of interconnected loops starting with the *foraging loop* (bottom left) in which data is gathered from the external environment and assembled into a more coherent body that can serve as evidence, and the *sensemaking loop* (top right) in which schematised evidence can be connected to hypotheses, and cases are built to inform decision making.

The loops themselves denote feedback in the respective parts of the process; further feedback loops exist between each pair of successive steps in the process. The progression of the process from left to right, and bottom to top, represents increasing effort on the part of the analyst, and increasing structure in the infor-

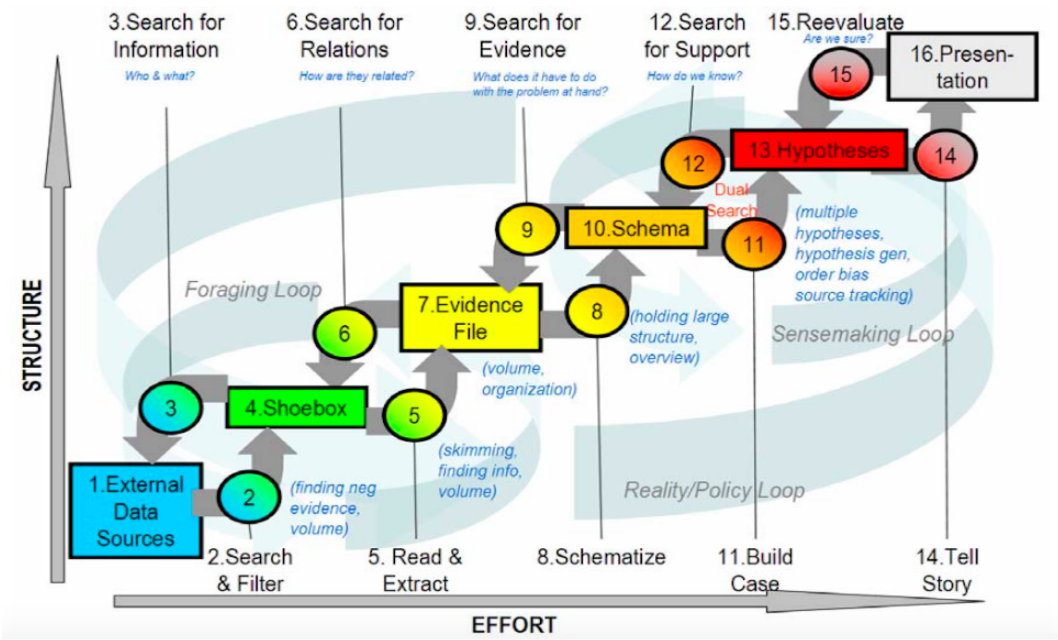


Figure 6.1: The sensemaking process for intelligence analysis (recreated based on [113]).

mation artefacts created. As the process progresses the structure can increase, and material created naturally flows from data to information and then knowledge. The ability to handle all these stages of refinement and increased specificity within a single environment is one of the goals of this research and is explicitly captured by *RQ3*. There is also a *reality/policy loop* which is very important to recognise as the link back to the real world but is not covered in detail here as it is currently dealt with outside of the system. However, insights from this higher-level loop could include process improvements that might lead to improvements in HAKF or additional Cogni-sketch plugins.

The ability for an analyst to rapidly traverse the process to serve any aspect of foraging or sensemaking at any time is of key importance, for example by applying semantic information retrospectively to simpler data foraged earlier in the exercise, or by defining a machine agent to do so based on training, or definition of relevant rules.

Similarly, machine processing can also be very valuable, for example in sifting large volumes of data to enable the extraction of potentially relevant information into the *shoebox* (step 2), whereas human insight is typically needed to set and fine-tune parameters for data collection (step 3). Similarly, between *shoebox* and *evidence file*, machine agents can assist with data analytics (step 5, e.g., anomaly detection to seek 'needles in haystacks'), complementing human judgement in seeking-out meaningful relationships (step 6).

In existing tooling, there is a tendency to either favour the upper parts of the model (schematization and case-building, often via formal representations) or the lower parts (pre-formalisation via shoeboxing and exploratory assembly of evidence) as discussed in Section 4.2. Many analysts fall back on generic mind-mapping or note-taking tools for the latter because of their ease of use and lack of formality (the digital equivalent of the classic evidence board[1]).

Whilst the goal of HAKF is to support any form of HAT based around the co-construction of task-relevant knowledge and information, a useful target application is sensemaking. Human users and machine agents must be able to contribute their information or knowledge at any stage in the sensemaking process and easily move between different stages of the process. It is uncommon for meaningful sensemaking to proceed in a linear fashion and environments that enforce too much structure early in the process will likely cause frustration, cognitive dissonance, or rejection for the human users through the enforcement of structure too early in the process or similar issues. The sensemaking process of Pirolli and Card [113] informs and inspires the sensemaking approach but is deliberately not encoded as specific functions or agents. In the remainder of this chapter, we describer a pilot exercise and subsequent experiment to attempt unified sensemaking in the simple Cogni-sketch environment.

[1]See `https://en.wikipedia.org/wiki/Evidence_board`.

6.3 Open source intelligence analysis: pilot exercise

This pilot use case is focused on human-led OSINT analysis with a focus on sensemaking. The pilot was applied to an ongoing real intelligence analysis operation that was successfully undertaken by a professional OSINT analyst using the Cogni-sketch environment to build a rich and detailed network of knowledge relating to their investigation as shown in Figure 6.2. All text, labels and images have been removed throughout to preserve anonymity.

6.3.1 Pilot objectives

To be successful in this activity it is imperative that the human users can share their insights and provide information back into the system to inform and configure, with other human team members, machine agents, or their future selves, being possible consumers of that knowledge graph information. This is the essence of the required capability for *knowledge fusion* through co-construction that is core to HAKF, as well as the human-facing capabilities of *agile information capture* and *visualisation and interaction*, as supported by the tellability flow within HAKF.

One additional purpose of the pilot was to investigate the potential for a more formal experiment in this space, inform the design of that, and improve the ability to support sensemaking activities within Cogni-sketch based on user feedback from the expert analyst.

The research objectives for this initial pilot exercise were: To support a real OSINT analyst in their attempts to use the Cogni-sketch platform to undertake an OSINT exercise, extending the environment during this pilot to better support the analyst in their goals. The ability to make extensions in real-time during the exercise, in conjunction with the analyst extending their own environment to support their needs, attempts to validate the flexibility of the environment and the ability to support rapid extensions. These research objectives were appropriate for the pilot, and led to a much improved platform with better instrumentation that enabled the definition of formal research objectives for the later evaluation with human participants as described in Section 5.3.1. The primary data created during the pilot which would enable evaluation of these objectives took two forms: The events generated within the platform as the analyst used the environment, and the artefacts they created on the canvas in the form of nodes, links and associated media or content. In addition to this quantitative data there were regular unstructured feedback sessions with the analyst throughout the pilot, with these mainly driving the various modifications and extensions to the platform.

6.3.2 Pilot method

The pilot exercise was carried out with one OSINT analyst over 3 months, following a detailed scoping exercise with this analyst and their colleague. The pilot consisted of sensemaking activities (creation or refinement of knowledge within the environment) on 19 separate days. Most sessions were for brief periods (circa 20 minutes), but 7 sessions were for extended periods of up to 4 hours, representing substantial development of the knowledge graph. During the pilot several refinements and improvements to the Cogni-sketch environment were made in response to feedback from the main analyst plus other informal users and collaborators, including the addition of *story* elements as an additional plugin capability.

There was regular contact with the analyst during this period, with feedback leading to improvements to the core system and relevant plugins in an iterative process. This represented a feedback loop between the operator user (the analyst) and the core and plugin creator user (the author of this thesis). This sparse but high-intensity profile for the pilot meant that the Cogni-sketch system could be upgraded with many of the requirements identified by the analyst during their usage, with other requirements being logged for future consideration. This constantly evolving baseline was one of the main reasons the pilot was informally executed, rather than featuring any detailed analytics of environment usage. The other factor that influenced this decision was the sensitive nature of the investigation.

This live and long-running pilot exercise highlights the way in which the Cogni-sketch system enables the analyst to participate in the various stages of the sense-making process, creating and consolidating relevant information. The analyst was able to build their own extensions to the palette, enabling a more nuanced expression of the types of information they were finding, creating markers for items they were seeking, as well as capturing some aspects of the intelligence gathering process itself. Many of these palette extensions could be reused in subsequent operations and shared with other users.

Cogni-sketch is designed to enable the human user to express themselves visually (in terms of layout and content style) and structurally (in terms of palette items, nodes, relationships and attributes with semantic meaning), with these distinct but related visual and semantic forms to serve different purposes to the cognitive activities of the analysts. This was an important capability during the pilot, the results of which can be seen in Figure 6.2.

Whilst the human user is enabled to capture and formulate their knowledge in forms that support analytical processes without being overly constrained to preexisting structures and formalisms, machine agents can also be used in this knowledge creation context. However, they were not widely used in this human-led intelligence analysis pilot beyond some simple experiments with Natural Language Processing (NLP) and Named Entity Recognition (NER) agents, and some specific productivity agents such as a pdf-to-text conversion agent.

6.3.3 Pilot results

The information shown on the Cogni-sketch canvas in Figure 6.2 represents the knowledge graph that was created by the OSINT analyst, working in an operator role. This operator user was able to successfully represent their knowledge through the creation of nodes, assigning meaningful relationships through links, with optional labels and the ability to directly embed multi-modal task-relevant information within the knowledge graph in the Cogni-sketch environment (text, multi-media, hyperlinks, documents, social media entries and more) to provide additional contextual information. Even with the labels and content removed, the structure of the graph can be easily seen, with the user choosing the layout to meet their cognitive needs and extending the palette with new palette items to represent specialised types as required.

Due to the ongoing development of the platform taking place during the pilot only basic analysis was able to be undertaken. The available data to support this quantitative analysis were: (a) the structure and (obfuscated) content of the canvas, i.e., the created knowledge graph, and (b) the event data from the actions undertaken by the user. The analyses in the following section are based on these quantitative data that were available for processing after the pilot exercise.

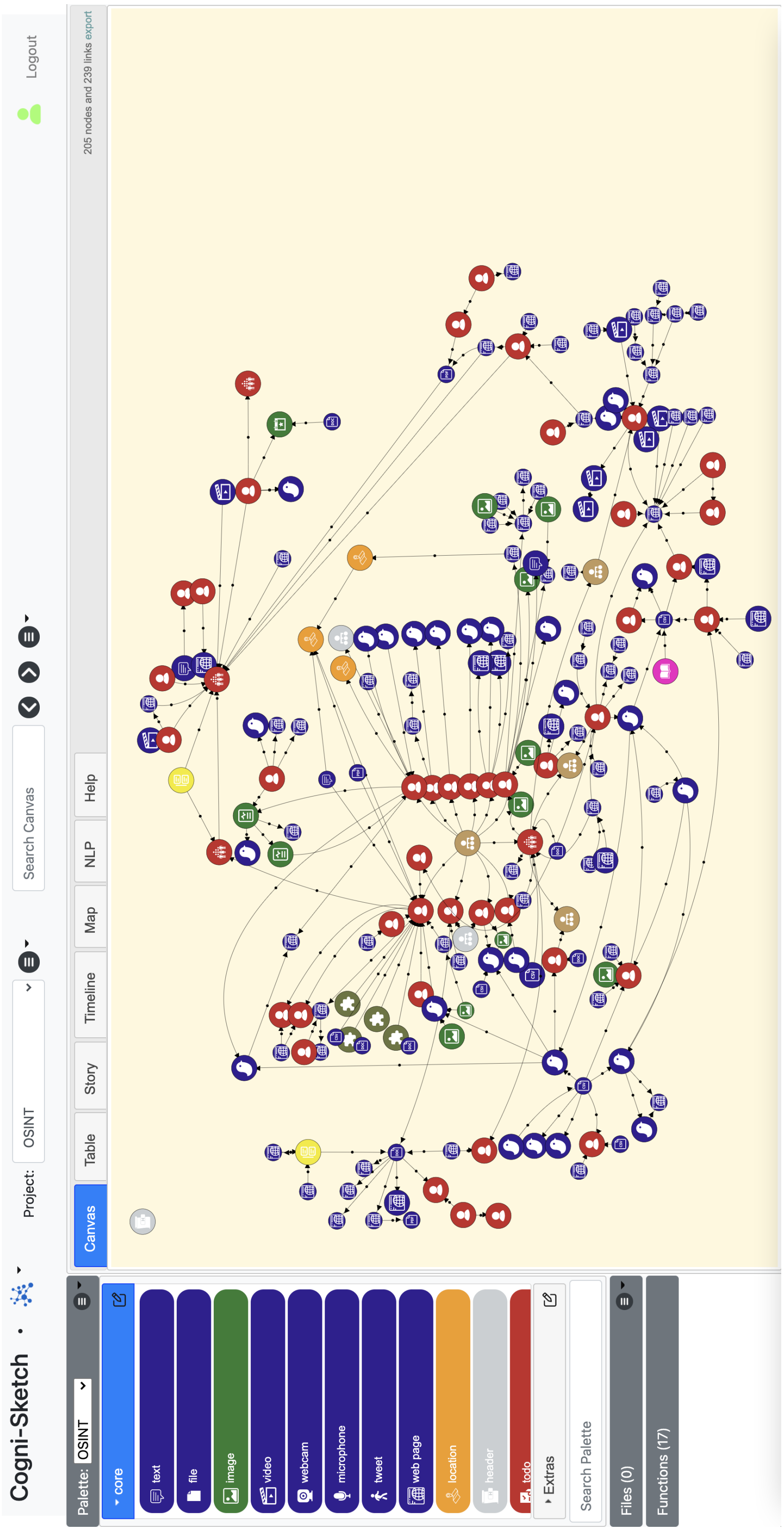


Figure 6.2: Artefacts created through OSINT analysis in Cogni-sketch

The analyst in this exercise was also able to work as a config creator user, to create specific palette items to mark open questions, and successfully capture partially formed thoughts within the knowledge graph, linking these to the relevant nodes, including the definition of explicitly *unknown* nodes of particular types that then served as goals for later investigation and foraging. These were useful and practical markers that could be created quickly and serve as reminders for missing information, or task other investigations or activities without breaking the focus of the analyst on their main investigation.

Informal feedback from the analyst who participated in this pilot was positive regarding this expressive freedom, particularly the ability to lay out their knowledge in a form to suit their own visual and cognitive needs, as well as the freedom to make knowledge modelling decisions and refine these in an iterative process (through ongoing modifications to the palette). The ability to quickly create new projects as separate knowledge graphs and copy/paste nodes and links between them was also valuable, enabling a curated master project to be managed alongside multiple less formal and more dynamic temporary projects. Typically, these were used more for shoeboxing [113], often feeding a subset of this separately curated knowledge graph back into the master project. The analyst mentioned that the expressive freedom triggered their spatial memory for where certain parts of the graph were located and found this experience similar to pen-and-paper notes and diagrams, but something they noted did not occur so strongly in more ad-hoc linear note-taking systems such as MS-Word etc.

Whilst the keyword search function and the different layout styles are useful for exploring the knowledge as it is created, the analyst particularly valued how the graph layout enabled a higher recall (within their head) of the location of data within their graph, supporting faster thought processes and clearer understanding for them.

For this pilot usability was not directly measured (e.g., through a qualitative measure such as a survey) but productivity was, with a total of 201 nodes and

235 links created, along with 9 new palette items to capture a range of specific additional needs for the analyst as shown in Figure 6.3.

The nodes created included: new items to represent more specialised classes of objects relevant to the exercise, items to mark observations or questions for future investigation, and one item to assist with the abductive reasoning process. In Figure 6.3 the growth of user created data within the environment over time for the top 5 data types is shown. For legibility the time series is normalised into generic 'time buckets' generally corresponding to the specific sessions in which the analyst interacted with the system.

Only the main project was analysed, so any temporary or transient work done in a separate project was ignored. The user defined node types are anonymised as *extra* to preserve the privacy of the exercise. As expected, the volume of all node types grow over time, but the creation and subsequent use of the user defined node type shows early recognition of the need for schematization by the analyst. In the earlier Figure 6.2 nodes corresponding to the user defined types can be seen in colours other than blue within the graph, for example the numerous red nodes across the graph as well as a variety of other node types. The other standard types can be seen as various blue nodes in the graph, and as palette items on the left.

The co-occurrence of the core data types and the user defined type, and the continued growth of both as the pilot progressed, indicates that both foraging, and schematization were occurring in parallel as intended. In total nodes from 13 data types were used by the analyst, with 5 of these being user defined extensions, some of which directly supported schematization. Figure 6.3 only shows the growth in nodes created over time, but additional information about link creation and property definition within the nodes and links is also available within the project but not reported for the pilot. Using this more precise data, additional information can be extracted as to the tasks the user was carrying out. For example: capturing raw new knowledge in the forms of text or imagery data,

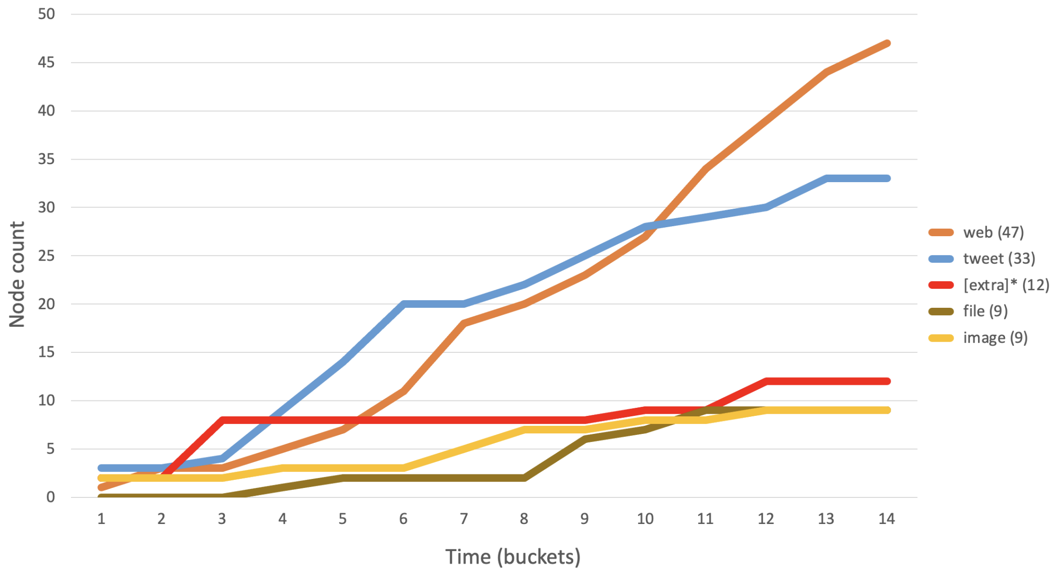


Figure 6.3: Nodes created over time during the pilot

web links or social media posts (shoeboxing), or refining the content of these nodes into richer structures, creating custom palette items, and linking material together with meaningful labelled relationships (schematizing).

6.3.4 Pilot discussion

The findings from this pilot exercise were promising. The positive reaction to the environment and the perceived value of the extensible knowledge-graph-based approach, coupled with the ability for the largely non-technical expert analyst user to be able to create and capture relevant data and extend the palette as needed was encouraging. This pilot experience informed the definition of a formal experiment to test whether multiple users in controlled conditions could create meaningful information in a constrained sensemaking setting and show progress in multiple areas of the sensemaking loops without significant technical training. Also, the ability to respond to the requirements of the analyst within the time frame of the pilot to see whether the functional updates delivered as plugins were beneficial and used by them.

Two specific findings arose from this pilot exercise, one of which was able to be addressed during the pilot itself to benefit the analyst in the creation of story-related material, and the second of which was a detailed understanding of the typical sensemaking tasks that were undertaken and how different events within the system could be mapped to these, to more easily measure user behaviour at a finer level of granularity. These are both described in the following subsections.

Specific support for storytelling

One important early finding from the pilot was the need for explicit support in the storytelling part of the process. This was identified by the tendency for the analyst to copy information out of the Cogni-sketch environment at certain points, and then provide additional contextual information outside of the environment, for example as annotations on MS-PowerPoint slides. The question then became how to add this additional information back into the knowledge graph in a convenient form for the analyst. A brief investigation of the situation identified that the analyst had a need to provide additional narrative information for parts of the knowledge graph and these didn't make sense being represented as just additional nodes since they could easily get lost within all the other information. Typically, they were attempting to convey contextual information that spanned multiple nodes and links, sometimes for their own intermediate purpose, and other times to communicate contextual information to others.

This insight and clear need motivated the development of the storytelling plugin to overlay a narrative story across the different portions of the graph. The storytelling capability is delivered as a custom plugin to the core Cogni-sketch environment, specifically to support the development of overlay narratives as part of the sensemaking process. These narratives can be thought of as 'paths across the graph', each of which represents a particular thought, insight or question, and can be reported and tracked by any analyst. In aggregate, these narrative story elements can be grouped together and sequenced to tell a story, representing a narrative arc across the broader knowledge graph and were partially inspired by the data-frame theory of sensemaking [73].

Currently these overlay story narrative elements are simple and contain only a set of nodes and links that represent a relevant subset of the whole knowledge graph along with a textual and/or graphical description provided by the user. These could be further extended to support the different kinds of data-frames envisaged by Klein et al. [73], thereby more deeply embedding the story elements into the sensemaking process itself. For example, by inserting placeholders for information still being sought, or theorised to exist based on expert frame types. Figure 6.4 shows the creation of a story pane within Cogni-sketch, with the drop-down list showing a list of twelve structured analytic techniques [56] that was used to provide simple guidance to story creators, corresponding to the various diagnostic, contrarian and imaginative thinking sensemaking activities that were a useful starting point for creating some structure around story goals[2].

[2]For a video demonstration of this pilot exercise please see video V7 in Appendix A.3.

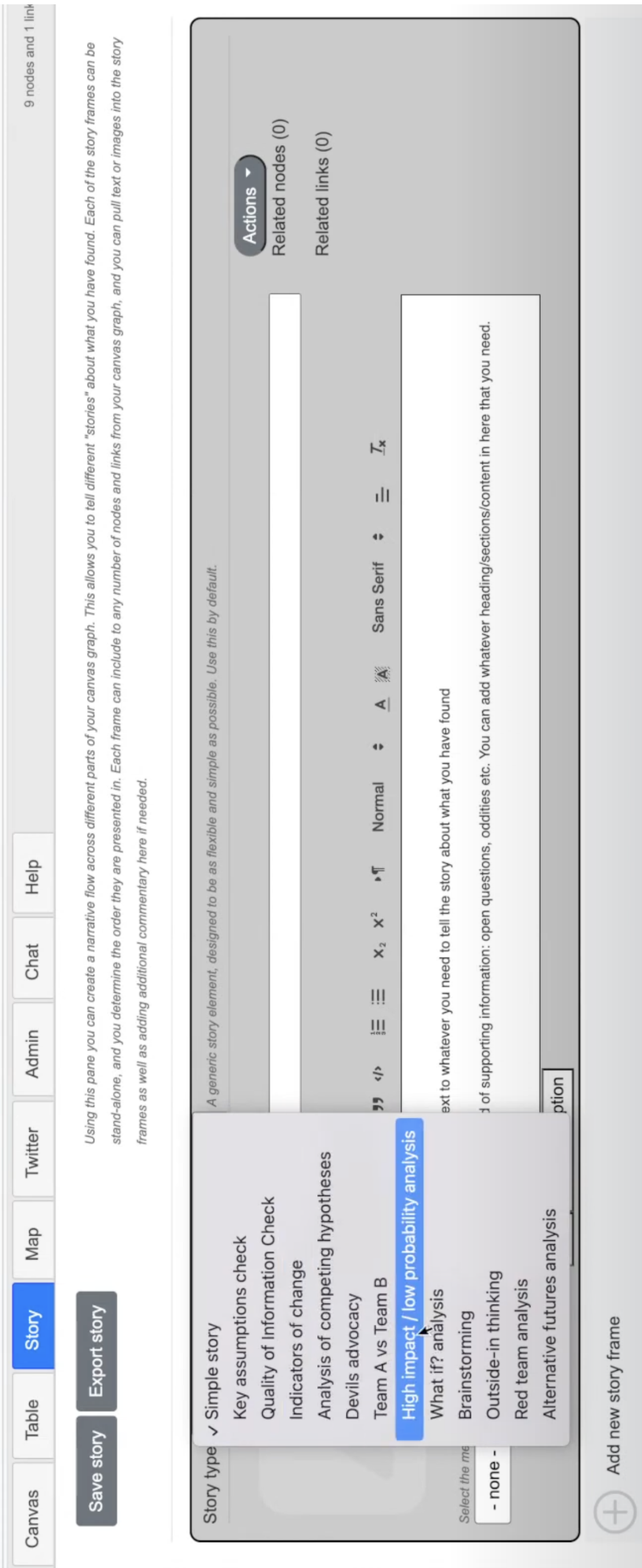


Figure 6.4: Creating a story element to capture narrative flow

The addition of the plugin storytelling pane was completed during the latter part of the pilot, based on feedback from the analyst with an initial basic version delivered during the pilot period, indicating that such functional extensions are credible within the time frame of a typical operation (assuming suitable plugin creators are available).

At the time of development, it was assumed that the storytelling plugin would be specific to sensemaking use cases, but since completion it has been used in other situations, for example see Section A.1.3.

Event instrumentation

The second finding from the pilot related to user behaviour and was anticipated but hard to detect or quantify without direct observation. One example of this was that the schematizing phase identified the need for new raw material in the form of additional text, images or links, but these were collected in a directed manner because of schematization, unlike the earlier open-ended shoeboxing behaviour. In other words, the concurrent activities of shoeboxing and schematising were reported to be occurring and could be proven through analysis of the creation timestamps for nodes and links, but this was a time-consuming process. This observation during the pilot directly informed a detailed event embedding and mapping exercise for the formal experiment, to enable a finer-grained record of user behaviour over time to be easily captured. This was implemented after the pilot, based on an analysis of the analyst activities during the pilot and was used during the subsequent experiment to provided detailed data for the participant actions, along with a mapping of the event types to different aspects of their sensemaking activity, as shown in Figure 6.5.

The Cogni-sketch environment was therefore instrumented with 102 distinct event types which are triggered and logged whenever any participant performs one of the instrumented activities. In addition, the set of interconnected sensemaking loops from Pirolli and Card [113] were condensed into four distinct categories that

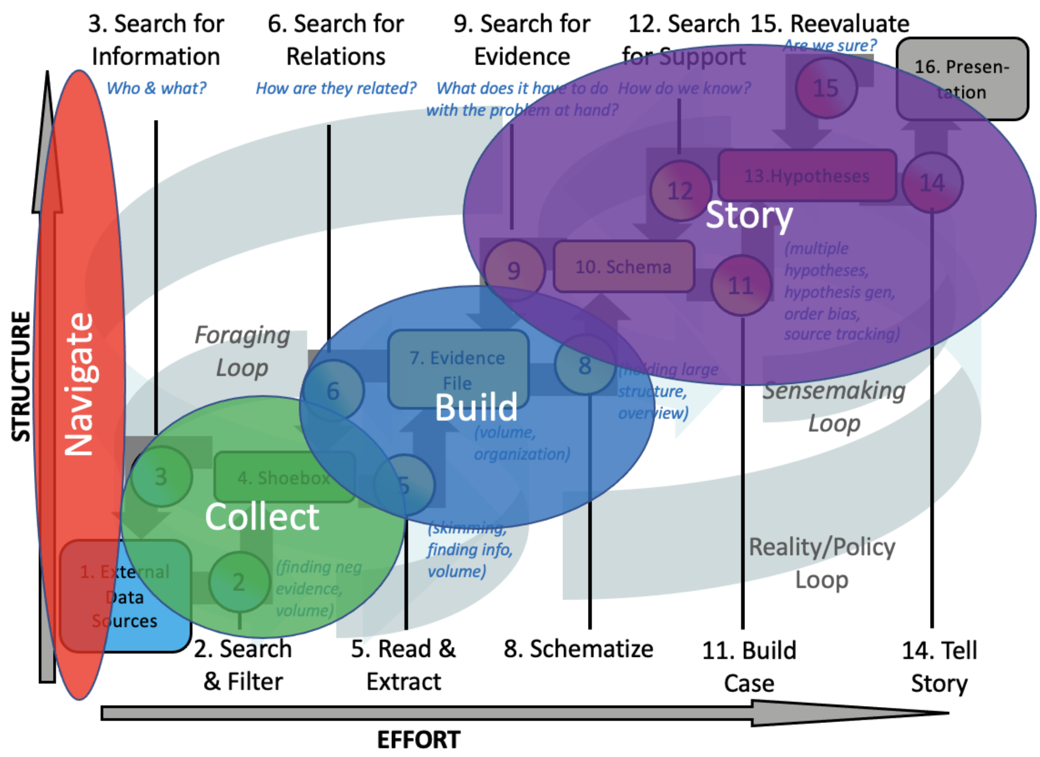


Figure 6.5: Mapping Cogni-sketch behaviour to Pirolli and Card sense-making.

are expressly supported by the Cogni-sketch environment, as shown in Figure 6.5. (Refer to the earlier Figure 6.1 for the original sensemaking diagram from Pirolli and Card [113]). The 102 event types were mapped to these four sensemaking categories.

In Figure 6.5 each of the four coloured overlays shows a different class of behaviour that can be detected within the Cogni-sketch environment through the event instrumentation. It is important to note that some of the foraging activity can happen outside the tool and is therefore not directly detectable since the instrumentation is focused on the participant activity only within the Cogni-sketch environment. For example, a user seeking information in their browser would be invisible until the user pastes the result into Cogni-sketch, and if nothing is found during external foraging and therefore nothing is added to the Cogni-sketch environment then no activity will be logged.

A substantial volume of user activity is represented in Figure 6.5 as *navigate* activity. This corresponds to the user navigation of the Cogni-sketch environment through interaction with nodes, links, panes and other resources, and whilst it cannot be consistently categorised into the *collect*, *build* or *story* classes it reflects the user building their understanding through exploration and usage. It is therefore mapped to the Pirolli and Card sensemaking process as a thin vertical area aligned with the *structure* axis to indicate that it can span any of the activities and represents low effort by the user. Navigation activity can then be separated in any subsequent analysis of user activity because it cannot be reliably tagged to any specific sensemaking activity.

Some event types are explicitly ignored, e.g., for certain highly repetitive activities such as panning/zooming. The mapping of events to the four sensemaking categories is shown in Table 6.2.

For a full list of the Cogni-sketch event types and their mapping to these sensemaking categories refer to the table in Appendix B.2. The pilot inspired the formal sensemaking experiment described in the following section, with the various upgrades for storytelling and event logging being completed before the experiment started.

6.4 Open source sensemaking experiment

Building on the results obtained from the successful pilot exercise, this section provides details for a formal experiment [1] into the usability of the Cogni-sketch environment when applied to OSINT analysis and sensemaking. The experiment is based on the exploration of a set of tweets from verified Twitter users relating to mask wearing during a 2-week period in the summer of 2021 whilst Covid-19 restrictions were in force.

Category	Description	# event types
ignore	Events that are captured but ignored as not relevant (e.g., panning or zooming).	9
navigate	Events relating to navigation of the environment.	10
collect	Events relating to data collection, as mapped broadly to the Pirolli & Card foraging region.	43
build	Events relating to knowledge building, as mapped broadly to the centre of the Pirolli & Card process.	34
story	Events relating to the construction of stories, as mapped to the Pirolli & Card sensemaking region.	6
total	All possible event types.	102

Table 6.2: Mapping user behaviour and event types to sensemaking categories.

6.4.1 Experiment objectives

There are two hypotheses for the experiment, arising directly from RQ1 (human creativity) and RQ3 (sensemaking support) for this research thesis. Whilst directed machine agents were present within the system to assist with search, exploration and sentiment analysis of the social media data, they are not the focus of the experiment, so *RQ2* is not stated to be directly relevant here.

H1 That untrained human participants can use the Cogni-sketch environment to create knowledge into the graph from a variety of sources and show human creativity *(RQ1)* in the visual and structural styles used.

H2 That the process of sensemaking[3] can be supported, with task-relevant information and knowledge being curated according to the foraging, schematising

[3]Specifically sensemaking as mapped to the Pirolli and Card sensemaking process [113].

and storytelling aspects of the sensemaking process *(RQ3)*, and explicitly that these can occur in any order, not being limited to flow only one way from foraging to storytelling.

These hypotheses have informed the design of the experiment that is described in the remainder of this chapter. The experiment used a large and messy real-world dataset, and with deliberately open-ended sensemaking questions posed to the participants.

6.4.2 Experiment method

Having chosen the domain of OSINT analysis as the target for the human-led experiment it was clear that a specific small set of target questions would be needed, alongside an easily accessible core dataset and a time-bounded period for the exercise. In the earlier pilot exercise, it was clear that this very open-ended, multi-month exercise would not be suitable for broader participation, both in terms of participant time and for meaningful analysis of the results across such an extended period. The target community of intelligence analysts is also small, and finding available time from a cohort of analysts for this exercise would be difficult. Therefore, the experiment was designed for novice users with little experience of intelligence analysis and would measure their ability to perform a sensemaking task based on instinct and a brief introduction, alongside a small amount of formal support within the environment.

Whilst Cogni-sketch is designed to be the integration point for multiple external data-sources in a variety of data formats it was clear that only using external resources would be too open-ended for this exercise, and would also have the side-effect of much of the foraging behaviour taking place outside of the Cogni-sketch environment and therefore being unavailable for post-experiment analysis since it cannot be easily instrumented without intrusive intervention in the user environment. It was therefore decided to collect a suitable rich and messy social media dataset as the basis for the experiment and use the plugin architecture for

Cogni-sketch to make this available as a specialised navigable set of data within the environment.

It was important that the data collection and processing did not take a large amount of development time but yielded a rich and varied dataset for the experiment. Also, that the development of the specialised data processing plugin was not a substantial and time-consuming development exercise either. This would be a useful informal assessment of the flexibility of Cogni-sketch to achieve this kind of capability without too much effort for a plugin creator user (the author of this thesis).

Based on operational experience and previous successful exercises (e.g., [119]) it was decided to collect a substantial set of Twitter data on a specific topic for a short period of time, and to develop a new 'Twitter data explorer' capability as a plugin pane to allow participants to explore, sort and summarise the available Twitter data directly within the Cogni-sketch environment as part of this experiment. This custom plugin pane is shown in Figure 6.6 with a variety of views and filters to explore the Twitter data as well as saving queries to the canvas (via the options menu).

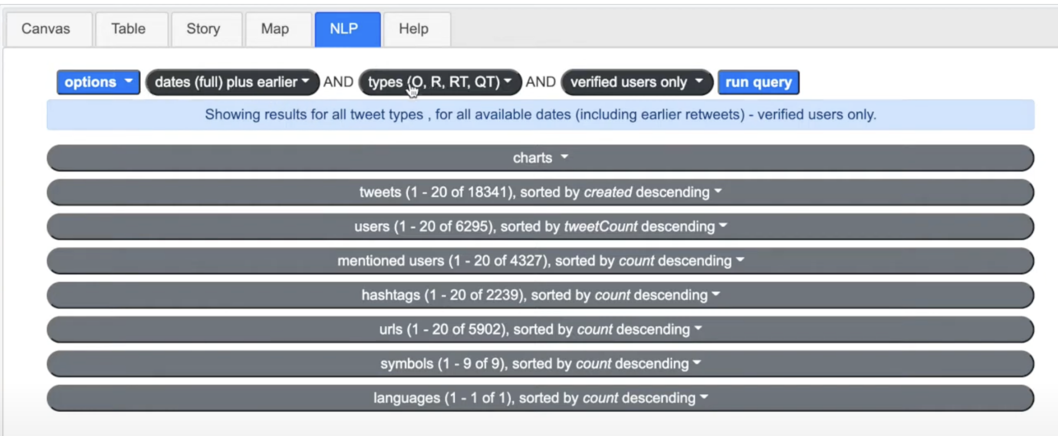


Figure 6.6: A custom pane to explore and query Twitter data

Dataset: Selection and characteristics

A specific but deliberately unoffensive, topic was desired for the Twitter data collection. This would support the answering of subsequent questions to be pursued by the experiment participants. It was mid 2021 when this experiment was being designed, and the U.K. was starting to exit a substantial period of social restrictions relating to Covid-19. To keep the focus away from explicit politics and obvious mis- or dis-information the dataset collection was chosen to be tweets related to the term ‘mask’ from 3rd July to 30th July 2021, a period of 27 days. During this period was Monday 19th July 2021 which was referred to as ‘Freedom Day’, headlined as a removal of all Covid-19 related restrictions in the U.K., but in reality “a move from a rules-based approach to one that manages risk in a more holistic way through our everyday behaviour” ([160]), and therefore placing the onus more on individual responsibility and behaviour. For this OSINT analysis experiment it gave an excellent opportunity for a clear dataset over a specific short period, and a range of possible target questions, all based on real data and real behaviour, and therefore indicative of a typical setting for use of a tool such as Cogni-sketch for sensemaking.

The data collection was run using existing Twitter data collection facilities at Cardiff University, yielding 1.1 million mask related tweets over the 27-day period. Of these, only 169,000 (15.4%) were *original* tweets written for the first time, while 235,000 (21.4%) were replies to other tweets, with the 712,000 majority (64.7%) being retweets, i.e. Twitter users amplifying an original tweet by re-sending it to their followers. For this exercise the retweets themselves were removed because they were duplicates, but the popularity statistic of the original tweet was unaffected. This left 412,000 tweets over the 27-day period. Within this refined subset of tweets there were 21,300 unique hashtags, 83,500 unique Twitter user mentions, and 106,000 separate Twitter user accounts that created the 412,000 tweets.

It was important for the sensemaking exercise to be realistic and that the participants operate in a real-world information environment that is known to be messy, inconsistent and vague, and that there are few definitive and correct answers. Therefore, by design, the questions they were tasked to answer did not necessarily have an explicitly right or wrong answer. The purpose of using Cogni-sketch as a flexible knowledge and information co-construction environment for OSINT analysis is to allow each participant to construct their own answers with evidence, their opinion and any other corroborating sources to explain their answers. However, it was useful to undertake some basic exploration of the data ahead of the exercise to ensure that any obvious issues or key features could be communicated to all participants. There was no attempt to remove any tweets by their content, for example known misinformation, offensive content, or any other features.

The task was aimed to be as realistic as possible and include the usual noise and other issues that come with typical open sources such as this. The selection of 'mask' as the core focus of the collection was deliberately vague and ambiguous to ensure the collection had these common issues. It was clear from a pre-experiment analysis of hashtags that there were a small number of very dominant hashtags and then a long tail of less frequently referenced ones. It was also clear that some hashtags suggested a 'pro mask' stance within the community (e.g., #WearAMask and #MaskUp) whereas others suggested an anti-mask stance (e.g., #TakeOffYourMask and #NoMasks) and many other hashtags were ambiguous or used by both stances. There were also several hashtags that showed some of the collected tweets were off topic (as planned and adding useful noise to the dataset). None of this was revealed to the participants, but the Twitter data explorer pane (shown in Figure 6.6) within the Cogni-sketch environment provided easy access to hashtags, Twitter usernames and other features such as number of replies or retweets that enable any participant to easily explore the data and draw their own conclusions, ideally capturing their results as new knowledge onto their Cogni-sketch canvas.

For the experiment the full set of tweets were reduced to original tweets from verified Twitter users, and all replies to these (from both verified and unverified Twitter users). Verified Twitter users at the time were public figures or well-known brands who had been validated by Twitter as officially representing that person or organisation.

Textual data support

Given the textual nature of core Twitter data it was also important to introduce some language specific capabilities that the participants may find useful in undertaking the experiment. Within the collected set of mask tweets there were many with embedded media, external hyperlinks and other non-textual content, all of which are potentially relevant and valuable to the human users, but no specific additional processing for these was included within the Cogni-sketch environment beyond the default support for a range of media types to be directly embedded and playable/renderable within the Cogni-sketch environment as part of the knowledge graph.

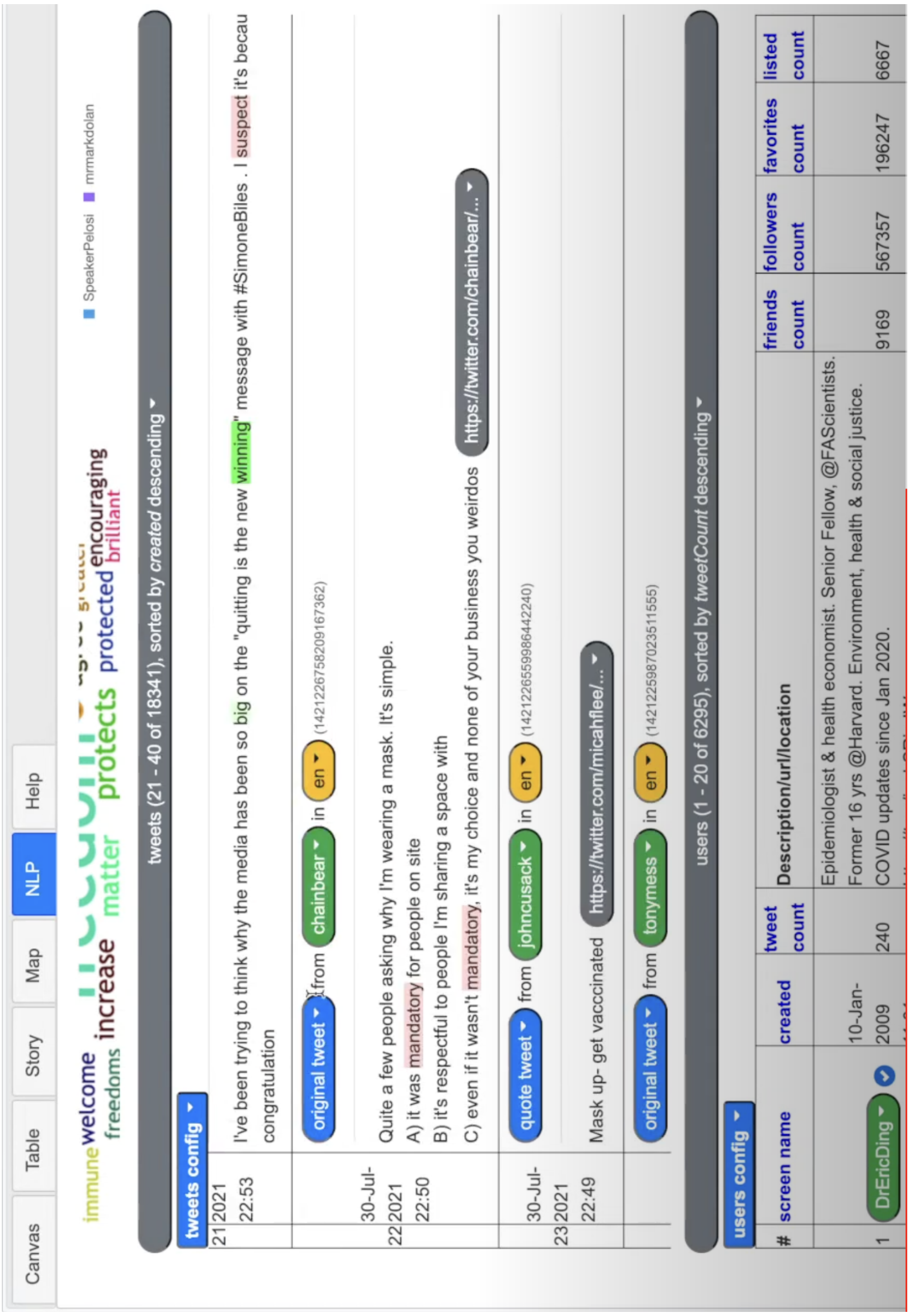


Figure 6.7: Interactive tweet explorer

The ability to navigate within the environment, from tweets to users, hashtags and other relevant features was important with an example of this navigable interface shown in Figure 6.7. This was both to provide a fluid and flexible directly embedded user experience but also to ensure that specific events for the browsing of these could be generated and logged for subsequent analysis. Otherwise, such navigation would happen outside the system, for example inside the general Twitter UI and the events would not be captured.

Sentiment analysis and Emojis

Three separate JavaScript-based sentiment classifiers were investigated, measuring the overall performance in terms of execution time and the range of sentiment scores for the tweets. Two were based on AFINN-165[4] and one on AFIN-111[5], and all had support for emoji sentiment analysis. Based on a detailed assessment, the wink-sentiment library was chosen to be used as part of the data analysis pipeline, meaning that each tweet available in the dataset had a sentiment score computed using this library, and the Cogni-sketch Twitter data plugin was able to search and sort, based on this sentiment score. Additionally, sentiment scores above 0 were classified as *positive* tweets, whilst sentiment scores of 0 or below were classified as *negative*, and these positive/negative categories were explicitly labelled within the environment to support optional filtering by the participants.

It should be noted that for this mask dataset the emoji for mask (😷) was treated as negative sentiment as in a typical setting a mask emoji would normally indicate some kind of negativity such as illness, but for this data collection there were cases when the mask emoji was used in both positive and negative (and ambiguous) contexts. However, this kind of issue is not unusual with sentiment analysis when applied in a particular context, for example sporting terms like: ‘smash’, ‘lob’, ‘attack’ having an often less-negative meaning than in everyday settings and requires a contextual understanding based on the task.

[4]See `https://www.npmjs.com/package/afinn-165`.
[5]See `https://www.npmjs.com/package/afinn-111`.

Word clouds

The other language specific capability that was added was the ability to generate real-time world clouds based on a subset of filtered tweets. This could additionally be filtered by positive or negative sentiment as shown in Figure 6.8. So, for example a participant could filter based on specific terms, hashtags, Twitter users and/or dates, and then have the content of all the matched tweets summarised into a single word cloud and choose positive or negative sentiment, or both. The ability to generate real-time bar charts for activity over time was also available (as shown in Figure 6.9) with the participant able to choose this based on various filters within the interface.

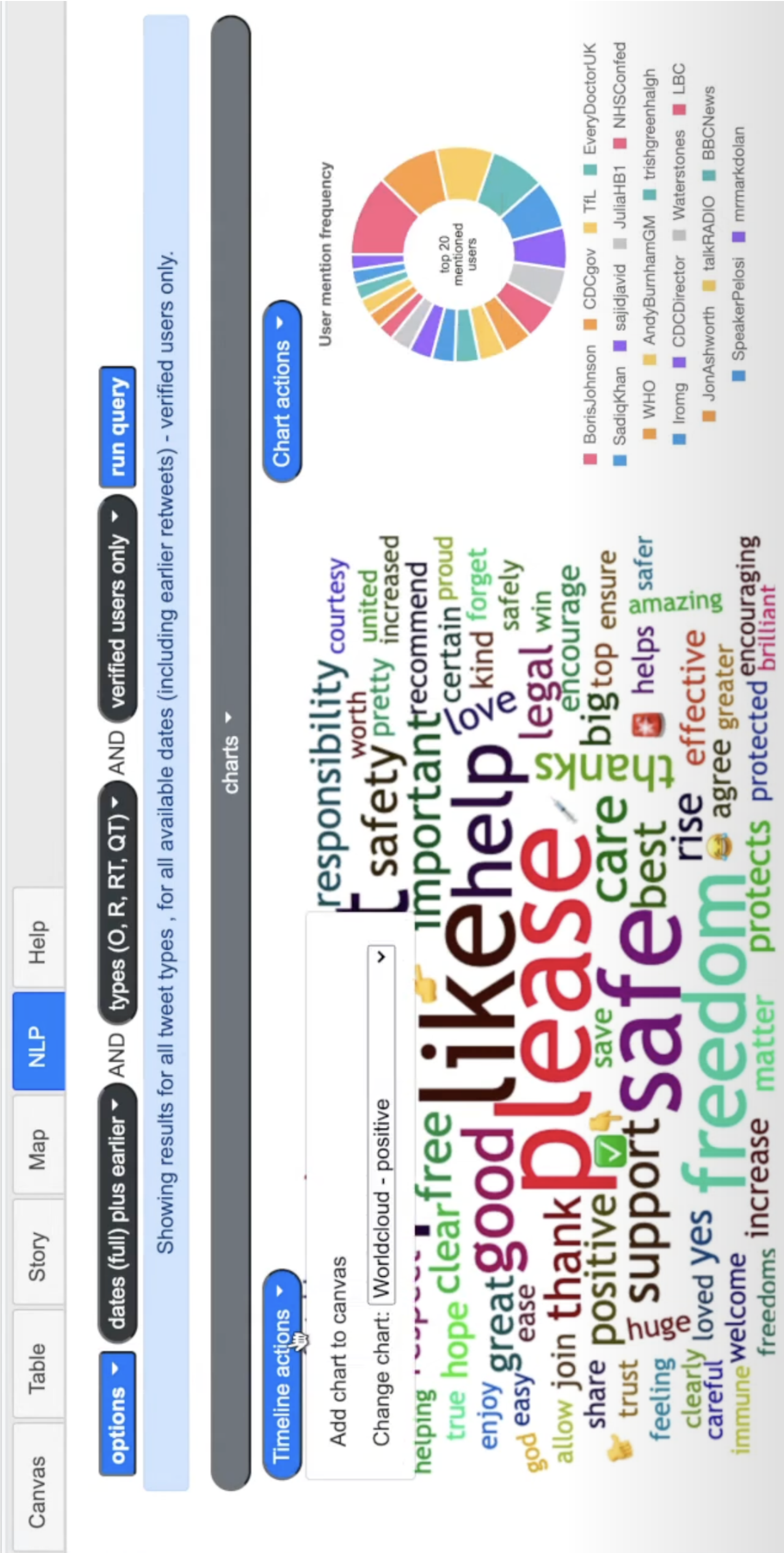


Figure 6.8: A dynamic word cloud showing positive sentiment

All these charts and word clouds, as well as the underlying filters and queries that drive them can be easily added to the canvas for the participant, enabling them to build their evidence base as part of the sensemaking exercise. They can also be used in simple combinations, for example by combining a keyword search with a filter on users, and a sort based on number of mentions.

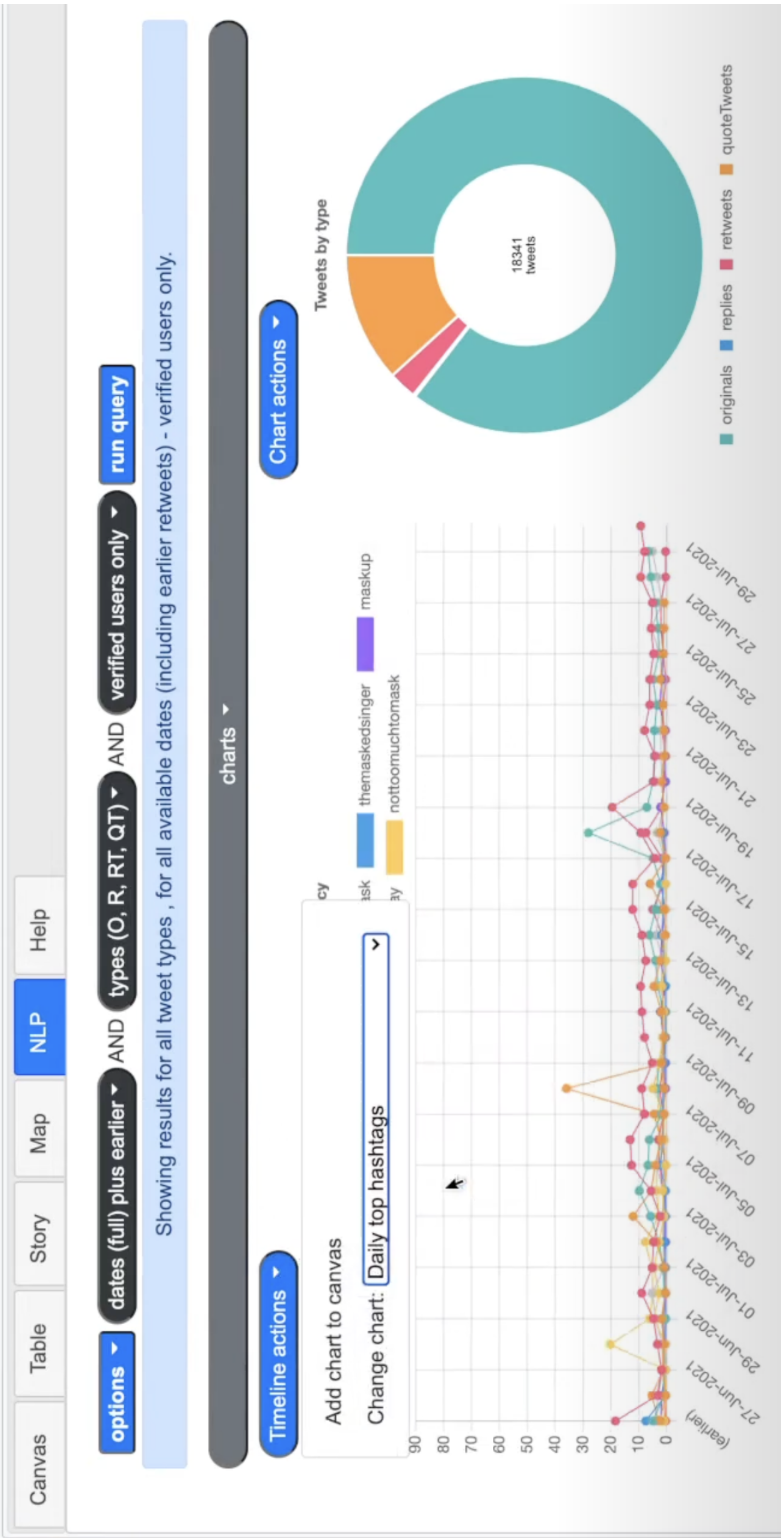


Figure 6.9: Filtering tweets and generating dynamic charts

The design and implementation of all these custom navigable views was able to be achieved within the plugin architecture of the core Cogni-sketch platform, with each of these being implemented as custom panes that invoked specific simple directed machine agent processes to search, filter or query the dataset and render the results. All these capabilities were essential to provide an instrumented set of simple capabilities to allow the experiment participants to carry out the foraging aspects of the exercise, and for this to be invoked within the Cogni-sketch environment to allow event generation and logging for participant behaviour analysis.

Participant guidance

At the start of the 2-hour experiment period participants were provided with some high-level scene setting to help them understand the context and goals of the exercise. This included a simple temporal analysis of the tweet data volumes over the 27-day period as shown in Figure 6.10.

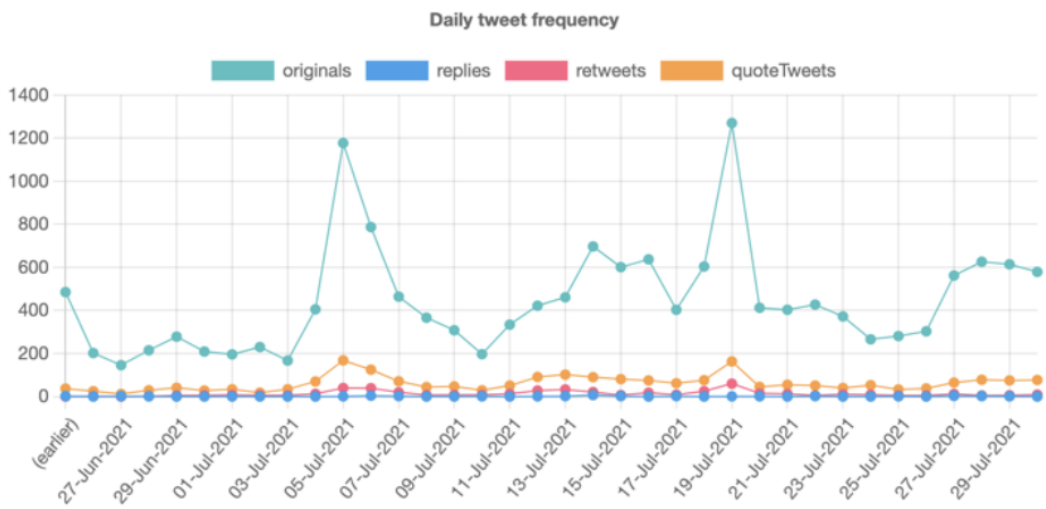


Figure 6.10: Selected tweet activity over time

The reduction of the full set of tweets to only those from verified users, alongside the removal of retweets means that the number of original tweets is the largest portion of the data shown in the graph. This is essential for ensuring the participants get a high volume of original material to review, and the retweet count for each tweet is preserved in the data available to them, even though the actual retweets are removed.

The graph in Figure 6.10 shows an interesting profile of overall activity, with Freedom Day (19th July) clearly shown as a peak in overall tweet volume, as might be expected. However, there is also a second earlier peak, and this simple temporal graph led to the two questions for the participants in the experiment. These were:

Q1: What caused the spike in volume for mask related discussion in the U.K. on social media on Monday 5th July?

Q2: How do the factors driving this first spike feature throughout the time period?

These questions were explicitly designed to be a combination of a well-bounded first question with the potential for a succinct and definite answer, with the second question being more subjective and open-ended and likely to be driven more by opinion. The goal of the exercise was to test the participants ability to use the Cogni-sketch environment for sensemaking, and to create and link knowledge rather than considering whether the answers they gave were right or wrong.

Refer to Appendix B.1 for a full copy of the two-page guide that each participant was given at the beginning of the experiment, including the timeline chart shown in Figure 6.10, and the two questions listed above.

Experiment design

All participants in the experiment were novice users of the environment, having never directly used Cogni-sketch before, and this inexperience is an important and deliberate aspect of the study. The typical human operator user that is the target of the Cogni-sketch environment is a non-technical domain expert (i.e., not a computer programmer or ontology expert, but someone who is an expert in the field they are using Cogni-sketch for). There were twelve anonymous participants in total, all of whom worked individually in their own Cogni-sketch environment and were able to create as many projects within this environment as they needed. Most (but not all) participants chose to stick with a single project environment, incrementally building knowledge as their investigation progressed.

The participants were sourced from a staff and students at IBM Research U.K. and the CSRI at Cardiff University. Whilst no participant had direct experience of the Cogni-sketch environment, there were a range of experience levels for OSINT analysis and sensemaking, with most participants having little-to-no experience of these areas, but three participants having some limited experience of OSINT analysis working alongside more experienced colleagues, capturing results using office tools such as MS-Word and MS-PowerPoint. No participant was an expert analyst.

The experiment was carried out as a series of 2-hour sessions between January and March 2022, with one or more participants active in each of these sessions. Each participant operated in their own separate environment and in only one 2-hour period. No collaboration between participants was enabled. The grouping of participants into specific sessions was purely to minimise the workload of the experiment organiser by enabling the support of multiple concurrent participants in a smaller set of total sessions. Prior to the start of the experiment none of the participants were aware of the task that they would be asked to complete.

In addition to standard ethics processes each participant was asked to confirm their availability for a particular 2-hour session and, at the start of the session, was provided a 2-page overview of the experiment and summary of the basic environment capabilities (See Appendix B.1 for a copy of this briefing material). This material was also created on the canvas for each of the participants as their starting point, providing an in-environment 'guidance graph' to ensure they have easy access to all relevant guidance material directly within the environment.

This overview also contained a link to a 6-minute YouTube video[6] outlining the basic capabilities of the environment, and each participant was sent their username and password for Cogni-sketch via Slack at the beginning of the session. These steps were to ensure that no participants could get advanced knowledge or experience of the environment or tasks ahead of the 2-hour experiment period, and that all activity for the experiment could be recorded directly within the 2-hour window. As a result, the first 10 minutes of the experiment were typically spent reading the material and watching the video, leading to a lower number of events for most participants in the first 10 minutes.

Because of Covid-19 restrictions the experiment subjects generally participated remotely from their homes. Support was offered for each participant via Slack (for text messages) and Webex (for video conferencing if needed) and any use of Slack or Webex was 1:1 with the experiment organiser, to prevent any cross-contamination of results by other participants observing the support being given to others. Some participants made use of the Slack environment to ask questions about functions and features within the Cogni-sketch environment, and no participants used the Webex video conferencing support. Overall, the need for support during the experiment was very low, and zero for many of the participants.

6.4.3 Experiment results

Based on the findings from the earlier pilot, data were collected using three separate techniques to support more advanced and sophisticated analyses after the evaluation of the Cogni-sketch platform. Substantial quantitative data were captured for each user based on their activities in using the platform during the evaluation. This event data was dynamically generated as a sequence of JSON event objects that were logged to the server for each user of the system for the duration of the exercise. These were processed after the experiment conclusion, with

[6] See `https://youtu.be/GsNq0EBpimU`.

analysis providing insight into the behaviour of each user through their generated activity log. These events were mapped to a sensemaking model (as described in Section 6.3.4), and aligned in time across the various evaluation sessions that were undertaken. Additionally, a usability survey was completed by each participant, providing their feedback on the experience. Finally a qualitative analysis of the created artefacts was undertaken for each participant. Each of these analyses are described in detail in the remainder of this section before discussion of the results and a comparison back to the findings from the earlier pilot.

All participants were able to use the Cogni-sketch environment successfully, and most participated for the majority of the 2-hour time window, with some completing the task early, with their activity duration being clearly identified within the collected data. Similarly, most participants took one or more short breaks during the exercise as advised by the participation guidelines. The data processing method for activity analysis consolidates data into 10-minute periods for consistent granularity of analysis, and to avoid the perception of any continuous performance monitoring of the participants.

Since this was a baseline experiment rather than a comparative exercise against an existing benchmark it was not possible to directly measure or quantify any *performance improvement* as captured within the original HAKF definition. However, the SUS usability exercise was able to give some insight into human user confidence in the system, and subsequent experiments could be designed to attempt to measure performance to detect any improvements (or otherwise) once suitable development of use case and machine agent support is achieved.

The results of this experiment were gathered in three ways, each of which is described in the following sub-sections, but summarised briefly here:

- First there is a detailed time-based quantitative analysis of **participant activity** based on fine-grained **event monitoring** within the Cogni-sketch environment. This enables a detailed analysis of the temporal activity for each participant, as well as for average participant behaviour. This data is consistently available for the duration of the exercise and can be broken down in various ways, primarily by mapping each event to one of the four sensemaking categories.

- Second is a **qualitative analysis of the sensemaking artefacts created** during the exercise by each of the participants, their ability to use different parts of the system, and whether they were able to construct stories as well as raw knowledge graph contents. See Appendix B.3 for a detailed summary of the artefacts created by each of the participants with corresponding screenshots of their canvas and story artefacts, as well as a detailed qualitative analysis for each participant. One example of this analysis is also included later in this section.

- Third is a **standardised assessment of general usability** using the SUS [30] which is an established industry-wide technique for assessing usability immediately after participants have been first exposed to a new system.

Quantitative results: participant behaviour

This analysis is for the sensemaking behaviour of the participants from the experiment and is specifically a statistical analysis of the detailed event data generated through instrumentation of the Cogni-sketch environment across the four main types of sensemaking behaviour shown earlier in Figure 6.5. The three sets of results that underpin the analysis in this section are shown below with a brief description for each.

1. Participant activity statistics

The participant activity statistics are shown in Table 6.3 - this provides summary data for each of the participants across the whole 2-hour duration of the experiment. This includes the number of events generated by each participant, the ratio of events in each of the four categories (and the ignored events) as well as the number of projects, nodes, links and stories created by each participant (with total, maximum, minimum and average values across all participants).

User	Projects	Nodes	Links	Stories	Actions					
					Total	ignore	nav	collect	build	story
1	2	22	10	2	192	24%	30%	39%	3%	4%
2	1	11	10	1	513	18%	15%	42%	25%	2%
3	1	33	33	3	592	16%	12%	20%	49%	3%
4	2	18	11	5	471	26%	20%	21%	28%	6%
5	2	30	28	3	524	18%	11%	23%	46%	2%
6	1	37	26	5	716	13%	14%	32%	39%	3%
7	1	17	0	2	545	9%	18%	59%	11%	3%
8	1	12	14	6	299	19%	20%	27%	27%	6%
9	2	35	30	5	1009	14%	16%	40%	28%	3%
10	1	26	7	2	682	18%	14%	45%	20%	3%
11	2	25	20	1	1050	19%	14%	39%	26%	1%
12	1	56	13	4	982	22%	13%	31%	32%	2%
Total	**17**	**322**	**202**	**39**	**7575**	**1326**	**146**	**2676**	**2219**	**208**
Min	**2**	**11**	**0**	**1**	**192**	**9%**	**11%**	**20%**	**3%**	**1%**
Max	**1**	**56**	**33**	**6**	**1050**	**26%**	**30%**	**59%**	**49%**	**6%**
Avg	**1.4**	**27**	**17**	**3**	**631**	**18%**	**16%**	**35%**	**28%**	**3%**

Table 6.3: Experiment participant activity statistics

2. Aggregate participant activity

Figures 6.11 and 6.12 show the average activity for all participants broken down into 10-minute periods throughout the duration of the 2-hour experiment. The first figure shows the absolute number of average events, whereas the second figure shows the ratio of the event total against all events. Both figures show data for each of the four sensemaking categories using a common marker style and colour between the figures.

3. Individual participant activity

Figures 6.13 and 6.14 show a set of twelve charts each; there is one chart per participant. Both figures show the individual participant activity for each 10-minute period of the 2-hour exercise. Figure 6.13 shows the cumulative total of all activities for each participant, whereas Figure 6.14 shows the activity for each

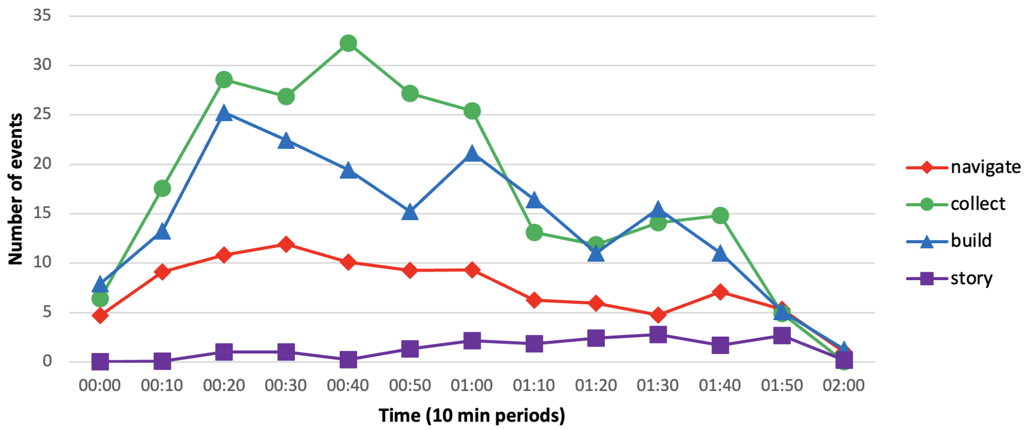


Figure 6.11: Average number of events, by category (10 min periods)

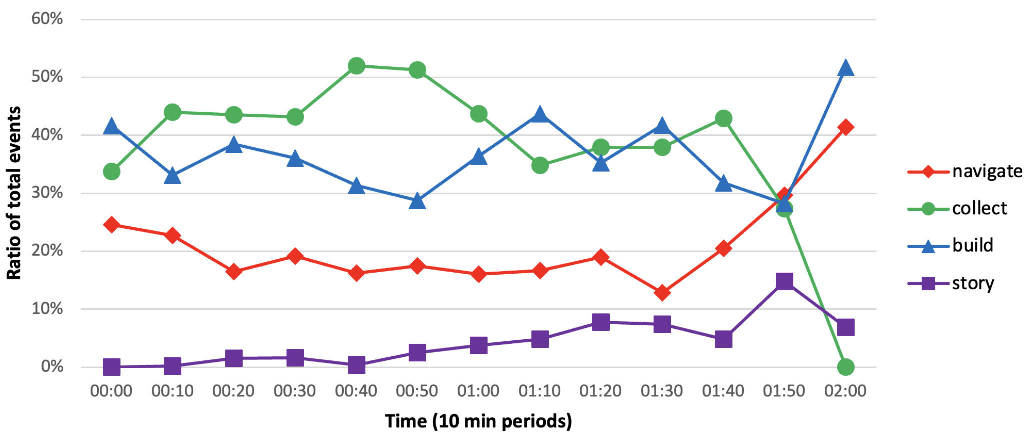


Figure 6.12: Average ratio of events, by category (10 min periods)

participant broken down by the sensemaking category (navigate, collect, build, story). The vertical axis is standardised across the two groups of charts to enable easier comparison (200 events per period for Figure 6.13, and 100 events per period for 6.14).

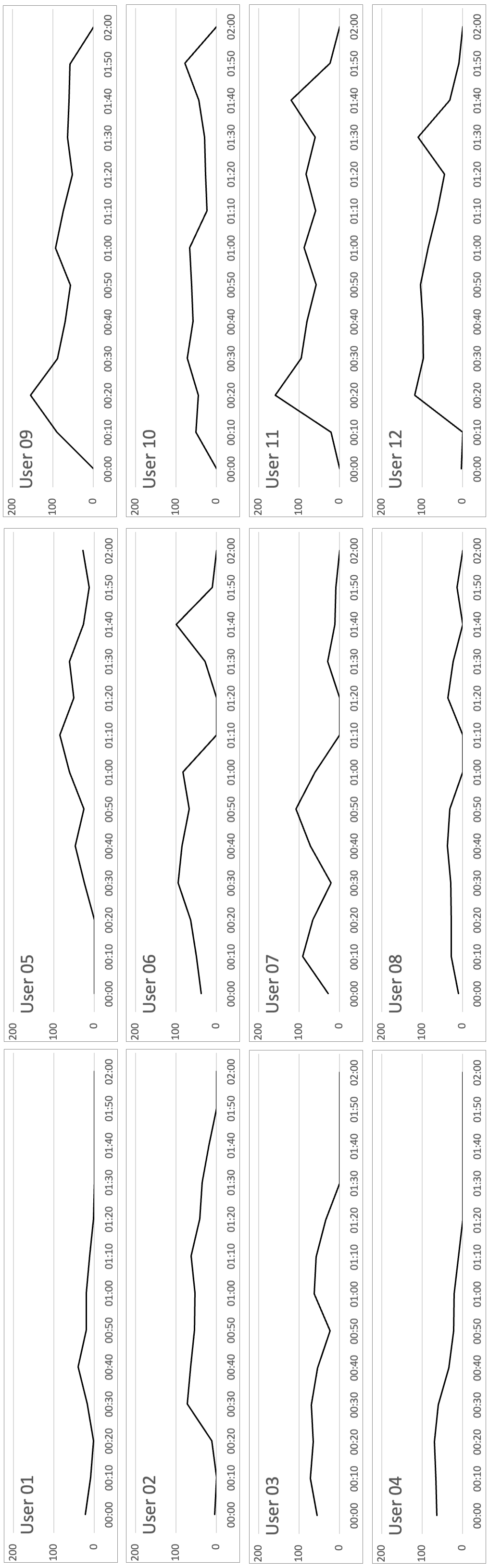


Figure 6.13: Total aggregate participant activity (10 min periods).

Y-axis is from 0-200 activity events in a period.

X-axis is each 10-minute period of the 2-hour exercise period.

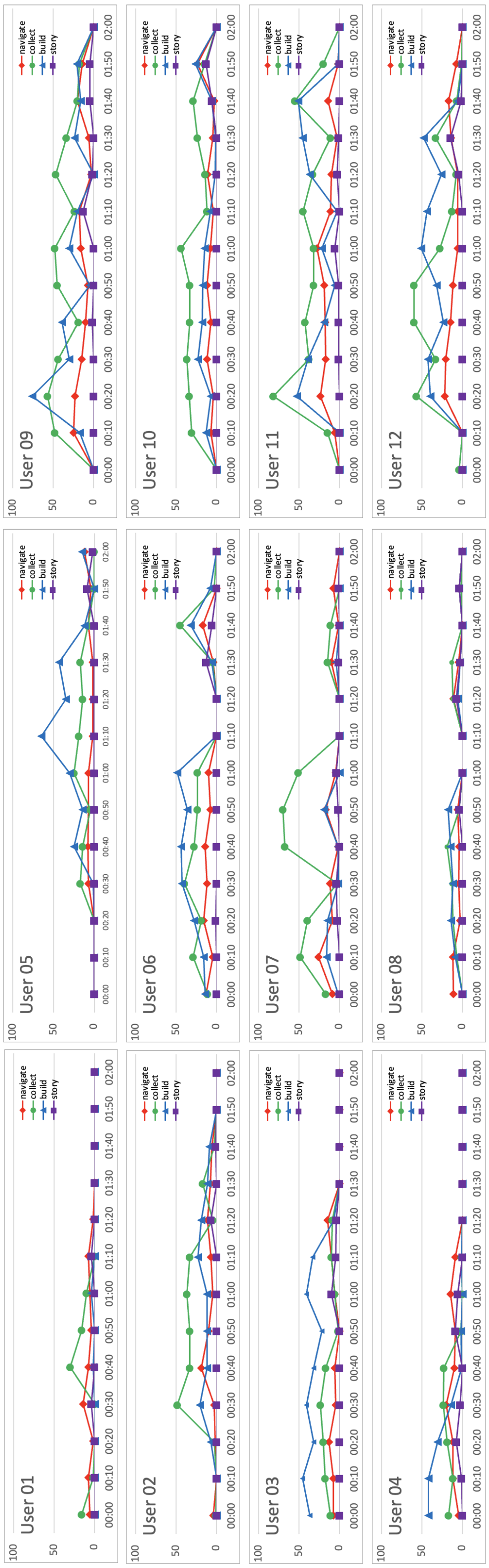


Figure 6.14: Total participant activity by type (10 min periods).

Y-axis is from 0-100 activity events in a period.

X-axis is each 10-minute period of the 2-hour exercise period.

Of the 7,575 events recorded across all participants: 18% (1,326) were ignored (as they were simple pan/zoom type events) and 15% (1,146) related to navigation of the Cogni-sketch environment. However, 35% (2,676) of all registered events related to the collect category, at the foraging end of the sensemaking loops, with 29% (2,219) events relating to the build category and representing the higher level of sensemaking, with just 3% (208) relating to the story category, and therefore the storytelling part of the process.

Analysing the data shown in these tables and graphs these basic observations emerge:

1. The twelve participants (identified as user 01 - user 12) were all able to interact with the environment for the duration of the experiment, with three participants finishing between 10 and 30 minutes early (users 01, 03 and 04).

2. There is no direct correlation between amount of activity and artefacts created. For example, for nodes, the average number of events per node created ranges from 9 to 46, and for links from 17 to 97 across participants. For stories the ratio is much greater (as expected due to the small number of stories relative to nodes and links), ranging from 94 to 1050 events per story element.

3. There is a wide range of activity difference across the participants, with one participant registering only 192 events within the 2-hour period (that participant finished at least 30 minutes early), whilst another participant registered 1,050 events. The average number of events was 631.

4. Whilst storytelling receives only a small percentage of activity (3%) the average number of stories created across all participants is three, with one participant creating six stories for their exercise. There are 39 story elements created across all participants, but only 208 events generated because the creation of stories is simple once the rest of the knowledge graph has been

built. Therefore, the volume of story events must not be conflated with the relative density and potential value of story elements.

5. Most participants continued to *navigate*, *collect* and *build* after they had created their first story, showing the desired non-linearity of the process. However, it is also clear that *story* activity increased towards the end of the experiment period.

6. In general, activity across the four categories increases rapidly from the start of the experiment, peaking at around 40 minutes (00:40), and then declining at 1 hour 10 minutes (01:10) before declining again just before the end of the experiment after 2 hours (See Figure 6.11).

7. As the experiment progresses the amount of storytelling increases (albeit at a low level of overall activity compared to the other categories) whereas all other categories decline (See Figure 6.11).

8. When considering relative activity there is a more stable profile across activity categories, without the general decline in activity observed towards the end of the experiment. However, the last two time-periods show varying average ratios as the overall volume of activity declines. It seems that some participants had completed the experiment and were exploring the system more generally, hence the relative rise in navigate activity in the last 30 minutes (See Figure 6.12).

9. The increase in story activity is more pronounced when looking at relative behaviour, with an overall rise from around 00:40 to 01:50 (See Figure 6.12).

Qualitative analysis: sensemaking artefacts

The qualitative analysis of the results is based on a post-hoc manual review and analysis of the artefacts created by each participant during the experiment. The full details for all twelve of these analyses alongside a copy of the Cogni-sketch canvas and story artefacts created can be found in Appendix B.3 with a short overall summary given here. Since these details are necessarily located in the appendix due to the length and level of detail, a single example of these three resources for a single participant are copied into the main body of the thesis here. The purpose is to ensure the reader has an example of these materials and can optionally refer to the appendix for all 12 of the participant details as needed.

Example outputs and qualitative assessment for an example participant (participant 05)[7]

[7]This participant was chosen because their canvas and story artefacts are a typical example of those created generally by participants within the experiment, but also because participant 05 rated their experience the lowest in the SUS usability assessment and therefore is chosen as a more critical user than average. For access to a high-resolution copy of the canvas and story images for this participant please refer to Appendix A.2.

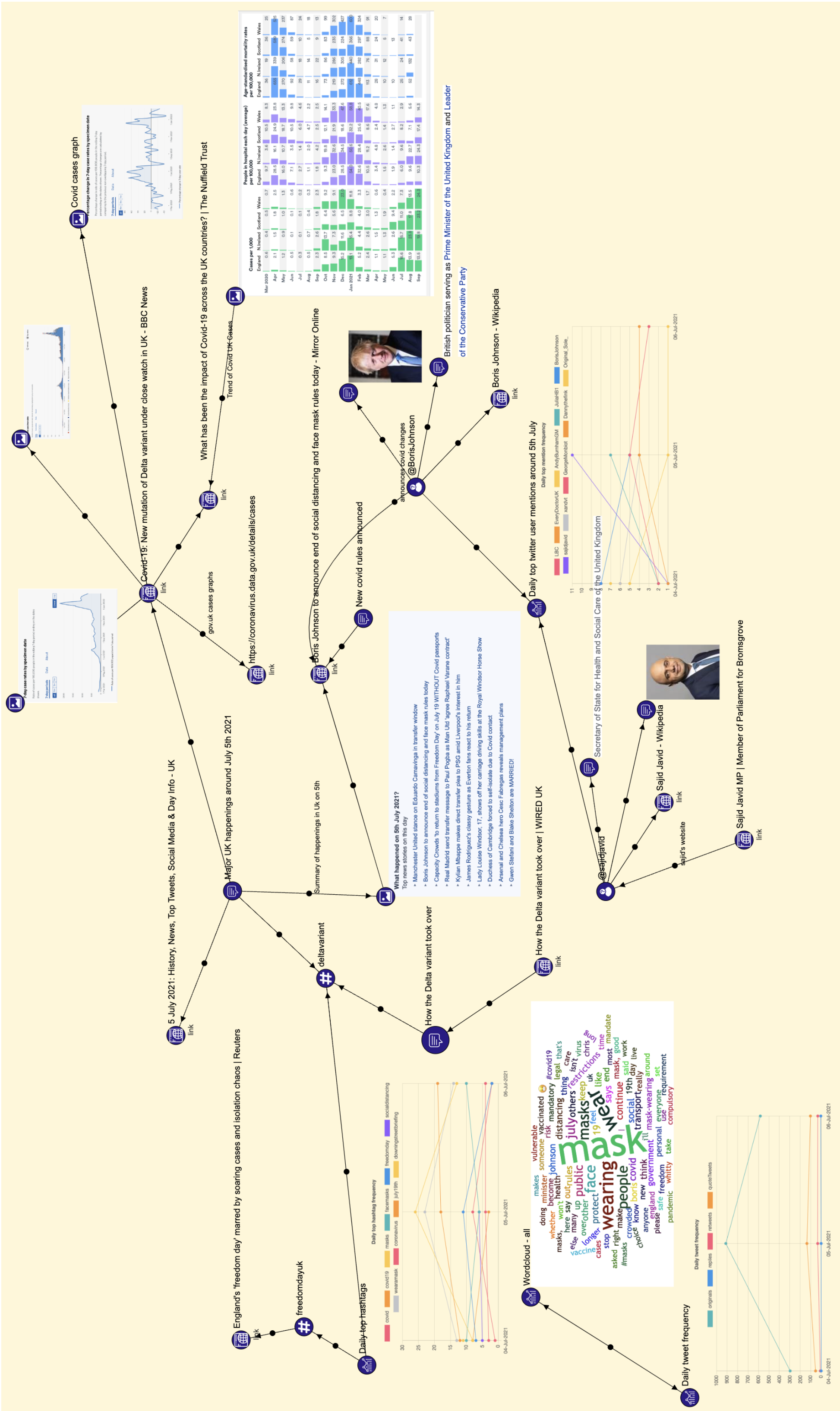


Figure 6.15: Canvas for example participant

Figure 6.16: Story for example participant

Qualitative assessment: Participant 05 created a reasonable number of nodes (30) and links (28). They drew extensively on external sources such as mainstream media articles and online medical resources as well as some images from web searches which they copied onto the canvas. They created two small clusters on the canvas for two individuals (with associated images, Twitter accounts and summary descriptions) via the chat interface. They created a small number of text nodes with some descriptive text to annotate some of their findings, mainly around the Delta variant, although this could be content sourced from the internet and copied into this text node. They also extensively labelled their links between nodes rather than leaving them unlabelled.

This participant created three story elements and used images from the canvas for two of these to help illustrate the story, but they did not link any nodes or links to the story nodes. In the first story element they noted that Delta cases were rising in early July, whereas in the second story element they noted that a major press announcement was made by the Prime Minister on 5$^{\text{th}}$ July. The third story element simply merged and summarised both points.

As expected, all the participants were all able to quickly contribute relevant information into the knowledge graph through the creation of nodes and links. The more advanced capabilities such as palette customisation and detailed property definition were less frequent than the basic knowledge creation activities, which is consistent with the novice experience level for these participants and the short 2-hour time window for the experiment. However, even with the small available time window and the lack of prior experience, all participants were able to show a broad variety of behaviours within Cogni-sketch as well as seeking resources and information from their external environment as well.

The basic observations from the qualitative analysis of participant results are:

1. All participants successfully used the environment to capture and relate task-relevant information they gathered.

2. A variety of visual and structural styles were used. Some favoured nodes and links (one participant only used nodes), whereas others put a lot of their content into stories. Some participants linked their stories to nodes on the canvas, whereas others left them standalone and provided only narrative textual descriptions in place of links.

3. Layout seemed to be an important consideration, with most participants creating visually meaningful graphs. Word clouds, dynamic charts and embedded tweets and other media were used by many participants.

4. All participants were able to create narrative story overlays on top of their knowledge graph, to explain their progress and conclusions.

5. All four classes of sensemaking behaviour were observed (*navigate*, *collect*, *build* and *story* as shown in Figure 6.5), with the majority being evidenced in the nodes and links created on the canvas (showing *collect* and *build* activities), as well as the *story* activities which resulted in the creation of the story elements.

6. Story creation tended to occur towards the end of the exercise as the participants considered and summarised their progress, but the creation of stories did not stop the other activities from continuing.

7. Some participants explicitly answered one or both questions, whereas others provided material that could answer them but did not explicitly state Q1, Q2 or to which of these their gathered material applied.

8. A small number of participants displayed more advanced behaviours, for example: labelling links, bending links to accommodate better layouts, creating non-standard properties on nodes and links, or creating new palettes.

9. Some participants used certain node types to indicate questions or hypotheses and then linked other nodes to these as they found supporting information.

10. From a stylistic point of view, some participants kept the guidance graph whereas others deleted it, with others creating a new blank project and working in that instead.

11. One participant used raw Twitter search as they couldn't figure out the embedded Twitter navigation pane[8], but later in the exercise they tried again and moved over to using the embedded pane and were able to then embed tweets and other artefacts accordingly.

The overall summary from the qualitative analysis is that the variety of approaches taken, and styles used, was broad and varied considerably across participants. The ability for a single environment to support a variety of visual and semantic styles is encouraging, implying that the participants did not feel constrained in their ability to create and place information within the environment. For this exercise, involving no independent machine agent support, the variety of

[8]They fed this back via the text in one of their stories.

styles used is not an issue, however in later experiments it would be interesting to see what kinds of assistance could be provided by independent machine agents to support deductive, abductive or inductive support [162], and the wide variety of styles and approaches used would need to be consistently consumable by these machine agents to support that.

The level of engagement from most participants was very good, along with the variety of sources consulted and content captured. Most participants created material relevant to the two questions posed as part of the exercise and some explicitly attempted to relate this material to the two questions. A wide variety of answers were given, especially for Q2 (the deliberately open-ended question), with investigations into football, Love Island, government briefings and hysteria being undertaken by various participants.

There were no errors reported by the participants and nothing reported in the logs for the duration of the experiments. No explicit support was requested for how to create different types of media within the environment, and some participants showed a high degree of visual layout ranging from traditional node-and-spoke type layouts to some participants creating a rows-and-columns style of layout on the canvas. Some participants attempted to directly answer the assigned questions and almost all participants attempted to tell a supporting narrative story alongside the knowledge graph they created, using the story element capability provided.

Usability assessment

Immediately after the experiment each participant was asked to complete a SUS feedback form, with a response rate of 100%. The results are shown in Table 6.4 for each of the twelve anonymous experiment participants, with their participant id shown in the first column[9].

[9]Participant ID is consistently used across the three analyses. i.e., *user 01* is the same anonymised participant in the quantitative, qualitative and usability analyses.

The SUS is widely used in contextual usability evaluation and a score of 68 is considered to be the average, with systems rated above 68 being deemed 'good' [75]. The SUS rating of the Cogni-sketch environment when used for this OSINT analysis exercise was rated as 72.1 on average, placing the usability experience well into the good category, and when considering the value as a typical academic result on a graded scale this corresponds to a solid 'B' grade on a typical A-F grade scale [130].

SUS consists of ten questions, five of which are phrased positively, and five of which are phrased negatively, each of which is scored on a five-point scale rating from *strongly disagree* to *strongly agree*. The algorithm to compute the SUS rating considers the positive and negative ratings and computes the aggregate score accordingly. SUS is structured in this way to mitigate against feedback that ignores the questions and always picks the same answer.

The ten SUS questions are:

Q1: I think that I would like to use this system frequently.

Q2: I found the system unnecessarily complex.

Q3: I thought the system was easy to use.

Q4: I think that I would need the support of a technical person to be able to use this system.

Q5: I found the various functions in this system were well integrated.

Q6: I thought there was too much inconsistency in this system.

Q7: I would imagine that most people would learn to use this system very quickly.

Q8: I found the system very cumbersome to use.

Q9: I felt very confident using the system.

Q10: I needed to learn a lot of things before I could get going with this system.

The results from the SUS survey can be seen in Table 6.4. The yellow and red cells indicate ratings that are below average for a given question. The SUS guidance states that the methodology prevents individual questions from being investigated in isolation but, in this particular experiment it is interesting to note that the few lower rated responses predominantly come from one participant within the cohort (participant 05)[10]. Two other participants also gave poor ratings for questions four and eight ('technical support needed', and 'the system is cumbersome'). These suggest that there is additional room for improvement for new/novice users either in terms of making capabilities more obvious, in improving the documentation, or in better techniques for gradually introducing complexity, for example, by adhering to the zero-overhead principle, where "no feature may add training costs to the user" ([107]).

#	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS
1	3	2	4	3	5	1	4	1	3	3	**72.5**
2	4	2	4	3	5	2	4	2	4	2	**75.0**
3	3	2	3	1	3	2	4	4	4	1	**67.5**
4	5	2	4	4	4	2	4	3	4	2	**70.0**
5	4	5	3	4	4	2	4	4	2	4	**45.0**
6	4	2	4	1	4	2	4	2	4	2	**77.5**
7	4	2	4	2	3	3	4	2	4	2	**70.0**
8	4	3	3	2	4	2	4	2	3	3	**65.0**
9	4	1	4	2	4	1	4	2	4	1	**82.5**
10	4	2	4	3	4	2	4	2	4	3	**70.0**
11	4	2	4	1	5	1	4	1	4	2	**85.0**
12	5	2	4	1	5	2	4	2	5	2	**85.0**
Avg	**4.0**	**2.25**	**3.75**	**2.25**	**4.17**	**1.83**	**4.0**	**2.25**	**3.75**	**2.25**	**72.1**

Table 6.4: System Usability Scale response summary

In conclusion, the usability assessment for this experiment, based on the

[10]See the previous subsection on qualitative assessment for a copy of the canvas and story detail created by participant 05. This participant was chosen as the example partly because of their lower results reported in the SUS.

widely used SUS survey technique, showed a good rating for the Cogni-sketch environment, with participants collectively giving an average rating of 72.1 corresponding to a B grade. One participant gave lower ratings in general but otherwise the results were generally consistent. This is a successful result for this first formal experiment with the Cogni-sketch platform being using to support OSINT analysis and sensemaking.

6.4.4 Experiment discussion

Considering the three dimensions of the analysis for this experiment it is clear that the Cogni-sketch environment was able to support untrained users in the task of sensemaking within a constrained time period. A variety of styles and layouts were observed in the qualitative assessment of the materials created, and the quantitative analysis for the participant activity shows a range of contributions both in terms of number of events as well as artefacts added into the knowledge graph. Every participant was able to create one or more story elements and the qualitative analysis reveals that every participant was able to create content that made sense in the context of the suggested investigation and target questions.

The first hypothesis for this experiment relates to *RQ1* for this thesis, and whether human creativity could be supported by the Cogni-sketch environment.

> **Hypothesis 1**: That untrained human participants can use the Cogni-sketch environment to create knowledge into the graph from a variety of sources and show human creativity *(RQ1)* in the visual and structural styles used.

The wide variety of stories and knowledge graphs created indicates that not only was human creativity being supported, but that it was happening in a variety of styles and formats.

The second hypothesis relates to *RQ3* and was that the environment could support the four categories of sensemaking defined in Figure 6.5 and that these could occur in any order without the need to flow only forwards from *collect* to *build* and then to *story*.

Hypothesis 2: That the process of sensemaking can be supported, with information and knowledge being curated in mechanisms aligned to the foraging, schematising and storytelling aspects of the sensemaking process *(RQ3)*, and explicitly that these can occur in any order, not being limited to flow only one way from foraging to storytelling.

The quantitative behaviour data supports this, showing concurrent *collect* and *build* activities for most participants across the duration of the two-hour experiment. Whilst the *story* events are fewer and therefore generate lower quantities of events, it is clear that story elements were created throughout the exercise (multiple participants use the story elements to document some aspects of their progress through the exercise) as shown in Figures 6.13 and 6.14.

Finally, the results from the SUS survey indicate a good level of feedback for the usability of the Cogni-sketch environment. This, coupled with the productivity range of the participants and the number of nodes, links and stories created shows that the Cogni-sketch environment was able to be successfully used for basic sensemaking and storytelling during this experiment.

6.5 Limitations and extensions

The pilot exercise identified that there are numerous ways in which the core environment could be further extended (through plugins) to more directly support the intelligence analysis process, based on different approaches and techniques in the literature (e.g., [64]).

6.5.1 Independent machine agent sensemaking support

Additional value to the human users could be achieved through implementation of commonly used forms and processes from existing techniques [65] and, through formal representations, the ability for machine agents to make use of the credibility and certainty of information contained in such reports. Both read and write access for independent machine agents (with suitable security permissions) are available, enabling the Cogni-sketch environment to be tailored with processes and agents for each specific need in the future.

The same opportunities relate to downstream support and integration with existing systems; with the Cogni-sketch system easily being extended to support generation of existing report structures and formats, providing the human operator with straightforward techniques to enable them to populate these with relevant data from the knowledge graph at a point in time. For a separate but relevant example of this refer to the 'Science Library' example reported in Section A.1.1 and published in [28]. Such reports can also be regenerated in the future as the operational picture, reflected in the knowledge graph, is updated, with all these iterations subject to version control if needed, with the version control information recorded in the knowledge graph and extracted for inclusion in the generated reports as needed.

During the pilot the analyst specifically requested the generation of a particular MS-Word format for briefings, but this was outside the scope of the broader research activity so was not implemented. The underlying data in the knowledge graph would have supported this, and the technique suggested for identifying which pieces of knowledge would go into which section of the report was to add additional palette items and link these into the knowledge graph to identify which nodes and links applied to the different parts of the requested report. This would enable the report to be directly generated from the graph and used within an existing established process without needing to modify that process.

Examples of possible future machine agents that could be included in the environment to directly assist with the sensemaking process are:

- A *deductive agent*

  Aware of the semantics of the items defined in the palette, including logical inference rules, can apply these rules in real-time as the human users create items in the knowledge graph, alerting them to relevant deductive inferences and adding new knowledge to the graph if permitted by the human users. Such capabilities, based on first order predicate logic are well known in reasoning systems, and a simple subset of inferences based on inheritance, node and link types, and simple logical inference rules can be supported (e.g., through a simple interface to a Controlled Natural Language (CNL) reasoning system [23]).

- An *abductive agent*
  Able to easily process the knowledge graphs as they are constructed, looking for repeated patterns in the data, advising the human users of this analysis and indicating parts of the graph where additional conclusions could potentially be made, based on other similar structural patterns within the graph that contain additional nodes. Machine support for abductive reasoning is possible (e.g., in a manner like [89]), but not yet directly explored in the Cogni-sketch environment but would be achieved in the same way through a typical machine agent plugin.

- General support for *inductive* reasoning
  Rather than propose a specific machine agent for inductive reasoning, instead it is likely that a set of support systems may be useful here, for example to summarise the nearby network within a graph to help the human user understand the context for a particular item of data. Another example could be the presentation of different combinations of supporting data to the human user to see whether any additional insights or inspiration is achieved by presenting these randomised pairs.

6.5.2 Additional support for sensemaking principles

Existing support for the 9 principles from Attfield et al [8] was described in Section 6.2.2, and there is more that could be added, as well as to support additional sensemaking processes and relevant approaches as described in Chapter 2. The list of all possibilities would be very long, but specifically for the 9 Attfield sensemaking principles there are a small number of relatively easy future extensions that could add further value, and these are listed in Table 6.5.

#	**Support in Cogni-sketch**
PR1	*Provide sufficient cues for sufficient sensemaking*: (1) Embed frames/cue-patterns within the environment to guide users based on previous paths. They could be implemented as complex structures available from the palette that are constructed on the canvas, or conversational interactions to fill slots to build the sub-graph structures. (2) Independent machine agent(s) to annotate the emerging graph with specific nodes added and labelled to suggest missing information or open questions etc. The agents can either be built to directly detect specific patterns, or trained on graphs more generally to detect typical patterns that occur.
PR2	*Support low cost information workflows*: For any more traditional workflows it would be better to use existing tools (e.g., Node-RED [49]). Fairly straightforward integration within the knowledge graph should enable substantial benefit from external workflow tools as well as predefined workflows, without the need to unnecessarily replicate that workflow construction environment within Cogni-sketch.

PR3	*Represent information quality and provenance*: For example, to enforce/encourage the human users: A simple visualisation plugin-in that can easily be added to the environment to visually depict the quality or certainty of information through capture of relevant certainty information for relevant nodes and/or links and highlight when it is missing. The addition of such information at the atomic level of individual nodes and links also allows easy inference of overall quality at higher levels of abstraction, based on various algorithms (average, weakest etc).
PR4	*Promote expertise and domain knowledge*: Better direct support for richer structures such as frames, or more advanced forms of stories that are directly mapped to different cognitive styles as discussed in Section 6.3.4.
PR5	*Allow time to acquire data/information to build an evidence-based and coordinated situation picture*: The ability to match different views or panes to different user roles that interact with the environment. Through this approach the same core data in the graph can be presented to different users in different ways and at different levels of granularity depending on their role and context.
PR6	*Use strategies for negotiation of sense*: Consider specific visual cues for high-value data, and possibly a custom pane for computing user contributions, for example in the style of a leaderboard for bounties gained; to drive competitive behaviour via gamification.

PR7	*Where appropriate, use strategies for frame enumeration and elimination*: (1) Additional example capabilities to show the potential, e.g., a "Devil's advocate" function, and a new pane that uses simple techniques to challenge the user to come up with alternative information or hypotheses. (2) The closest current capability to this is the story elements; a future simple extension would be to allow different cognitive frames to be represented by a story type, building on the current simple description-based approach. This would allow the user to quickly switch between frame types to see what fits and what is missing, and the options could be ranked according to "best fit" or vice-versa.
PR8	*Provide explanatory context for actions, orders and requests*: The ability to achieve the capture of this contextual information already exists but is basic and without constraints. An additional pane like the storytelling pane described in Section 6.3.4 could be used for tracking actions and relevant contextual information to more easily summarise what is being requested and shared.
PR9	*Minimise the costs of achieving and maintaining common ground*: There are already plans for further improvements to the collaborative capabilities and the specific modes of collaboration that they can support (See Section 7.2.1). Achieving and maintaining common ground should be one (of many) requirements that inform these future collaboration capabilities.

Table 6.5: Future support for the Attfield 9 principles

6.6 Chapter Summary

This chapter introduced the need for sensemaking as a specific use case to test the tellability aspects of HAKF with human users through the Cogni-sketch environment. It started with a definition of sensemaking activities and included analysis of recent material from a small number of sources motivating the need for capabilities to support specific principles for sensemaking. A brief assessment of how Cogni-sketch can support these principles was given. The useful Pirolli and Card model of sensemaking [113], as a series of interconnected loops was chosen as the process against which to evaluate behaviour. These loops increase in structure and effort as they rise from foraging to sensemaking to storytelling. The OSINT analysis and sensemaking use case that underpins *RQ3* was aligned to this model of sensemaking and was the basis for a 3-month pilot with an expert analyst using Cogni-sketch to undertake a real OSINT analysis exercise, with obfuscated results report in Section 6.3 along with various lessons learned and improvements made along the way.

During this pilot exercise it became clear that user behaviour could be instrumented according to the sensemaking categories, and a formal user experiment was designed. This experiment was defined with two hypotheses, each corresponding to *RQ1* and *RQ3* to explore the potential for untrained human users to perform sensemaking. The results from this experiment are reported in Section 6.4. The experiment was carried out with twelve participants who each used the Cogni-sketch environment to investigate two specific questions using a set of carefully curated social media data. The duration of the exercise was two hours, with each participant being new to the environment and only receiving a short video and textual description of the task. The experiment results were analysed three ways:

1. Through a detailed quantitative analysis of event data generated by each participant.

2. By qualitative assessment of the results created by each participant.

3. Through a formal SUS survey to measure usability.

The results of the experiment showed that the stated hypotheses were viable, with supporting evidence for each coming out of the assessment of results.

Chapter 7

Conclusion and future work

7.1 Summary of contributions

I have published the research reported in this thesis in various peer reviewed publications, comprising the formalisation of the HAKF concept, the development and open-source release of the Cogni-sketch platform, and the design and execution of pilot, experiment and evaluation activities. These activities were all my individual contributions and formed a key and novel set of capabilities supporting a broader collaborative effort as described in earlier chapters. These have spanned a seven-year period and were largely performed as part of the now completed DAIS ITA research program. Figure 7.1 shows a visual representation of the timeline for these various activities and how they relate to the completion of this thesis.

The four research contributions listed below were stated at the start of this thesis in Section 1.4. They are revisited here, with a brief narrative for each identifying the nature of the contribution, to what extent it has been addressed, and where the relevant details can be found in this thesis. These contributions were motivated by the research question and three specific sub-questions, also described in Section 1.3. The research questions and corresponding contributions[1] are visually summarised in Figure 7.2.

[1]The research contribution descriptions have been abridged in the diagram.

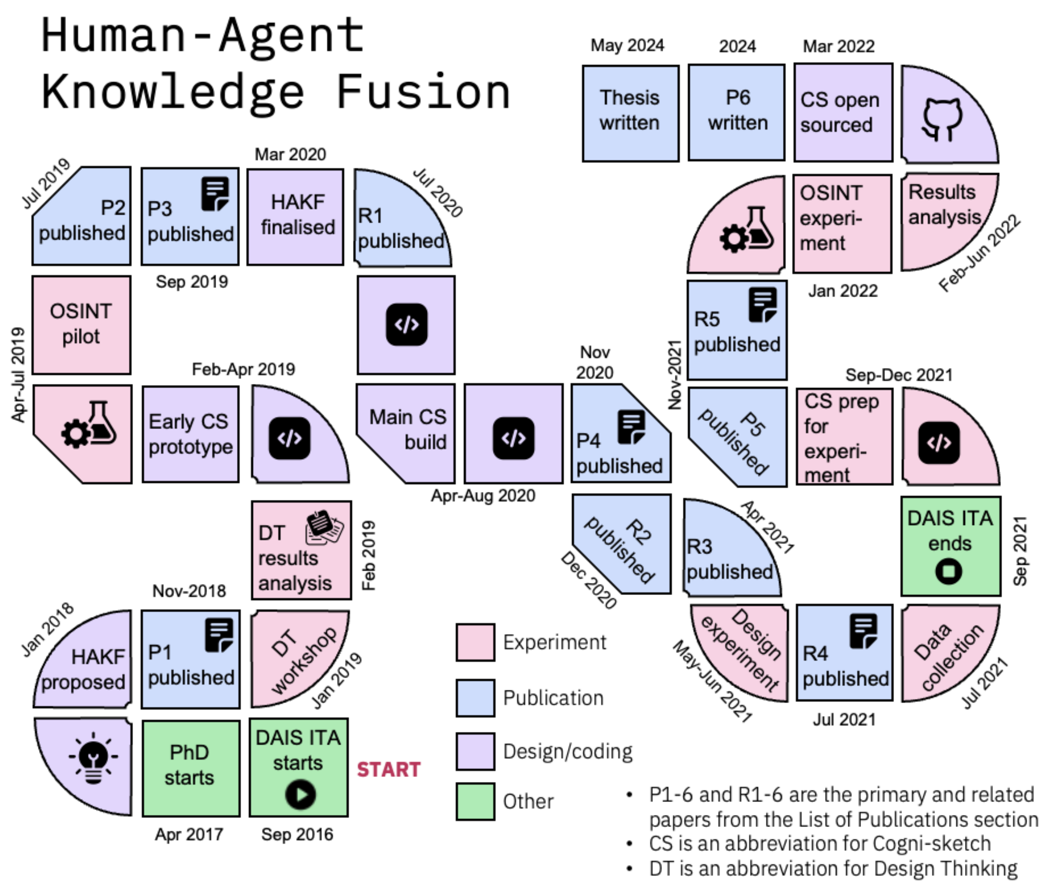


Figure 7.1: Timeline for HAKF research activities

7.1.1 Definition of the HAKF concept

This contribution was originally stated as:

> *Definition of HAKF* as an under-pinning concept to support the agile human-agent collaborative environment that is outlined in this thesis (specifically to support *RQ1* and *RQ2*), comprising the tellability and explainability flows, and how these can support collaboration through increased confidence and context-aware knowledge and configuration.

HAKF has been defined, along with specific *required capabilities* that are necessary for any implementation of HAKF to be viable. The primary flows of tellability and explainability are defined as well as the various roles of human users

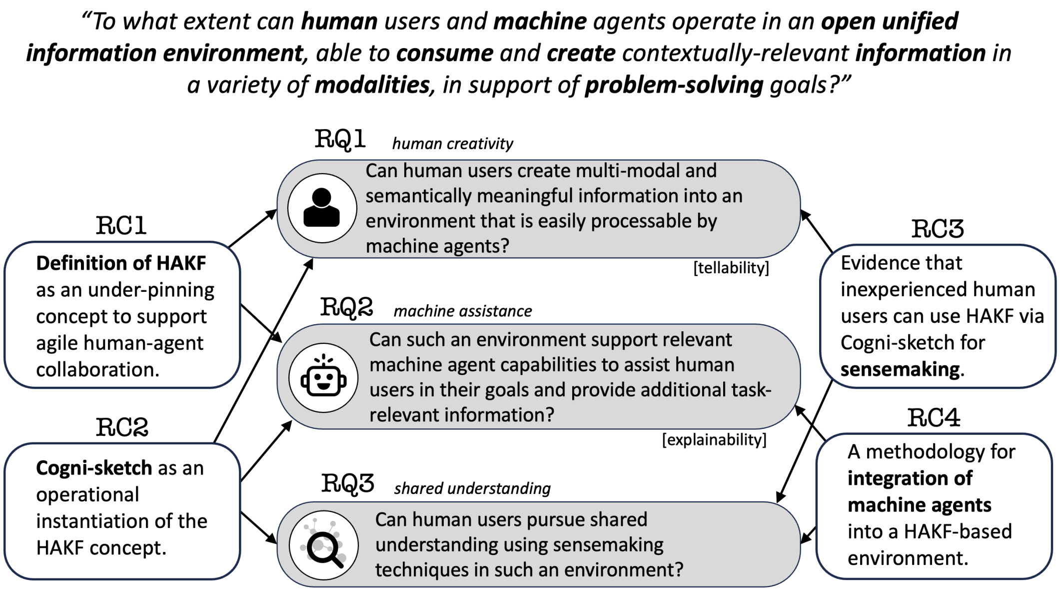


Figure 7.2: Summary of research questions and contributions

and machine agents that can interact with a HAKF system. In defining HAKF the relevant literature was reviewed, especially for noted gaps or calls to the community for attributes needed in collaborative and HAT settings. The focus on sensemaking as a motivating use case also identified a small number of specific requirements for enabling human users and ensuring that their cognitive needs are accounted for. HAKF was also directly informed by our own DT workshop exercise with a team of U.K. military stakeholders alongside representatives from government and industry who were tasked with considering what AI assistance and XAI needs would likely be for future military support systems. The combination of all these factors is captured in the definition of HAKF as a distillation of the relevant factors and minimum required set of capabilities. Additional earlier research into conversational interactions for XAI systems informed aspects of explainability for HAKF. Chapter 3 provides a summary of these required capabilities that must be fulfilled to achieve HAKF as an underlying definition for the creation of more applied implementations (such as Cogni-sketch).

7.1.2 Development and release of Cogni-sketch

This contribution was originally stated as:

> *Creation of an operational instantiation* of the HAKF concept for human-agent collaboration, enabling experimentation and general usage. Specifically, this resulted in the development and open-source release of the Cogni-sketch environment along with specific plugins to implement various machine agent capabilities (*RQ2*), human problem-solving and visualisation capabilities (*RQ1*), and features specific to sensemaking and SU (*RQ3*).

The Cogni-sketch platform has been successfully created and was released as open-source on GitHub under an MIT license in March 2022[2]. A full list of all plugins that have been created so far can be found in Appendix A (Section A.4) and a small set of these have also been released as open-source software, with additional plugins able to be added in the future as appropriate. The Cogni-sketch platform underpins the evaluations, pilot and formal experiment reported in Chapters 5 and 6 and all the additional examples reported in Section A.1. It has been successfully used in several additional projects at IBM Research U.K. and Cardiff University. Cogni-sketch represents an operationalisation of HAKF and has been designed to support an expanded set of HAKF user roles and features a number of well-defined plugin extension points to allow easy contribution and sharing of new capabilities for specific needs. Cogni-sketch forms the technical basis for the experimental aspects of this research thesis and has been used wherever possible to demonstrate the feasibility of the wide range of human user and machine agent interaction types described throughout.

The Cogni-sketch platform has been used to assess the three research questions

[2]The core Cogni-sketch platform is available at `https://github.com/dais-ita/cogni-sketch` and a small subset of the stable plugins are also available, at `https://github.com/dais-ita/cogni-sketch-plugins`.

(RQs) previously identified, and the answer to these RQs are given against the most relevant contribution in the remaining sub-sections below.

7.1.3 Human sensemaking exercise and results

This contribution was originally stated as:

> *Evidence that inexperienced human users can successfully use HAKF* as embodied in Cogni-sketch for sensemaking and communicate their findings. This is achieved through the execution of a formal human user experiment into the use of the Cogni-sketch environment for a simulated sensemaking exercise based on analysis of social media sources (*RQ3*). This includes the answering of two predefined intelligence questions, making use of the Pirolli and Card [113] sensemaking loops including foraging, sensemaking and storytelling to communicate the conclusions of each participant (*RQ1*).

Details of this contribution are extensively reported in Chapter 6. A long-running OSINT pilot was run with an intelligence analyst performing sensemaking within the Cogni-sketch environment. The duration of the pilot enabled substantial improvements and extensions to be made, with the ability to perform storytelling being a key insight and plugin capability. The pilot also informed the design of a formal experiment to evaluate sensemaking within Cogni-sketch and identified over 100 event types that could be instrumented within the platform to enable a detailed behavioural assessment of participants and identification of which part of the sensemaking process they were performing.

The experiment showed that all users were able to successfully use the Cogni-sketch environment to carry out sensemaking and storytelling activities. It was observed that these activities were carried out non-linearly, meaning that the participants were able to perform different parts of the sensemaking process when needed and were not forced into a linear flow through the system. They did not need to use additional tools for note-taking, foraging or any other aspect of the experiment and the SUS usability evaluation reported a good perception of Cogni-sketch across the participants in a brief survey directly after the exercise. In addition to the usability survey, the results of the experiment were extensively analysed quantitatively for participant behaviour, and qualitatively through an assessment of the artefacts created, the questions answered, and the visual styles used.

The human sensemaking exercise (and corresponding pilot exercise) reported in Chapter 6 demonstrated that human users can create multi-modal and semantically meaningful information into the Cogni-sketch environment *(RQ1)*, and the work demonstrating the integration of explainable machine agents reported in Chapter 5 demonstrated that information in this same environment can be easily processed by machine agents *(RQ2).* The human sensemaking exercise also showed that human users can pursue shared understanding using sensemaking techniques in this same environment *(RQ3).*

7.1.4 Practical demonstration of machine explanations

This contribution was originally stated as:

> *A methodology for integration of machine agents* into the HAKF environment, specifically through co-construction of machine generated information into the knowledge graph. This takes the form of various explanations and other contributions from machine agents operating within the Cogni-sketch environment (*RQ2*) in a variety of SU demonstrations (*RQ3*), with development and integration of these machine agents carried out by the author of the tool and this thesis as well as other collaborators, with the latter helping to validate the simplicity and extensibility of the architecture for the insertion of machine agents.

Chapter 5 describes each of these integrations and the ways in which the independent and directed machine agents operate in the Cogni-sketch environment and can contribute to the sensemaking activities with the human users of the system. There are a variety of evaluations reported, including an end-to-end pilot integration of multiple external machine agents using the Cogni-sketch environment and a subsequent evaluation using a form of workflow orchestration, but in the style of a mind map with semantically relevant material being contributed (and consumed) by the machine agents. There are also examples of a conversational interaction pattern with machine agents providing different forms of explanations to human users based on analysis of imagery, video and other sources.

In these, considerations of security and privacy are present, and various metadata is defined to enable confidence information to be communicated. There is also an informal evaluation of the ability to achieve integrations such as these in a real-time setting using a locally convenient example with a live feed from a webcam and dynamically generated provenance information based on entity detection. The purpose of this evaluation is to determine whether the previously defined examples can be recreated in a real-time operational setting, via the existing extension points within the Cogni-sketch environment, and whether this can be done with a small amount of effort in a short time frame. The types of explanation are described along with the co-construction basis for their addition to the knowledge graph, enabling other users to read them but also provide additional annotations, extensions or corrections. Full details of all these exercises can be found in Chapter 5.

These various machine agent integrations demonstrate that the Cogni-sketch environment can support relevant machine agent capabilities to assist human users in their goals, and provide additional task-relevant information *(RQ2)*.

The additional minor contributions that were listed are:

1. Exploration of methods for Situation Understanding (SU) with human and

machine agents.

2. Summary of stakeholder requirements for interactive AI systems.

3. Application of HAKF and Cogni-sketch in real OSINT analysis exercises.

Each of these is reported within this thesis, with (1) reported in [21] and discussed in Chapter 3. (2) yielded insights for potentially valuable future capabilities in the area of human-agent collaboration with AI and XAI systems as a result of the DT workshop held with military stakeholders, reported in Chapter 3. Finally, (3) is reported in Chapter 6 following the use of the Cogni-sketch platform in an OSINT investigation with an OSINT analyst.

7.2 Future work

This section lists a small set of possible future requirements and builds on the functional extensions related to sensemaking described in Section 6.5. The content of this section falls into two broad categories: (1) a small set of non-functional enhancements that should be made as part of the operationalisation of the Cogni-sketch environment, and (2) a set of exciting future opportunities for integration with Large Language Model (LLM) based machine agents which could provide a substantial advance in potential machine agent capabilities.

7.2.1 Future non-functional enhancements

For this experimental implementation of Cogni-sketch some simple technical solutions have been chosen to make adoption and installation simpler, but which have been done in a way that supports easy upgrades to a more scalable equivalent in the future. This is an important consideration and represents a trade-off in development effort for the Cogni-sketch Minimum Viable Product (MVP) verses a larger development effort to support scalability when needed.

Collaboration is minimally implemented as described in this thesis. Specifically, there are two forms of collaboration that are currently supported (See Section 4.3.3 for the full description). The support for broader collaboration options have been explicitly avoided since *RQ1* is focused on human creativity and usability, and the solution to a useful and understandable collaboration solution for genuine multi-user concurrent knowledge graph building will require substantial additional investigation to achieve properly. Experiences with similar systems suggest that the typical free-for-all approach to collaboration can be disconcerting and a level of control and permission-granting or zoning may be required by the overall project owner [68]. The ability to achieve broad collaboration is not technically challenging as the platform has been designed with this in mind, but careful thought relating to the impact on human cognitive processes is needed before implementing anything specific. Collaboration styles and their impact on user behaviour could be a substantial focus area for future work and the Cogni-sketch environment could provide a useful experimental basis for this.

There are also additional future non-functional requirements that are important, but less fundamental than the collaboration example highlighted above:

- **Data storage**
  As mentioned previously, for ease of implementation and installation the data for the Cogni-sketch server is simply stored as files. It is very easy to upgrade this on the server to use a document database or similar and the change will be invisible to all users. This should only be done if larger volumes of usage or very large graphs need to be interacted with, or if advanced cross-project or cross-user collaboration or search capabilities are needed.

- **Authentication**
  The current MVP implementation uses the 'passport' JavaScript library which provides a wide range of authentication options, including the use

of external identity providers which would be the eventual goal for Cogni-sketch. The current implementation uses locally encrypted passwords stored in a file so users must be manually created by a system administrator (there is an admin UI to make this easy for them). This is the preferred solution whilst the solution remains more tightly controlled and therefore user access should be granted and managed centrally. To enable the full solution with self-registration and external identity management is very straightforward.

- **Internationalisation**
  The 'i18next' library has been used to provide internationalisation, and most user-facing text has been abstracted into a language file to make subsequent internationalisation straightforward. Currently only the English language is supported, but the process has been tested with a few German translations. Should translation resources become available in the future there will be little technical effort required to achieve this apart from populating each of the language files with the correct translations.

- **Encryption**
  See above for comments about user password encryption. Also, the platform has been tested extensively with https protocol to ensure that information in transit is encrypted. Currently the saved knowledge graphs and activity logs are not encrypted but this is best done when migrating away from simple file-based storage to a document database or similar, and enabling encryption within the storage system that is used. This will require no additional code changes to the Cogni-sketch platform beyond moving to a database solution.

- **Scalability**
  Like collaboration, this is another area where the relatively easy technical scalability of the solution should not be confused with the much more complex issues surrounding human consumption of large-scale graph data,

especially when considering *RQ1*. Even without explicitly supporting scalability there are several techniques available to the user around hiding nodes, creating zones or areas within their graph, or splitting the graph into different sub-graphs (as separate projects). As noted in Pienta et al [112]: "approaches (are required that) have strong potential in handling large graphs, algorithmically, visually, or interactively" and "multiple research fields — data mining, machine learning, human-computer interaction, information visualization, information retrieval, and recommender systems — to underline their parallel and complementary contributions to graph sensemaking". This is a complex area and investigating these kinds of scalability are not in scope for this thesis but provide excellent opportunities for future work.

7.2.2 Other potential enhancements

There are also some additional opportunities for future investigation mentioned in this thesis and related publications:

- **Further refinement of the explanation meta-model**

  In [21] a basic meta-model to support the integration of services and explanations was introduced and is mentioned in Chapter 5. This is experienced by the Cogni-sketch users as an extensible palette within the environment with node types for explanations, services, events and other types that can be used to define rules for consumption by a machine agent. It is straightforward to add new sensors as additional palette items (as sub-types of the main sensor type) and to define additional types and map them to the machine agent, but a substantial improvement would be the definition of a broader set of predefined explanation types. These can define expected properties and have custom rendering styles to more effectively communicate the explanation to human users and the presence of additional specific properties for the explanation types will make integration with machine agents simpler too.

- **More intuitive and expansive uncertainty information**

  What types of visualisation (or description) can be used to express uncertainty information in a manner that is able to be recognised and understood by human users? Building on the work reported in [21] this may encompass multiple dimensions of uncertainty from different sources (both human and machine). Design of visual techniques for clearly conveying this uncertainty to the human decision-maker [135] can be explored, and the embedding of this as a node decoration onto the Cogni-sketch canvas. Examples include: inherent uncertainty associated with the data feed, the location, the fidelity, any time-delay, or even the trust associated with the partner that is providing that feed.

  There may also be inherent uncertainty in the core classification decision [59, 104], or the explanation [33], as well as potential ambiguity or uncertainty in the alignment of that core classification outcome to any higher-level domain concepts. All these sources of uncertainty may be potentially important to the decision-maker (or other agents within the system), but how can this be conveyed in a consumable form that does not add substantial cognitive burden? Various visualisation techniques could be defined and assessed in different settings with users, along with a conversational (or visual) interaction to allow the users to explore more deeply when needed.

- **Empowered autonomous machine agents**

  It is straightforward to integrate autonomous agents supporting humans in their activities (although it will not be straightforward to build the agents themselves). Consider, for example, the case where one team member needs privileged access to some resources or a data stream provided by another member. Assuming that such data streams can be guarded by autonomous agents following a set of guidelines the humans provided them with, we can thus embed in our HAKF layer an autonomous agent negotiating with the autonomous guardian(s) of service level agreements involved in accessing the data stream. For example, a request to access a real-time full-definition video stream from a surveillance camera mounted on an unmanned autonomous vehicle (UAV) whose location might be sensitive and cannot be disclosed. In this example the embedded agent would negotiate within the constraints (e.g., a time delay must be imposed), seeking to obtain information relevant to their mission. In previous research we worked on enabling autonomous negotiation in shared settings [152], and this could be the basis for such an integration, alongside fine-grained policy application for accurate information sharing.

- **Explicit support for defined roles**
  Cogni-sketch can be extended to more explicitly support role-based experiences through the definition of predefined but flexible roles for all users. These would be specialisations of the roles already defined (e.g., operator, creator etc). This role information can then be used to determine which panes, palettes and agents are available to them, and the way in which they experience them. For example, specific explanation formats can be defined for each role, and the core explanation. Information (available in the graph) could be transformed into the role-specific format prior to sharing with the user. This would be achieved through the creation of meta-models for the roles and their typical (but generic) needs, and the ability for functions to declare capabilities aligned to those needs.

Since the Cogni-sketch platform has been released as open-source software and is publicly available on GitHub it is also hoped that extensions and additional plugins will be provided by other members of the community and any pull requests received with such contributions will be gladly considered.

7.2.3 Large Language Model (LLM) opportunities

As mentioned earlier in this thesis, the biggest recent advance in capability relevant to this work is the high-profile emergence of new foundation model and LLM capabilities that are then made available in tools such as ChatGPT and others. The ability to achieve high-quality outputs on a diverse set of tasks through only prompt-engineering or prompt-tuning of the same model is a step-change in capability and versatility with many high-value uses already possible and many more anticipated to come. The tantalising aspect of these new LLM capabilities is the potential for a deep and rich integration between the cognitive and human aspects provided by HAKF and Cogni-sketch combined with the powerful but often contextually unaware capabilities of LLMs. Through inclusion of relevant knowledge graph information into LLM prompts it may be possible to quickly achieve a level of integration not commonly seen in typical LLM use cases today.

The work reported in this thesis was largely complete by the time these models became widely available so no specific interactions or implementations have yet been undertaken but there are numerous potentially high-impact capabilities that can be investigated as future work:

- **Provenance tracking for prompt edits**
  When using LLM services like ChatGPT, a typical user will often try many different prompts before they get an ideal or preferred response. Tracking the changes between versions of these prompts can be ad-hoc (or non-existent) and there is opportunity to forget why changes were made, or what the impact of specific changes were. For both prompt tuning and prompt engineering it would be useful to have the ability to record your progress and the impact of changes made so that you can more easily return to earlier versions or try different 'branches' of change etc. Even without modification the Cogni-sketch environment can already be used for this kind of provenance tracking, for example as reported in [28]. You can also easily append comments or any other information that is relevant to the exercise.

A deeper integration may also be valuable by creating a specific set of palette items for different kinds of prompts or responses for example enabling the user to categorise the kind of change they made and whether the response was good or bad, improved or declined etc.
An integration such as this could be valuable for people exploring the impact of prompt tuning on running systems where regular prompt tuning is needed to keep responses on track as the inputs and desired outputs change over time.

- **Palette generation for a domain**
When Cogni-sketch is used for knowledge modelling it is typical for a config creator user to manually create a palette for the domain in which the knowledge modelling is being used. For example, this may be healthcare or scientific research, news and current affairs or a specific scientific domain. The ability for LLMs such as ChatGPT to generate structured data such as JSON means that it is credible to create a new LLM-based function to process a body of data such as a web page or local text file and generate a suitable Cogni-sketch palette based on the various entities and types mentioned in those documents. Even if the generated palette is not quite right or is incomplete it could still be a very valuable additional capability in bootstrapping the palette creation process.
For now, it is assumed that the human user will still manually choose the icons and colours used for the palette items as well as any schema data (node properties and data types) but it is plausible that even this level of detail could be proposed by the LLM in the future, including a mapping to existing ontologies.

- **Assistance with storytelling**
In the sensemaking experiment reported in Chapter 6 a key part of the exercise was storytelling and specifically whether the human users were

able to do this, and if so, how they chose to do it. One of the common tasks a LLM is used for is summarisation, and part of the storytelling process is to summarise a subset of nodes and links on the knowledge graph to convey the meaning for that subset. Not all storytelling descriptions are summaries but for those that are it would be very feasible to have a 'summarise' button on the storytelling pane that would invoke a model like ChatGPT with the collective data for all the selected nodes and links and use this to generate an example summary which could be reviewed and edited by the human user before saving.

- **Upgraded chat mode**
  This is an obvious example where the text interpretation and summarisation capabilities of these LLMs can be used to substantially upgrade the existing simple chat implementation that allows users to interact with their knowledge graph via a chat interface. Through few-shot examples in the prompt it is anticipated that it would be simple to map a small but powerful set of Cogni-sketch actions related to retrieving data from the knowledge graph or creating exported copies for summarising or sharing. The local knowledge graph is also ideal for implementing Retrieval Augmented Generation (RAG) [82] for the LLM, but it would be easy to add additional external sources to seek data from and provide phrases or buttons within the chat UI to easily add these into the graph.

- **Plan and execute agents**
  Specifically to harness the ability for LLM agents to plan and execute their own workflows to achieve specific goals based on the context of the knowledge graph and the position within the sensemaking process. This is currently a large area of focus within the LLM community with these kinds of self-planning agents able to perform well in a wide range of tasks, and

increasingly enabled by code libraries such as LangChain[3]. The potential for such LLM agents to co-construct relevant knowledge with other machine agents and human users via Cogni-sketch could offer an improved extensibility and agility for HATs using these kinds of agent-planned workflows and it would be a simple exercise to build a basic implementation to evaluate the potential here.

The LLM-related ideas listed above are deliberately limited to just the top five based on a combination of feasibility and potential impact, however there will likely be many more. The knowledge graph implementation for Cogni-sketch and the explicit support for easy inclusion of machine agents should make it a very powerful environment to augment the capabilities of LLMs and provide advanced opportunities to casual users beyond what can be achieved in the typical 'prompt studio' settings that are used to interact with these LLMs. In the interests of space the examples are limited to just text-based cases but there are other modalities such as image generation that could provide valuable use cases too.

[3] See `https://www.langchain.com/`.

Bibliography

[1] Leonard Adelman. Experiments, quasi-experiments, and case studies: A review of empirical methods for evaluating decision support systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(2):293–301, 1991.

[2] Mohammad Aliannejadi, Manajit Chakraborty, Esteban Andrés Ríssola, and Fabio Crestani. Harnessing evolution of multi-turn conversations for effective answer retrieval. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 33–42, 2020.

[3] Fraser Allison, Ewa Luger, and Katja Hofmann. Spontaneous interactions with a virtually embodied intelligent assistant in Minecraft. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2337–2344, 2017.

[4] Yasmeen Alufaisan, Laura R Marusich, Jonathan Z Bakdash, Yan Zhou, and Murat Kantarcioglu. Does explainable artificial intelligence improve human decision-making? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6618–6626, 2021.

[5] Eeva Aromaa, Päivi Eriksson, Jean Helms Mills, Esa Hiltunen, Maarit Lammassaari, and Albert J Mills. Critical sensemaking: Challenges and promises. *Qualitative Research in Organizations and Management*, 14(3):356–376, 2018.

[6] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020.

[7] Mark Lawrence Ashwell. The digital transformation of intelligence analysis. *Journal of Financial Crime*, 24(3):393–411, 2017.

[8] Simon Attfield, Shailey Minocha, Bob Fields, Emmeline Elliott, James Harryman, Rob Hutton, and Chris Baber. Enhancing sensemaking: Supporting distributed groups in the future operating environment. *Defence Human Sciences Symposium*, 2020.

[9] Alex Banks. From notetaking to Neuralink. `https://research.contrary.com/reports/from-notetaking-to-neuralink`, May 2023. Accessed: 2024-05-10.

[10] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. Beyond accuracy: The role of mental models in human-AI team performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7(1):2–11, October 2019.

[11] Brett A Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. Programming is hard - or at least it used to be: Educational opportunities and challenges of AI code generation. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pages 500–506, 2023.

[12] Richard Bellman. *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company, 1978.

[13] Nigel Bevan and Ian Curson. Methods for measuring usability. In *Human-Computer Interaction INTERACT'97*, pages 672–673. Springer, 1997.

[14] Yali Bian. *Human-AI Sensemaking with Semantic Interaction and Deep Learning*. PhD thesis, Virginia Tech, 2022.

[15] Erik Blasch, Élio Bossé, and Dale A Lambert. *High-Level Information Fusion Management and Systems Design*. Artech House, 2012.

[16] Erik Blasch, Chee-Yee Chong, Wolfgang Koch, Henry Leung, Tien Pham, Dave Braines, and Tarek Abdelzaher. Machine learning/artificial intelligence for sensor data fusion – Opportunities and challenges. *IEEE Aerospace and Electronic Systems Magazine*, 36(7):80–93, 2021.

[17] Erik Blasch and Susan Plano. Level 5: User refinement to aid the fusion process. In *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003*, volume 5099, pages 288–297. International Society for Optics and Photonics, SPIE, 2003.

[18] Erik Blasch, Steven Rogers, Hillary Holloway, Jorge Tierno, Eric Jones, and Riad Hammoud. QuEST for information fusion in multimedia reports. *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, 2(3):1–30, 2014.

[19] Grady Booch, Francesco Fabiano, Lior Horesh, Kiran Kate, Jonathan Lenchner, Nick Linck, Andreas Loreggia, Keerthiram Murgesan, Nicholas Mattei, Francesca Rossi, and Biplav Srivastava. Thinking fast and slow in AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15042–15046, 2021.

[20] Jeffrey M Bradshaw, Patrick Beautement, Maggie R Breedy, Larry Bunch, Sergey V Drakunov, Paul J Feltovich, Robert R Hoffman, et al. Making agents acceptable to people. *Intelligent Technologies for Information Analysis*, pages 361–406, 2004.

[21] Dave Braines, Federico Cerutti, Marc Roig Vilamala, Mani Srivastava, Alun Preece, Lance Kaplan, and Gavin Pearson. Towards human-agent knowledge fusion (HAKF) in support of distributed coalition teams. *AAAI Fall Symposium Series, AI in Government & Public Sector*, 2020.

[22] Dave Braines, Eunjin Lee, Gavin Pearson, and Alun Preece. Exploring the future of explainable AI solutions with military stakeholders. In *Proceedings of the 4th Annual Fall Meeting (AFM) of the Distributed Analytics and Information Science International Technology Alliance (DAIS ITA)*, Sep 2019.

[23] Dave Braines, David Mott, Simon Laws, Geeth de Mel, and Tien Pham. Controlled English to facilitate human/machine analytical processing. In *Next-Generation Analyst*, volume 8758, page 875808. International Society for Optics and Photonics, 2013.

[24] Dave Braines, Nick O'leary, Anna Thomas, Daniel Harborne, Alun Preece, and William Webberley. Conversational homes: A uniform natural language approach for collaboration among humans and devices. *International Journal on Advances in Intelligent Systems*, 10(3/4):223–237, 2017.

[25] Dave Braines and Alun Preece. Open source intelligence: Sensemaking evaluation for human-agent knowledge fusion. *Unpublished*, 2024.

[26] Dave Braines, Alun Preece, and Daniel Harborne. Multimodal explanations for AI-based multisensor fusion. *NATO SET-262 RSM on artificial intelligence for military multisensor fusion engines. NATO*, 2018.

[27] Dave Braines, Alun Preece, Colin Roberts, and Erik Blasch. Supporting agile user fusion analytics through human-agent knowledge fusion. In *2021 IEEE 24th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2021.

[28] Dave Braines, Jane Stockdill-Mander, and Eunjin Lee. The science library: Curation and visualization of a science gateway repository. *Concurrency and Computation: Practice and Experience*, 33(19):e6100, 2021.

[29] Dave Braines, Richard Tomsett, and Alun Preece. Supporting user fusion of AI services through conversational explanations. In *2019 22nd International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2019.

[30] John Brooke. SUS - a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[31] Mario Bunge. The relations of logic and semantics to ontology. *Journal of Philosophical Logic*, 3:195–209, 1974.

[32] Supriyo Chakraborty, Alun Preece, Moustafa Alzantot, Tianwei Xing, Dave Braines, and Mani Srivastava. Deep learning for situational understanding. In *International Conference on Information Fusion (Fusion)*, 2017.

[33] Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer M Rao, Troy Kelley, Dave Braines, Murat Sensoy, Chris Willis, and Prudhvi Gurram. Interpretability of deep learning models: A survey of results. In *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1–6. IEEE, 2017.

[34] Xiaojun Chen, Shengbin Jia, and Yang Xiang. A review: Knowledge reasoning over knowledge graph. *Expert systems with applications*, 141:112948, 2020.

[35] Gobinda G Chowdhary. Natural language processing. *Fundamentals of artificial intelligence*, pages 603–649, 2020.

[36] Michael Chromik and Andreas Butz. Human-XAI interaction: A review and design principles for explanation user interfaces. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II 18*, pages 619–640. Springer, 2021.

[37] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. What makes a good conversation? challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12, 2019.

[38] Natalie Clewley, Tim Forsyth, Lorraine Dodd, and Jeremy Hilton. A framework for systems thinking practice. In *OR64: OR for a better world together: The Operational Research Society's Annual Conference*. Cranfield University, 2022.

[39] R Jordon Crouser and Remco Chang. An affordance-based framework for human computation and human-computer collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2859—-2868, 2012.

[40] Missy Cummings. Man vs. machine or man + machine? *IEEE Intelligent Systems*, 29(5):62–69, 2014.

[41] Missy Cummings. *Artificial Intelligence and the Future of Warfare*. Chatham House for the Royal Institute of International Affairs London, 2017.

[42] Praveen Damacharla, Ahmad Y Javaid, Jennie J Gallimore, and Vijay K Devabhaktuni. Common metrics to benchmark human-machine teams (HMT): A review. *IEEE Access*, 6:38637–38655, 2018.

[43] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 135–144, 2017.

[44] Stephen L Dorton and Robert A Hall. Collaborative human-AI sensemaking for intelligence analysis. In *International conference on human-computer interaction*, pages 185–201. Springer, 2021.

[45] Dstl. Human-centred ways of working with AI in intelligence analysis. `https://www.gov.uk/government/publications/human-centred-ways-of-working-with-ai-in-intelligence-analysis`, 2023. Accessed: 2024-05-10.

[46] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, et al. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9):1–33, 2023.

[47] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. In *Situational awareness*, pages 9–42. Routledge, 2017.

[48] Dieter Fensel, Umutcan Şimşek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, Alexander Wahler, Dieter Fensel, et al. Introduction: What is a knowledge graph? *Knowledge graphs: Methodology, tools and selected use cases*, pages 1–10, 2020.

[49] Katalin Ferencz and József Domokos. Using Node-RED platform in an industrial environment. *XXXV. Jubileumi Kandó Konferencia, Budapest*, pages 52–63, 2019.

[50] Bruna Ferreira, Williamson Silva, Edson Oliveira, and Tayana Conte. Designing personas with empathy map. In *SEKE*, volume 152, 2015.

[51] Kenneth Forbus, Emmett Tomai, and Jeffrey Usher. Qualitative spatial reasoning for visual grouping in sketches. In *Proceedings of the 17th International Workshop on Qualitative Reasoning*, pages 131–138. Brasilia, Brazil, 2003.

[52] William W Gaver. Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 79–84, 1991.

[53] J. J. Gibson. *The senses considered as perceptual systems.* Houghton Mifflin, 1966.

[54] Michael Glassman and Min Ju Kang. Intelligence in the internet age: The emergence and evolution of open source intelligence (OSINT). *Computers in Human Behavior*, 28(2):673–682, 2012.

[55] Vinicius G Goecks, Nicholas Waytowich, Derrik E Asher, Song Jun Park, Mark Mittrick, John Richardson, Manuel Vindiola, Anne Logie, Mark Dennison, Theron Trout, Priya Narayanan, and Alexander Kott. On games and simulators as a platform for development of artificial intelligence for command and control. *The Journal of Defense Modeling and Simulation*, page 15485129221083278, 2022.

[56] US Government. A tradecraft primer: Structured analytic techniques for improving intelligence analysis. CIA Center for the study of intelligence, `https://www.stat.berkeley.edu/~aldous/157/Papers/Tradecraft%20Primer-apr09.pdf`, 2009. Accessed: 2024-05-10.

[57] Herbert P Grice. Logic and conversation. In *Speech acts*, pages 41–58. Brill, 1975.

[58] Jia-Chen Gu, Zhen-Hua Ling, and Quan Liu. Interactive matching network for multi-turn response selection in retrieval-based chatbots. In *proceedings of the 28th ACM international conference on information and knowledge management*, pages 2321–2324, 2019.

[59] Prudhvi Gurram, Suproyo Chakraborty, Richard Tomsett, Raghuveer Rao, and Lance Kaplan. Uncertainty in explanations. In *Proceedings of the 5th Annual Fall Meeting (AFM) of the Distributed Analytics and Information Science International Technology Alliance (DAIS ITA)*, Sep 2020.

[60] Andrea L Guzman. What is human-machine communication, anyway? *Human-machine communication: Rethinking communication, technology, and ourselves*, pages 1–28, 2018.

[61] Daniel Harborne, Chris Willis, Richard Tomsett, and Alun Preece. Integrating learning and reasoning services for explainable information fusion. In *International Conference on Pattern Recognition and Artificial Intelligence (CPRAI)*, May 2018.

[62] Barbara Hayes-Roth. A blackboard architecture for control. *Artificial intelligence*, 26(3):251–321, 1985.

[63] Sam Hepenstal, Leishi Zhang, and BL William Wong. An analysis of expertise in intelligence analysis to support the design of human-centered artificial intelligence. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 107–112. IEEE, 2021.

[64] Richards J Heuer. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, 1999.

[65] Adrian James. *Understanding Police Intelligence Work*, volume 2. Policy Press, 2016.

[66] Sérgio Jesus, Catarina Belém, Vladimir Balayan, João Bento, Pedro Saleiro, Pedro Bizarro, and João Gama. How can I choose an explainer? an application-grounded evaluation of post-hoc explanations. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 805–815, 2021.

[67] Ulla Johansson-Sköldberg, Jill Woodilla, and Mehves Çetinkaya. Design thinking: Past, present and possible futures. *Creativity and innovation management*, 22(2):121–146, 2013.

[68] Mark St John and Harvey S Smallman. Staying up to speed: Four design principles for maintaining and recovering situation awareness. *Journal of Cognitive Engineering and Decision Making*, 2(2):118–139, 2008.

[69] Audun Jøsang. *Subjective Logic: A Formalism for Reasoning Under Uncertainty.* Springer Publishing Company, Incorporated, 2018.

[70] Daniel Kahneman. *Thinking, Fast and Slow.* Macmillan, 2011.

[71] Harmanpreet Kaur, Eytan Adar, Eric Gilbert, and Cliff Lampe. Sensible AI: Re-imagining interpretability and explainability using sensemaking theory. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 702–714, 2022.

[72] Gary Klein. Naturalistic decision making. *Human factors*, 50(3):456–460, 2008.

[73] Gary Klein, Jennifer K Phillips, Erica L Rall, and Deborah A Peluso. A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making*, pages 113–155. New York, NY: Lawrence Erlbaum Assoc Inc, 2007.

[74] Gary A Klein. A recognition-primed decision (RPD) model of rapid decision making. *Decision making in action: Models and methods*, 5(4):138–147, 1993.

[75] Brandy Klug. An overview of the system usability scale in library website and system usability testing. *Weave: Journal of Library User Experience*, 1(6), 2017.

[76] Anna Knack, Richard Cater, and Alexander Bubuta. Human-machine teaming in intelligence analysis: Requirements for developing trust in machine learning systems. Report, Centre for Emerging Technology and Security, 2022.

[77] Alistair Knott. Job sharing between human professionals and chatbots: How should 'handovers' happen? *Journal of AI, Robotics & Workplace Automation*, 2(2):121–130, 2023.

[78] Margarita Konaev and Husanjot Chahal. Building trust in human-machine teams, Feb 2021.

[79] Stefan Kopp and Nicole Krämer. Revisiting human-agent communication: The importance of joint co-construction and understanding mental states. *Frontiers in Psychology*, 12:580955, 2021.

[80] Cynthia F Kurtz and David J Snowden. The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM systems journal*, 42(3):462–483, 2003.

[81] Jieun Kwon, Younghyun Choi, and Yura Hwang. Enterprise design thinking: An investigation on user-centered design processes in large corporations. *Designs*, 5(3):43, 2021.

[82] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[83] Anthony Liew. Understanding data, information, knowledge and their inter-relationships. *Journal of knowledge management practice*, 8(2):1–16, 2007.

[84] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018.

[85] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. SSD: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[86] Ewa Luger and Abigail Sellen. "Like having a really bad PA" The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5286–5297, 2016.

[87] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[88] Joseph B Lyons, Kevin T Wynne, Sean Mahoney, and Mark A Roebke. Trust and human-machine teaming: A qualitative study. In *Artificial intelligence for the internet of everything*, pages 101–116. Elsevier, 2019.

[89] Jiefei Ma, Alessandra Russo, Krysia Broda, and Keith Clark. DARE: A system for distributed abductive reasoning. *Autonomous Agents and Multi-Agent Systems*, 16(3):271–297, 2008.

[90] Yogesh Malhotra. Expert systems for knowledge management: crossing the chasm between information processing and sense making. *Expert Systems with Applications*, 20(1):7–16, 2001.

[91] Matthew Marge, Claire Bonial, Stephanie Lukin, and Clare Voss. Bot language (Summary technical report, Oct 2016–Sep 2021). `https://apps.dtic.mil/sti/trecms/pdf/AD1194759.pdf`, 2023. Accessed: 2024-05-10.

[92] Patricia McDermott, Cindy Dominguez, Nicholas Kasdaglis, Matthew Ryan, Isabel Trhan, and Alexander Nelson. Human-machine teaming systems engineering guide. Technical report, Mitre Corp, Bedford MA, United States, 2018.

[93] Lucas Memmert and Eva Bittner. Complex problem solving through human-AI collaboration: Literature review on research contexts. In *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022.

[94] Declan Millar, Dave Braines, Erik Blasch, Doug Summers-Stay, and Iain Barclay. Semantically-guided acquisition of trustworthy data for information fusion. In *2021 IEEE 24th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2021.

[95] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[96] UK Ministry of Defence. Joint doctrine publication 04 (JDP 04) Understanding and decision making. `https://www.gov.uk/government/publications/jdp-04-understanding`, 2010. Accessed: 2024-05-10.

[97] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4):1–45, 2021.

[98] David Mott, Cheryl Giammanco, Michael C Dorneich, and Dave Braines. Hybrid rationale for shared understanding. In *The Fourth Annual Conference of the International Technology Alliance. London, England*, pages 13–17, 2010.

[99] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. Patterns for how users overcome obstacles in voice user interfaces. In

Proceedings of the 2018 CHI conference on human factors in computing systems, pages 1–7, 2018.

[100] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, 1994.

[101] Don Norman. *The Design of Everyday Things: Revised and expanded edition*. Basic books, 2013.

[102] Alistair Nottle, Daniel Harborne, Dave Braines, Moustafa Alzantot, Santiago Quintana-Amate, Richard Tomsett, Lance Kaplan, Mani B Srivastava, Supriyo Chakraborty, and Alun Preece. Distributed opportunistic sensing and fusion for traffic congestion detection. In *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1–6. IEEE, 2017.

[103] Chris Olah, Arvind Satyanarayan, Ian Johnson, Shan Carter, Ludwig Schubert, Katherine Ye, and Alexander Mordvintsev. The building blocks of interpretability. *Distill*, 3(3):e10, 2018.

[104] Pablo Ortega, Richard Tomsett, and Lance Kaplan. Evidential deep learning with recurrent neural networks. In *Proceedings of the 5th Annual Fall Meeting (AFM) of the Distributed Analytics and Information Science International Technology Alliance (DAIS ITA)*, Sep 2020.

[105] Frans PB Osinga. *Science, Strategy and War: The Strategic Theory of John Boyd*. Routledge, 2007.

[106] Steven Ovadia. Markdown for librarians and academics. *Behavioral & social sciences librarian*, 33(2):120–124, 2014.

[107] DJ Patil. Building for the enterprise - the zero overhead principle. `https://techcrunch.com/2012/10/05/building-for-the-enterprise-the-zero-overhead-principle-2/`, 2012. Accessed: 2024-05-10.

[108] Cathy Pearl. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, Inc., 2016.

[109] Gavin Pearson and Tien Pham. The challenge of sensor information processing and delivery within network and information science research. In *Defense Transformation and Net-Centric Systems 2008*, volume 6981, pages 34–42. SPIE, 2008.

[110] Tien Pham, Gregory Cirincione, Ananthram Swami, Gavin Pearson, and Christopher Williams. Distributed analytics and information science. In *2015 18th International Conference on Information Fusion (Fusion)*, pages 245–252. IEEE, 2015.

[111] P Jonathon Phillips, Carina A Hahn, Peter C Fontana, David A Broniatowski, and Mark A Przybocki. Four principles of explainable artificial intelligence. *Gaithersburg, Maryland*, 2020.

[112] Robert Pienta, James Abello, Minsuk Kahng, and Duen Horng Chau. Scalable graph exploration and visualization: Sensemaking challenges and opportunities. In *2015 International conference on Big Data and smart computing (BIGCOMP)*, pages 271–278. IEEE, 2015.

[113] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, 2005.

[114] Alun Preece, Dave Braines, Federico Cerutti, Jack Furby, Liam Hiley, Lance Kaplan, Mark Law, Alessandra Russo, Mani Srivastava, Marc Roig Vilamala, and Tianwei Xing. Coalition situational understanding via explainable neuro-symbolic reasoning and learning. In *Artificial Intelligence and*

Machine Learning for Multi-Domain Operations Applications III, volume 11746, pages 453–464. SPIE, 2021.

[115] Alun Preece, Federico Cerutti, Dave Braines, Supriyo Chakraborty, and Mani Srivastava. Cognitive computing for coalition situational understanding. In *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1–6. IEEE, 2017.

[116] Alun Preece, Chris Gwilliams, Christos Parizas, Diego Pizzocaro, Jonathan Z Bakdash, and Dave Braines. Conversational sensing. In *Next-Generation Analyst II*, volume 9122, pages 165–178. SPIE, 2014.

[117] Alun Preece and Winston R Sieck. The international technology alliance in network and information sciences. *IEEE Intelligent Systems*, 22(5):18–19, 2007.

[118] Alun Preece, Will Webberley, and Dave Braines. Conversational sensemaking. In *Next-Generation Analyst III*, volume 9499, pages 121–129. SPIE, 2015.

[119] Alun Preece, Will Webberley, and Dave Braines. Tasking the tweeters: Obtaining actionable information from human sensors. In *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VI*, volume 9464, page 946402. SPIE, 2015.

[120] Alun Preece, William Webberley, Dave Braines, Erin G Zaroukian, and Jonathan Z Bakdash. SHERLOCK: Experimental evaluation of a conversational agent for mobile information tasks. *IEEE Transactions on Human-Machine Systems*, 47(6):1017–1028, 2017.

[121] Aung Pyae and Tapani N Joelsson. Investigating the usability and user experiences of voice user interface: A case of google home smart speaker. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct*, pages 127–131, 2018.

[122] Roy Radner. Satisficing. In *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974 6*, pages 252–263. Springer, 1975.

[123] Kellyn Rein and Joachim Biermann. Beyond situation awareness: Considerations for sense-making in complex intelligence operations. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 1859–1865. IEEE, 2018.

[124] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[125] Chet Richards. Boyd's OODA loop. *Necesse*, 5:142–165, 2020.

[126] Avi Rosenfeld and Ariella Richardson. Explainability in human-agent systems. *Autonomous Agents and Multi-Agent Systems*, 33(6):673–705, 2019.

[127] Donald Rumsfeld. *Known and unknown: A memoir*. Penguin, 2011.

[128] Stuart J Russell. *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc., 2010.

[129] Arthur L Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.

[130] Jeff Sauro. Measuring usability with the system usability scale (SUS), May 2016.

[131] Florian Schaurer and Jan Störger. The evolution of open source intelligence. Report 3, Association of Former Intelligence Officers, 2010.

[132] Annika Schoene and Geeth de Mel. Pooling tweets by fine-grained emotions to uncover topic trends in social media. In *2019 22nd International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE, 2019.

[133] Jan Maarten Schraagen, Susan F Chipman, and Valerie L Shalin. *Cognitive Task Analysis*. Psychology Press, 2000.

[134] James Self, Mark Evans, and Eun Jin Kim. A comparison of digital and conventional sketching: Implications for conceptual design ideation. *Journal of Design Research*, 14(2):171–202, 2016.

[135] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, pages 3179–3189, 2018.

[136] Priti Shah and Akira Miyake. *The Cambridge Handbook of Visuospatial Thinking*. Cambridge University Press, 2005.

[137] Elisa Shahbazian, Dale E Blodgett, and Paul Labbé. The extended OODA model for data fusion systems. *Fusion*, 2001.

[138] Jennifer E Sims. Decision advantage and the nature of intelligence analysis. *Oxford Handbook on National Security Intelligence*, 2010.

[139] Paul R Smart, Ali Bahrami, Dave Braines, Duncan McRae-Spencer, Jun Yuan, and Nigel R Shadbolt. Semantic technologies and enhanced situation awareness. In *The First Annual Conference of the International Technology Alliance. London, England*, 2007.

[140] Paul R Smart, Trung Dong Huynh, David Mott, Katia Sycara, Dave Braines, Michael Strub, Winston Sieck, and Nigel R Shadbolt. Towards an

understanding of shared understanding in military coalition contexts. In *3rd Annual Conference of the International Technology Alliance (ACITA'09)*, 2009.

[141] Paul R Smart, David Mott, Katia Sycara, Dave Braines, Michael Strub, and Nigel R Shadbolt. Shared understanding within military coalitions: A definition and review of research challenges. In *Knowledge Systems for Coalition Operations, Southampton, UK*, 2009.

[142] Lauro Snidaro, J Garcia-Herrera, James Llinas, and Erik Blasch. Context-enhanced information fusion. *Boosting Real-World Performance with Domain Knowledge*, 2016.

[143] David J Snowden and Mary E Boone. A leader's framework for decision making. *Harvard business review*, 85(11):68, 2007.

[144] Keith Stenning and Michiel van Lambalgen. A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, 28(4):481–529, 2004.

[145] Niranjan Suri, Kelvin M Marcus, Casper van den Broek, Harrie Bastiaansen, Piotr Lubkowski, and Mariann Hauge. Extending the Anglova scenario for urban operations. In *2019 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–7. IEEE, 2019.

[146] Harrison Taylor, Liam Hiley, Jack Furby, Alun Preece, and Dave Braines. VADR: Discriminative multimodal explanations for situational understanding. In *23rd International Conference on Information Fusion (FUSION)*, Jul 2020.

[147] George Terzopoulos and Maya Satratzemi. Voice assistants and smart speakers in everyday life and in education. *Informatics in Education*, 19(3):473–490, 2020.

[148] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(5):675–691, 2005.

[149] Richard Tomsett, Dave Braines, Daniel Harborne, Alun Preece, and Supriyo Chakraborty. Interpretable to whom? a role-based model for analyzing interpretable machine learning systems. *ICML Workshop on Human Interpretability in Machine Learning (WHI 2018)*, 2018.

[150] Richard Tomsett, Alun Preece, Dave Braines, Federico Cerutti, Supriyo Chakraborty, Mani Srivastava, Gavin Pearson, and Lance Kaplan. Rapid trust calibration through interpretable and uncertainty-aware AI. *Patterns*, 1(4):100049, 2020.

[151] Paddy Turner and Lorraine Dodd. Developing the cognitive and social aspects of military 'understanding capability'. In *Proceedings of the 21st International Command and Control Research Technology Symposium*, 2016.

[152] Sam Vente, Angelika Kimmig, Alun Preece, and Federico Cerutti. Increasing negotiation performance at the edge of the network. In *Proceedings of AT2020*, 2020.

[153] Marc Roig Vilamala, Dave Braines, Federico Cerutti, and Alun Preece. Visualizing logic explanations for social media moderation. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pages 3056–3058, 2023.

[154] Paul Watzlawick, Janet Beavin Bavelas, and Don D Jackson. *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes.* WW Norton & Company, 2011.

[155] Jing Wei, Tilman Dingler, and Vassilis Kostakos. Understanding user perceptions of proactive smart speakers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–28, 2021.

[156] Karl E Weick, Kathleen M Sutcliffe, and David Obstfeld. Organizing and the process of sensemaking. *Organization science*, 16(4):409–421, 2005.

[157] John Wenskovitch, Corey Fallon, Kate Miller, and Aritra Dasgupta. Beyond visual analytics: Human-machine teaming for AI-driven data sensemaking. In *2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)*, pages 40–44. IEEE, 2021.

[158] Michel Wermelinger. Using GitHub Copilot to solve simple programming problems. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pages 172–178, 2023.

[159] Graham White, Simon Pierson, Brian Rivera, Maroun Touma, Paul Sullivan, and Dave Braines. DAIS-ITA scenario. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006, pages 505–524. SPIE, 2019.

[160] Simon Williams and Susan Michie. Covid-19: One year on from "Freedom Day," what have we learnt?, 2022.

[161] Tianwei Xing, Marc Roig Vilamala, Luis Garcia, Federico Cerutti, Lance Kaplan, Alun Preece, and Mani Srivastava. DeepCEP: Deep complex event processing using distributed multimodal information. In *IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 87–92. IEEE, 2019.

[162] Chong Ho Yu. Abduction? deduction? induction? Is there a logic of exploratory data analysis? Technical report, University of Oklahoma, 1994.

[163] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. MetaGraph2Vec: Complex semantic path augmented heterogeneous network embedding. In *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part II 22*, pages 196–208. Springer, 2018.

Appendices

Appendix A

Details of the Cogni-sketch environment

This appendix contains various details about the design and implementation of the Cogni-sketch environment, including links to external resources such as videos which may be of interest to the reader.

A.1 Examples of Cogni-sketch usage

The most substantial and formally evaluated examples of Cogni-sketch usage can be found in Chapters 5 and 6 of this thesis. However, the scope of HAKF applications is broader than just sensemaking or SU and the goal is to support a wide range of collaborative behaviours. Cogni-sketch has been used in a wide variety of use cases and reported in various publications. Those examples not covered in Chapters 5 and 6 are therefore briefly described in the following subsections to show the wider variety of uses for the Cogni-sketch environment which have been carried out by the author and other collaborators who are using the platform.

There is also a fourth example which was created but was not published as the DAIS ITA programme finished shortly after. It focuses on the use of Cogni-sketch as an environment to explore the processing of ML models as would typically be done in an environment such as a Jupyter notebook but a technical ML expert user. Because it was not hardened into an example suitable for publication is it is not listed here as one of the main examples, but a video demonstration is available, see video V15 in Section A.3.

A.1.1 Science library

For both the NIS ITA [117] from 2006 to 2016, and the DAIS ITA [110] from 2016 to 2021 we curated a *science library* of all publications from these long-running research programs, containing publications from hundreds of researchers. The science library itself was a stand-alone application developed separately to Cogni-sketch and predating it significantly. However, the effort required to maintain these science libraries was not in the publication of each paper but in the tracking of papers through their life cycle from draft to submitted, to accepted, and then to published. This, plus management of the provenance for each paper, such as a link to the published version, confirmation of attribution to the program and checking that all authors and organisations had been accurately captured were key to the stakeholder confidence in the accuracy of the data. Maintaining these details was time-consuming and complex to manage between the small group of administrators who ran these science libraries. In 2021 we created a small set of plugins and a custom palette to enable the provenance of each publication to be collaboratively tracked and agreed between the administrators, as well as defining the metadata for each publication so that the science library record for each could be automatically generated. This substantially reduced the admin burden and automated the process of publication as reported in [28]. It is also a nice example for the extensibility of Cogni-sketch with a custom palette, visualisations, forms, and a directed machine agent.

Figure A.1 shows an example of a simple network visualization for a paper in the Cogni-sketch interface. This paper can be created in a number of ways, for example: by using the custom form to create the graph by automatically parsing and analysing the PDF and seeking user confirmation for all of the values that are extracted according to the schema for a paper node within the palette, or by manually drawing the components using the palette on the left-hand side of the Cogni-sketch interface, or by editing the underlying spreadsheet that can be imported (or any combination of these).

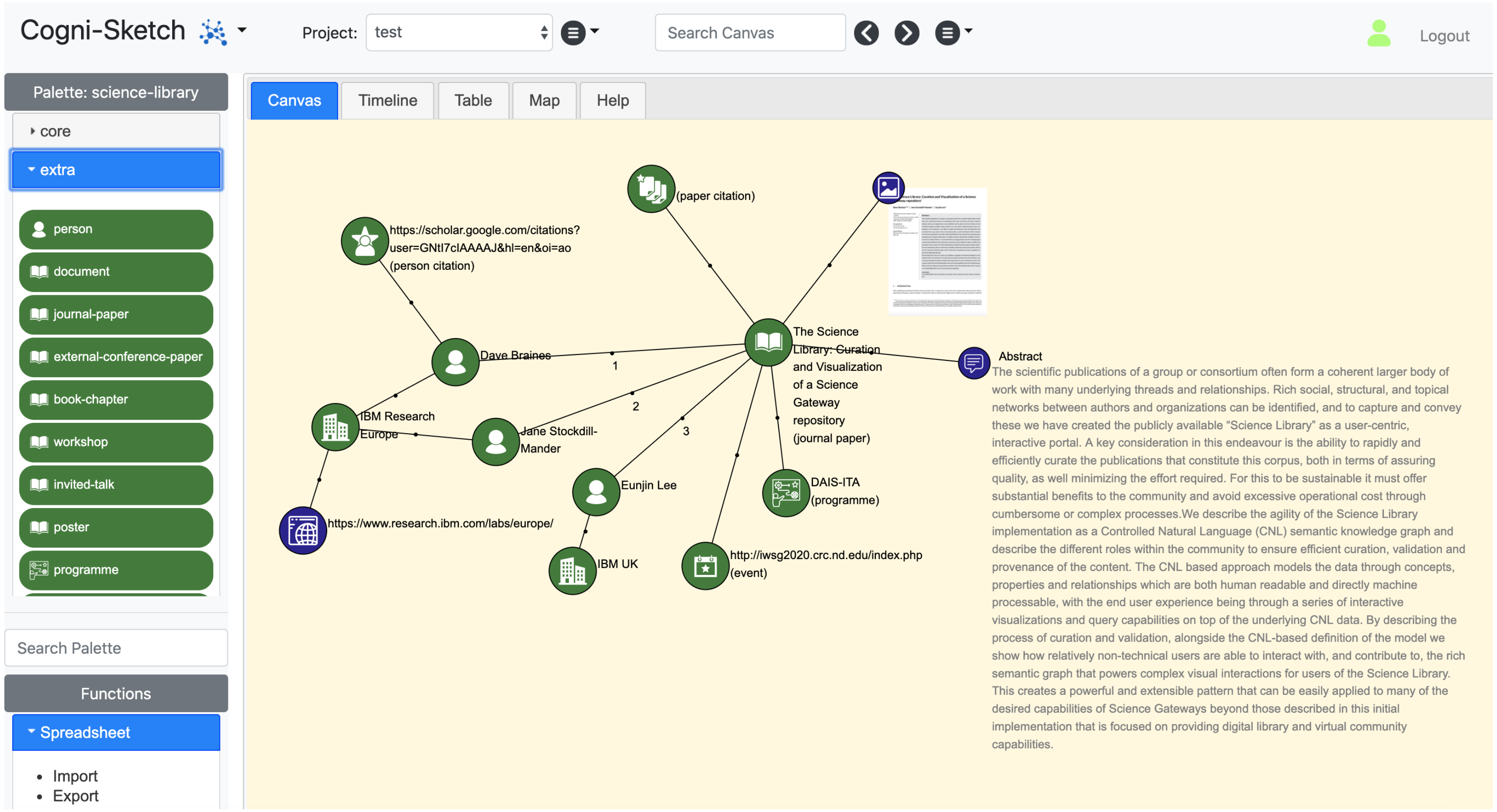


Figure A.1: Network visualization of an example paper and related material

The administrators can explore the publication network as well as other views such as showing all the publications for an author, including material that has not yet been pushed to public science library. They can also attach notes and comments, and share these with other administrator users, as well as paste in material from external sources such as email correspondence with the authors. This can be useful in validating any questions relating to the publications that are confirmed directly by the authors. All this additional supporting information is available within the Cogni-sketch environment, but is not published to the public science library, nor included in the generated knowledge graph; it is for the administrators to document and maintain their provenance for the papers.

By creating this simple but highly extensible environment for collaborative administration of the science library publications we have enabled a substantial improvement in the efficiency of the administrators, enabling them to publish updates more regularly and spend less time on onerous manual tasks such as seeking citation data and cross-validation with citation databases. The flexibility and extensibility of the Cogni-sketch environment helps to retain the overall goal of agility in conceptual model extensions in the future, albeit with the caveat that the custom forms will require modifications if the conceptual model is extended (or the additional information required must be captured manually using the generic tabular or graphical interfaces).

A.1.2 Plutchik's wheel of emotions

In December 2021, during the Covid-19 pandemic we organised a Christmas hack-day to help with team morale and to give people a 2-day period in which to try a new idea in a team working with people they don't usually work with. The Cogni-sketch platform was suggested as an interesting dynamic environment in which to attempt a real-time rendering of the emotional content of text in the form of Plutchik emotional representation, and specifically an animated version of Plutchik's Wheel [132] based on previous work from the team. This was implemented as a standard machine agent within the Cogni-sketch environment which was built entirely within the two-day period and successfully demonstrated working on a variety of example text sources within the Cogni-sketch environment at the end of the event. Once built it was trivial to apply the agent to any text to see the results.

See Section A.3 (video V12) for a video demonstration of this agent dynamically rendering Plutchik emotional data for different textual content within the Cogni-sketch environment.

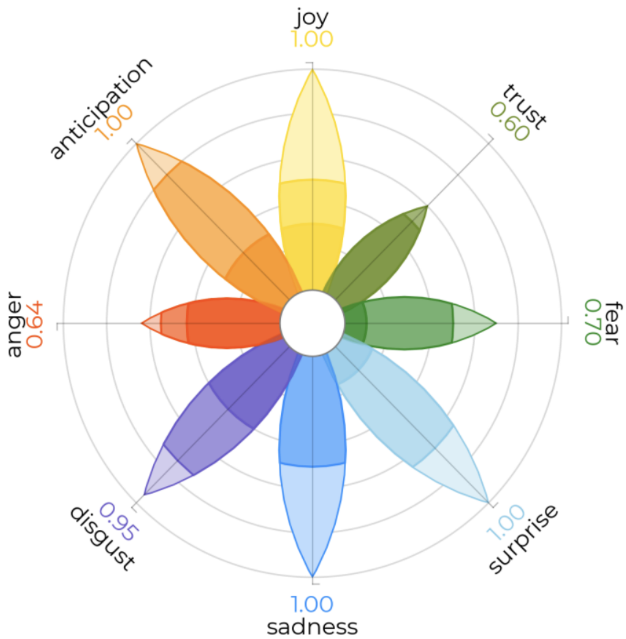


Figure A.2: Typical two-dimensional visualisation of Plutchik's wheel

This was able to be achieved with a small team of four researchers in a two-day period with no modifications to the core Cogni-sketch code base, with much of the time spent creating the Plutchik analysis and animated visualisation, and relatively little time spent on Cogni-sketch integration. By creating a very simple machine agent plugin that could be dropped onto any existing text node it would dynamically generate an animated Plutchik wheel which would pulse to show the emotion as an animation of the text that flowed through. Because of the open and extensible canvas pane for Cogni-sketch this could be run in parallel on multiple bodies of text at the same time to see a comparison of textual emotion. For example, we showed the difference between the two versions of the moon landing text prepared for the U.S. president to read after the attempted landing in 1969, with one for success and the other lesser known one for failure.

This example is included here both because it provides an unusual example of a rich emotion processing capability exposed into an environment such as this, but also because it was able to be achieved in just two days with no modifications to the Cogni-sketch core code and is a good example of the intended ease and agility with which such agents can be created.

A.1.3 Meaningful paths in semantic vectors

Related work in the DAIS ITA research program was investigating the role of semantic vectors in helping users to understand complex graphs. In this work we proposed that significant value can be added by annotating *meaningful paths* through a knowledge graph so that when skip-gram based heterogeneous graph embedding algorithms, such as MetaPath2Vec [43] and MetaGraph2Vec [163], are used to compute the semantic vector space, particular emphasis can be placed on the paths through the graph that are meaningful based on the user-defined annotations that reflect what the graphs represent. These approaches require explicitly defined meta-paths as input to inform the heterogeneity-aware biased random-walks used by these models. Identifying these paths is a task much better suited to human annotation than machine processing since for any typical knowledge graph there will be many thousands of permutations of possible paths and only a small subset of these will be meaningful.

By using the Cogni-sketch environment we were able to define a simple custom palette to allow the schema for a knowledge graph to be drawn, with all possible links between different node types as shown in the top half of Figure A.3. The story pane plugin that was developed for the OSINT analysis experiment reported in Chapter 6 was then used to enable the user to pick a subset of nodes and links and provide a textual description for each, as shown in the lower part of Figure A.3.

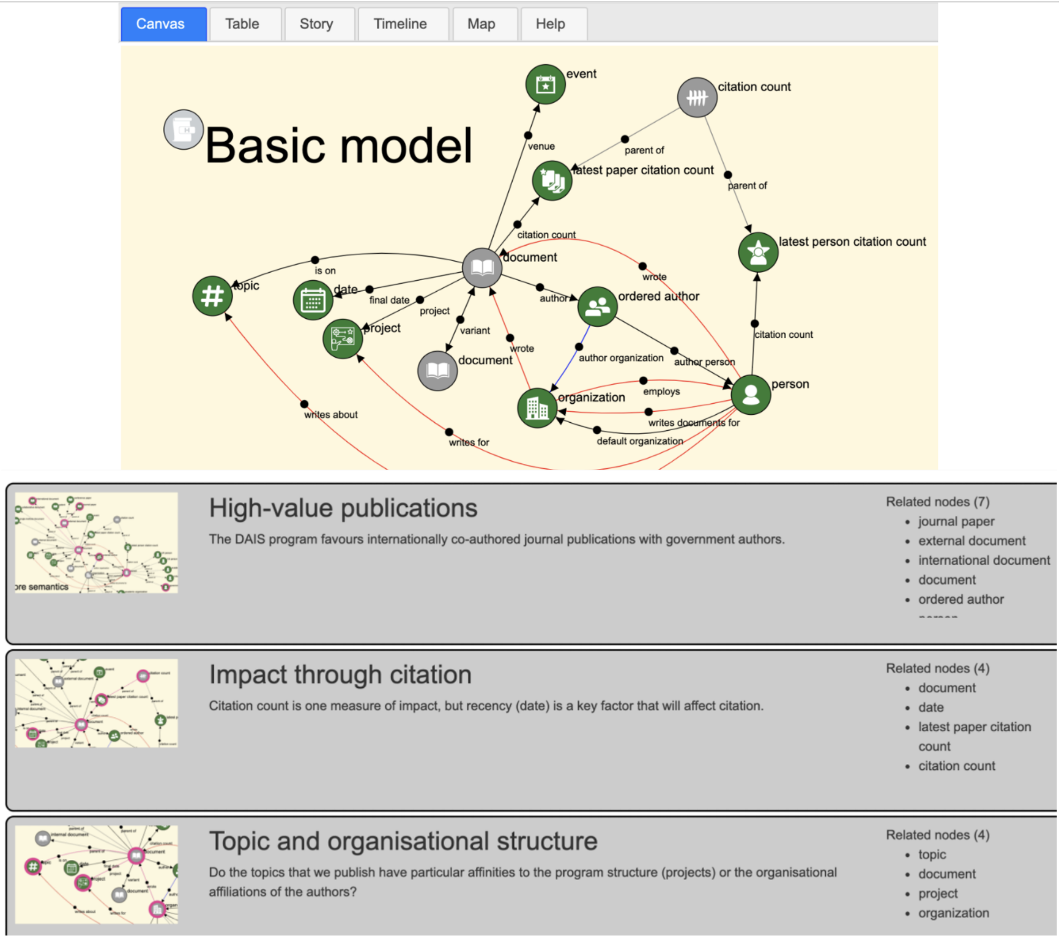


Figure A.3: Defining meaningful semantic vector paths using Cogni-sketch.

The results from this exercise are published in [94] and show how Cogni-sketch can be successfully used for defining the meaningful paths through an existing knowledge graph so that skip-gram based heterogeneous graph embedding algorithms can be used. These abilities were able to be achieved without any extensions to the core Cogni-sketch code and successfully reused the story pane to explain which nodes and links are relevant, reusing a pattern created to support storytelling, but which was equally applicable to this use case.

A.2 Data

The data for each of the experiments and examples described within this thesis can be found at `https://bit.ly/cogni-sketch-data`. For the experiment results each of the created knowledge graphs for the twelve anonymised participants can be downloaded and easily loaded into a Cogni-sketch environment by importing a project from the project menu and specifying the JSON file for that participant as the project definition. These specific steps should be followed:

- Navigate to the public data folder (`https://bit.ly/cogni-sketch-data`).
- Go to the *OSINT_experiment* folder.
- Download the **palette** JSON file in this folder and import it into Cogni-sketch (assuming you haven't already done so).
- Identify and download the user project zip file that you wish to load (some users created multiple projects during the experiment).
- Load the project into Cogni-sketch by importing it.
  You can do this either with the whole zip file, or with just the main project. JSON file after extracting it from the zip file. It is better to import the whole zip file as you will also get all images, files and other attachments that were created during the exercise.

The other data located in this folder relates to the pilot and evaluation described in Chapter 5 and the various videos listed later in this appendix. Details of the OSINT pilot are not shared, as explained in Section 6.3.

A.3 Video demonstrations

This section lists all the publicly available video demonstrations in the order they were created, along with a link to the video on YouTube or elsewhere. Some of these videos are mentioned as footnotes throughout the thesis where they correspond to the text and provide useful additional (optional) context. They were a useful mechanism for sharing updates with interested parties during the Covid-19 restrictions. The creation date and duration (in mm:ss) are also shown.

- V1 Cogni-sketch 1 - introduction [06:12] - `https://www.youtube.com/watch?v=KmaheXO6D9M` (27-Apr-2020).
- V2 Cogni-sketch 2 - functions and files [06:48] - `https://www.youtube.com/watch?v=wAzjZeG3jWc` (27-May-2020).
- V3 Cogni-sketch 3 - semantics [08:00] - `https://www.youtube.com/watch?v=GOXGj_Dcvfw` (8-Jul-2020).
- V4 Cogni-sketch 4 - chat [12:43] - `https://www.youtube.com/watch?v=FHOff6S2-NY` (23-Jul-2020).
- V5 Human-Agent Knowledge Fusion at AAAI FSS 2020 [20:26] - `https://www.youtube.com/watch?v=kZ3YE6bxGJM` (13 Nov 2020).
- V6 Cogni-sketch latest progress [12:17] - `https://www.youtube.com/watch?v=Hi7uXXqTJg8` (2-Feb-2021).
- V7 Cogni-sketch story telling [04:27] - `https://www.youtube.com/watch?v=jNGE737n3RA` (19-Apr-2021).
- V8 Cogni-sketch for sensemaking and Intelligence Analysis [05:57] - `https://www.youtube.com/watch?v=i6PudsmgRaw` (26-May-2021).
- V9 Dave Braines PhD Research Retreat 2021 [06:40] - `https://www.youtube.com/watch?v=iFQxAozOCzY` (28-Jun-2021).
- V10 FUSION 2021 - Supporting Agile User Fusion Analytics through Human-Agent Knowledge Fusion [14:58] - `https://www.youtube.com/watch?v=QiCOYQywjs8` (21-Oct-2021).

V11 Cogni-sketch experiment guide [06:32] - https://www.youtube.com/watch?v=zDSeWRxdLPw (18-Jan-2022).

V12 Plutchik emotion dynamic graphs [01:11] - https://youtu.be/ksJi92aShes (5-Apr-2023).

V13 Enabling rapidly formed human-agent coalition teams through extensible information exchange [13:15] - https://dais-legacy.org/1c01/ (10-Aug-2021).

V14 Adapting AI systems to recognise new patterns of distributed activity [15:23] - https://dais-legacy.org/1c16/ (10-Aug-2021).

V15 AI based Analysis of Terrorism Events [15:47] - https://dais-legacy.org/3a13/ (10-Aug-2021).

A.4 Cogni-sketch plugins

This section lists all the plugins currently developed for the Cogni-sketch environment. Those that have already been released as open-source software are indicated (using *) and can be found on GitHub at https://github.com/dais-ita/cogni-sketch-plugins. Most others are generically useful and can be open sourced in the future once additional documentation has been created.

- cogni-sketch-contrib-ce
- cogni-sketch-contrib-chat
- cogni-sketch-contrib-complex-event-processing
- cogni-sketch-contrib-exif
- cogni-sketch-contrib-google-scholar
- cogni-sketch-contrib-json

- cogni-sketch-contrib-language-translate
- cogni-sketch-contrib-map (*)
- cogni-sketch-contrib-media
- cogni-sketch-contrib-nlp
- cogni-sketch-contrib-nlu
- cogni-sketch-contrib-object-detection
- cogni-sketch-contrib-pdf
- cogni-sketch-contrib-plutchik
- cogni-sketch-contrib-rdf
- cogni-sketch-contrib-science-library
- cogni-sketch-contrib-sl-generator
- cogni-sketch-contrib-spreadsheet
- cogni-sketch-contrib-story
- cogni-sketch-contrib-timeline
- cogni-sketch-contrib-twitter (*)
- cogni-sketch-contrib-visual-recognition

A.5 Storytelling

In supporting the OSINT analyst during the sensemaking pilot the story plugin was created. The text in this section was shared with them to explain the capability and capture possible future extensions:

There is a new optional pane named ‘story’ that allows you to write narrative text around groups or collections of existing nodes. There can be any number of these story nodes created on your canvas and they are the ‘story elements’ or ‘story frames’ that define the story. The story pane provides a custom view for interacting with these where they can be re-ordered and better visualised. Each story element has a label, and description, an image, and a list of nodes that they relate to (all these properties are optional).

There is a short (4 minute) YouTube video outlining the story capability, `https://youtu.be/jNGE737n3RA`

When you first start you won’t have the new ‘story’ type defined in your palette, but when you first visit the story pane there is a link there to create that for you:

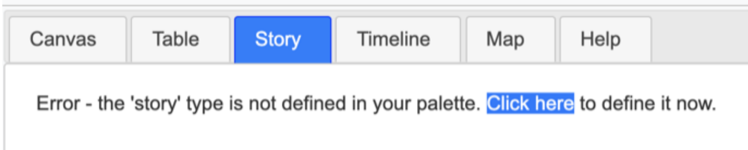


Figure A.4: Undefined story node type

(Note that you cannot modify the default palette, so you must first create your own palette before you can create the new story type).

When that is done you get the new story type appearing at the bottom of your ‘core’ palette section:

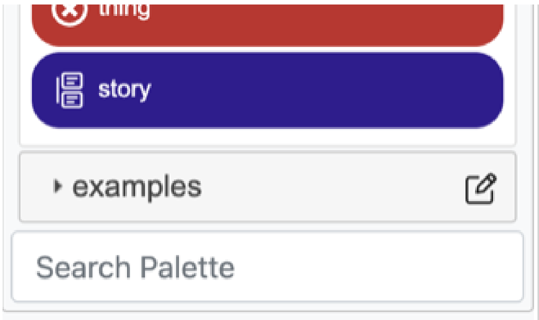


Figure A.5: Story node shown in the palette

You can now create stories in two different ways:

1. By dragging out the new story node and dropping it onto the canvas and editing/linking to other nodes in the usual way.

2. By using the new story pane.

Users already familiar with the Cogni-sketch environment should know that this is a 'normal' palette type and can be used in the same ways as other palette types.

Creating story elements

On the story pane you can click on 'Add new story frame' to create a new story frame on this page:

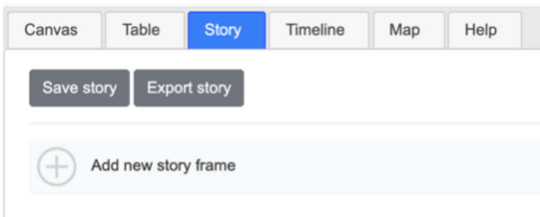


Figure A.6: Correctly activated story pane

This creates a new empty story element in edit mode for you to provide the label, description and image. If you have nodes selected on the canvas when you create a new story element, then these will be listed as related nodes in the new story element. A new story element with a label and description, an image and some related nodes is shown below:

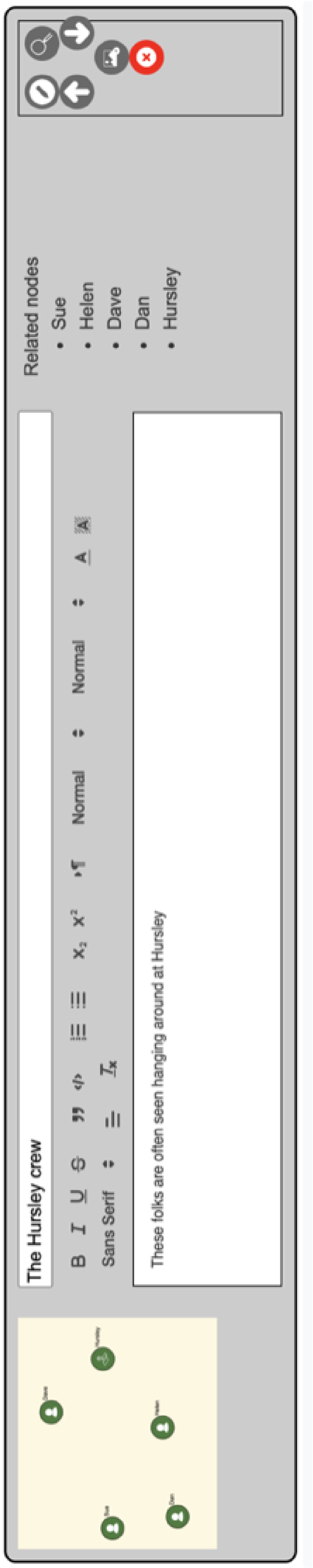


Figure A.7: An example story element

At this point the story is not saved, so if you navigate back to the canvas, you won't see any changes, but if you click on 'Save story' then this new story node is created and linked to all the related nodes:

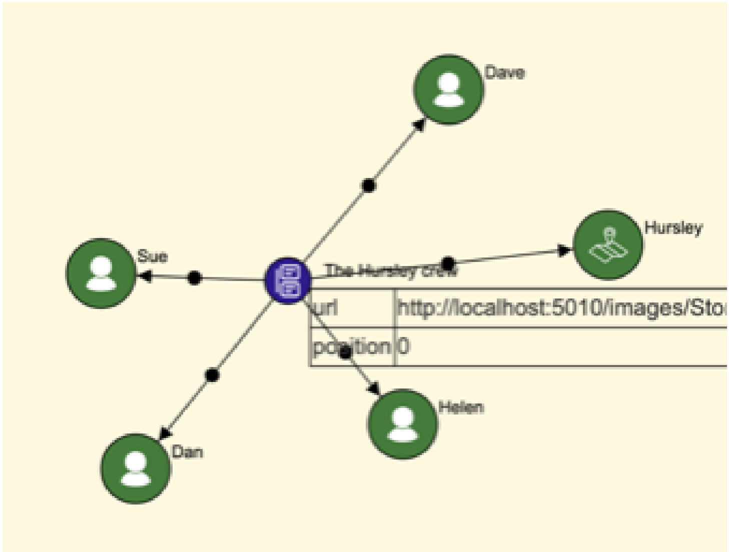


Figure A.8: Story element shown as raw nodes on the canvas

If you add or remove links, then the story pane will be updated to reflect the current state of the canvas. Links can only be made on the canvas, not in the story pane.

You can create as many story elements as you like and provide as much or little information for each as is required. You can click on the names of the related nodes to be shown them on the canvas, and the magnifying glass icon for the story element itself.

For each story element you can switch between view and edit mode and use the arrow buttons to move the story element up or down in the narrative flow. You can also delete the story element, which has the same effect as deleting the node on the canvas.

Thoughts and future upgrades

This is an initial implementation and can be substantially extended if needed. The following are short term low-cost changes that could easily be added:

1. Allow images to be uploaded (using the upload button) or dropped onto the story element. Currently images can only be pasted.
2. Better synchronisation between the canvas and story pane. Currently the story pane can get out of sync with edits made on the canvas, but it is easily fixed by saving the project and reloading the browser page.
3. Allow user suppression of properties in the table on the canvas. It is not that useful to see the URL and position values for the story nodes and they should be suppressed.
4. Better centralisation of the story node between the nodes that it connects.
5. Complete the 'Export to MS-Word' function.

It may also be possible to achieve some of the following if desired:

1. Automatic generation of the image, based on copying the relevant part of the canvas that contains all the selected nodes.
2. Story sections – enabling multiple stories to be shown on the story pane.
3. Enable non-story nodes to be included in the story pane, e.g., if you have already created a node that summarises everything about a person or other entity then it might be nice to include that directly in the story rather than copying the information out into a new story node.
4. Allow user configuration of what properties are summarised on the story pane, enabling them to add their own properties to the story nodes and have them appear in the story pane summary.

Appendix B

Open source sensemaking experiment supporting information

This appendix contains supporting information relevant to the open source sensemaking experiment described in Section 6.4.

B.1 Participation guide

The material in this section is a verbatim copy of the participation guide that was given to each of the participants in the sensemaking experiment.

Please read this document thoroughly as it defines the scope and resources for this exercise. The exercise is **limited to 2 hours**, and you can take as many breaks as needed. The purpose is to measure the ability to find and capture relevant knowledge and insights, building artefacts to answer the question(s) defined below. Please be available from [**hh:mm**] – [**hh:mm**] on [**date**]. You need an **internet connection** and **google chrome** running on a **desktop** (not touchscreen).

Resources

The following resources may be useful during this exercise:

- The live Cogni-sketch environment - `https://dais-cogni-sketch.org/` [1]

  Your username and password will be provided separately.

- Support for questions/advice during this exercise (use whichever work best for you):

  – Webex - `https://ibm.webex.com/join/dave_braines`
  (for live video chat or screen sharing).

  This will be open for the duration of the experiment and you can optionally join if needed.

  – Slack - `https://slack-s0y1245.slack.com/archives/C02PPAT266T`
  (for text interactions).

  This is a dedicated private channel, not visible to anyone except you and the experiment lead.

Background

Monday 19th July 2021 was declared "Freedom Day" by the U.K. government, with almost all COVID-19 restrictions being lifted in England. A social media collection task was initiated from 27th June to 29th July 2021, capturing U.K. twitter activity from verified users related to the term "mask". This gathered over 18,000 tweets which are available on the Cogni-sketch "twitter" pane, with the ability explore, filter and generate different charts, and add any of these easily to the canvas. These tweets plus the wider public discourse through news articles, web pages and other materials give an insight into what the public was experiencing during this period.

Task

Figure 6.10 shows the frequency of verified Twitter user tweets during this period with two clear spikes: the 5th July, and the 19th July. The latter spike, on

[1] This website was live for the duration of the experiment but is now offline.

"Freedom Day", was anticipated, but the first spike was not.

Please explore public data sources to answer these questions:

1. What caused the spike in volume for "mask" related discussion in the U.K. on social media on Monday 5th July?
2. How do the factors driving this first spike feature throughout the time period?

Please record all relevant information and supporting evidence in the Cogni-sketch environment, creating nodes and links as required, and using whatever media is needed (text, imagery, video, web pages etc).

Cogni-sketch capabilities

The Cogni-sketch environment provides a broad set of simple capabilities to enable you to capture and link information that is relevant to your investigation.

There is a short (6 minute) overview video here - `https://youtu.be/GsNq0EBpimU`

The capabilities you are likely to need during this exercise are:

- General:
  - The main canvas drawing area.

    Nodes can be created, edited, extended and linked together.
  - Projects.

    Any number of projects can be saved and switched between using the drop-down menu.
  - Search.

    A simple keyword search for all nodes and links on the canvas.
  - Files and Functions.

    Not used in this exercise, but any files you save during the exercise will be listed.

  - Table and Map tabs.

    Generic tabs for showing data in tabular and spatial (map) format.

  - Help tab

    Lots of useful guidance for the Cogni-sketch environment.

- Palette:

  - A predefined set of node types.
  - The ability to add new palette sections and/or palette items.

    Into the existing palette, or by saving a new palette copy.

- Creating nodes:

  - Drag and drop

    e.g., browser pages or local files. The "best fit" palette item will be chosen.

  - Copy and paste

    e.g., text, images, screen capture images etc. The "best fit" palette item will be chosen.

  - Drag from palette

    This creates a new empty node of that type on the canvas.

- Creating links - *Holding the shift key whilst dragging a node starts a new link. Dropping this link over the target node will create a new link. Double clicking the blob at the center of the link allows you to edit to add a label or properties.*

- Twitter tab - *This allows you to explore the mask tweets sourced from twitter. You can filter and sort by various criteria and see various data types (tweets, Twitter users, hashtags, URLs etc) as well as different predefined charts. All items can easily be saved to the canvas via drop-down menus, along with the ability to save the current query as a canvas node for easy recreation of that query.*

- Story tab - *Useful to provide a narrative summary of your graph, creating a linear summary of your findings. You can add nodes from the canvas here, choosing which properties to highlight, adding descriptions as needed, and choosing the order of the story.*

- Keyboard shortcuts (The full list is defined on the help tab, but these may be useful):

  - Ctrl +/- *to zoom in/out (you can also zoom in/out with scroll wheel or 2-fingers on touchpad).*
  - Arrows - left / right / up /down *to scroll the canvas (you can also scroll by dragging the canvas).*
  - Ctrl e or o *to zoom to show all nodes (e), or to reset to the original zoom size (o).*
  - Backspace or delete *to delete selected nodes and links.*
  - Ctrl a *to select all nodes + links (or use the shift key to select in a rectangle, or click nodes individually).*
  - Ctrl d *to duplicate selected nodes and links.*
  - Ctrl s *to save the canvas and palette.*

B.2 Event type to sensemaking behaviour category mapping

Section 6.2.3 describes the mapping of Cogni-sketch event types to four sensemaking categories. The full list of all event types and their mapping to the corresponding category is shown here for completeness.

Event	**Category**	**Description**

canvas:abandonPartialLink	ignore	User chooses not to complete a link, by not dropping it onto a suitable node.
canvas:collapseOrExpand	build	User chooses to expand or collapse one or more nodes on the canvas.
canvas:deleteSelected	collect	User chooses to delete one or more nodes from the canvas.
canvas:deselectNode	collect	User deselects a node on the canvas.
canvas:duplicate	collect	User duplicates one or more nodes on the canvas.
canvas:editLink	build	User double clicks on a link to edit the properties.
canvas:editNode	build	User double clicks on a node to edit the properties.
canvas:finishPartialLink	build	User completes the creation of a new link by dropping it onto a valid node.
canvas:linkAnchorMoved	build	User drags a link anchor to move it along the link line.
canvas:linkBent	build	User drags a link anchor to bend the link line.
canvas:pan	ignore	User panned (scrolled) the canvas using mouse or keys.
canvas:pasteImage	collect	User pasted an image from their clipboard onto the canvas.
canvas:pasteText(canvas)	collect	User pasted text from their clipboard onto the canvas.

canvas:pasteText(failed)	collect	User attempted to paste text from their clipboard onto the canvas, but it was rejected because the selected node already had text.
canvas:pasteText(node)	collect	User pasted text from their clipboard onto an existing node on the canvas.
canvas:selectNode	collect	User selects a node on the canvas.
canvas:startPartialLink	build	User starts to draw a partial link from a node on the canvas.
canvas:zoom	ignore	User zooms in or out on the canvas using mouse or keys.
canvas:zoomToOriginal	ignore	User chooses to zoom the canvas to the original starting position and zoom.
canvas:zoomToSelected	ignore	User chooses to zoom the canvas, so the selected nodes fill the space.
chat:addNodesToCanvas	collect	User asks for nodes to be created on the canvas from the chat interaction.
chat:receivedAnswer	collect	A message is received from the chat application.
chat:sendMessage	collect	A message is sent to the chat application.
chat:showHelpAnswer	navigate	The specific "help" chat message is shown to the user.
chat:showWelcome	navigate	The "welcome" chat message is sent to the user.

chat:showWelcomeBack	navigate	The “welcome back” chat message is sent to the user.
createEmpty	collect	User creates a new empty node on the canvas.
createFull	collect	User creates a new full node on the canvas.
cs:changePane	navigate	User changes the main pane of the user interface.
drop:browser	collect	User drops a URL from another browser/tab onto the canvas, creating a node, or onto an existing node.
drop:imageFile	collect	User drops a file from their local environment onto the canvas, creating a node, or onto an existing node.
qdrop:palette	collect	User drops a palette item onto the canvas, creating a new node of that type.
link:addTextProperty	build	User adds a new text property to an existing link.
link:changedBidirectional	build	User changes an existing link to be bi-directional.
moveNode	build	User moves a node on the canvas using mouse or keys.
nlp:addDoughnutToCanvas	collect	User adds a specific doughnut chart from the twitter pane to the canvas.

nlp:addTimelineToCanvas	collect	User adds a specific timeline chart from the twitter pane to the canvas.
nlp:addToCanvas	collect	User adds a tweet element (URL/user/tweet etc) to the canvas.
nlp:changedDateFilter	collect	User changes the date filter on the twitter pane.
nlp:changedDoughnut	collect	User changes the doughnut chart type on the twitter pane.
nlp:changedIgnoreRemainder	collect	User changes the "ignore remainder" mode for the date filter on the twitter pane.
nlp:changedIncludeEarlier	collect	User changes the "include earlier" mode for the date filter on the twitter pane.
nlp:changedShowLegend	collect	User changes the "show legend" settings for the chart capability on the twitter pane.
nlp:changedTimeline	collect	User changes the timeline chart type on the twitter pane.
nlp:changedTypeFilter	collect	User changes the tweet type filter on the twitter pane.
nlp:doQuery	collect	User executes a query with the specified filters on the twitter pane.

nlp:doSlice	collect	User moves to the next or previous page of results for any entity on the twitter pane.
nlp:doSort	collect	User sorts the list of any entity by any relevant feature on the twitter pane.
nlp:filterByHashtag	collect	User adds a hashtag filter on the twitter pane.
nlp:filterByMention	collect	User adds a mention filter on the twitter pane.
nlp:filterBySymbol	collect	User adds a symbol filter on the twitter pane.
nlp:filterByUrl	collect	User adds a URL filter on the twitter pane.
nlp:filterByUser	collect	User adds a twitter user filter on the twitter pane.
nlp:filterToRetweets	collect	User adds a filter to show only retweets from a specific twitter user on the twitter pane.
nlp:invertedFilter	collect	User inverts any filter on the twitter pane.
nlp:keywordSearch	collect	User adds a keyword search filter on the twitter pane.
nlp:openInTwitter	collect	User chooses to open a tweet in twitter from the twitter pane.
nlp:openUrlInBrowser	collect	User chooses to open a URL in a new browser window from the twitter pane.

nlp:removedFilter	collect	User removes any filter from the twitter pane.
nlp:saveQuery	build	User saves a specific query from the twitter tab to the canvas.
nlp:showSentiment	collect	User toggles the show/hide sentiment option on the twitter pane.
node:addNormalProperty	build	User adds a normal property to a node on the canvas.
node:addTextProperty	build	User adds a text property to a node on the canvas.
node:changedPropertyType	build	User changes the type of an existing node property.
node:changedPropertyValue	build	User changes the value of an existing node property.
node:removeProperty	build	User removes an existing property from a node on the canvas.
node:showType	build	User toggles the “show type” property for a node on the canvas.
palette:cancelSectionPopup	ignore	User cancels the section popup for the palette.
palette:cancelTypePopup	ignore	User cancels the type popup for the palette.
palette:changePalette	build	User switches to a different palette.
palette:createNewPalette	build	User creates a new palette based on the currently selected palette.
palette:editItem: changed:iconAlt	build	User changes the icon name for an existing type on the palette.

palette:editItem: changed:name	build	User changes the name of an existing type on the palette.
palette:editItem: changed:schema(new)	build	User adds schema information to an existing type on the palette.
palette:editItem: changed:schema(type)	build	User edits schema information for an existing type on the palette.
palette:endAddItem	build	User completes the creation of a new type on the palette.
palette:endEditItem	build	User completes the edit of an existing type on the palette.
palette:filterPalette	navigate	User specifies search terms to filter the palette.
palette:startAddItem	build	User begins the process of adding a new item to the palette.
palette:startEditItem	build	User begins the process of editing an item on the palette.
palette:startEditSection	build	User begins the process of editing a section on the palette.
popup:closePopup	ignore	Any popup window is closed either by saving changes or cancelling.
project:changeProject	navigate	User switches to a different project.
project:reload	navigate	User reloads the current project.
project:resetCanvasPosition	navigate	User resets the canvas position to the original starting point.
project:save	ignore	User manually saves the project.
search:search	collect	User uses local search to search the nodes and links on the canvas.

search:next	collect	User selects the next matching node or link in the local search.
story:addNewFrame	story	User adds a new frame to their story.
story:deleteFrame	story	User deletes a frame from their story.
story:editStoryFrame	story	User edits an existing story frame within their story.
story:findLinkOnCanvas	navigate	User follows a link from their story frame to the location on the canvas.
story:findNodeOnCanvas	navigate	User follows a node from their story frame to the location on the canvas.
story:saveStory	story	User saves their story.
story:showOnCanvas	story	User clicks to show a story frame on the canvas, by finding and highlighting all related nodes and links.
story:updateRelatedNodes	story	User edits a story frame and updates the list of related nodes and links based on those selected on the canvas.
updateLinkLabel	build	User updates the label for a link.
updateNode	build	User updates the properties of a node.
updateNodeLabel	build	User updates the label for a node.
updateNodeType	build	User changes the type of a node.

updateToEmpty	build	User removes a property from a node so that it is reverted to "empty".
updateToFull	build	User adds a property to a node so that it becomes a "full" node.

Table B.1: Cogni-sketch event type mapping to sensemaking categories

B.3 Participant canvases and stories

This section contains a brief qualitative analysis of the canvas and associated story components for each participant that participated in the experiment. A screenshot of the canvas and story components are also included for each participant should the reader wish to correlate the summary provided here with the raw output the participants created during the experiment. Note that the included screenshots are as large and high-resolution as possible but due to the large number of nodes and links created, it is not always possible to read the full details of each user canvas within this thesis document[2].

[2]Full extracts of the data created by each user within the experiment are available as high-quality images for the canvas and story elements as well as in JSON format to enable the projects from the participants to be easily recreated, see Section A.2 for details of how to access this data.

B.3.1 Participant 01

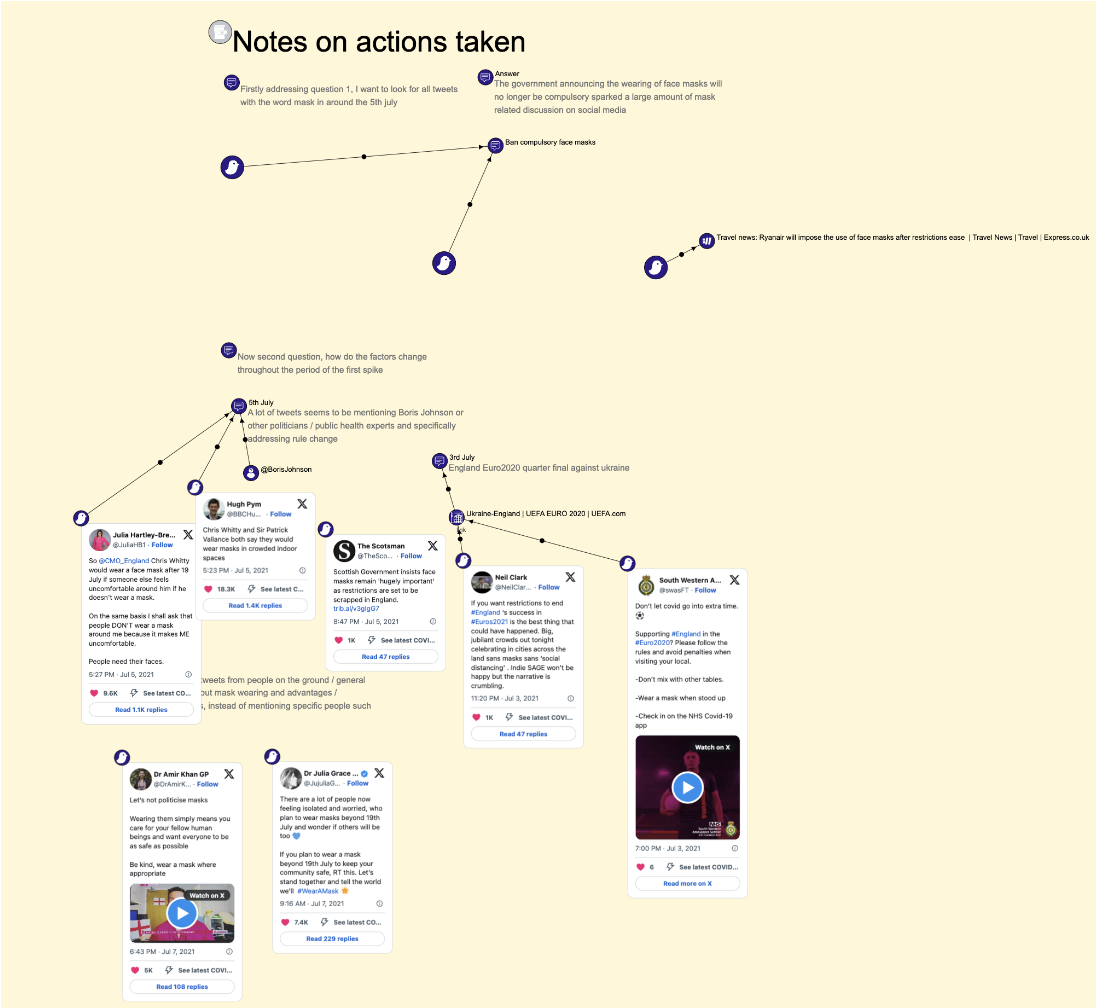


Figure B.1: Canvas for participant 01

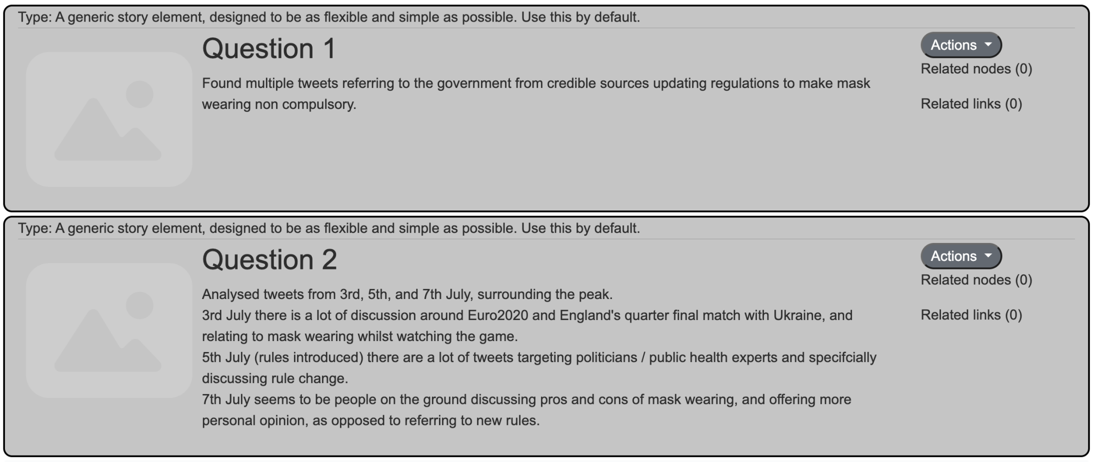


Figure B.2: Story for participant 01

Participant 01 was able to identify that the answer to Q1 was the government announcement of Freedom Day and identified tweets and possible trends over time to answer Q2. They created a separate project for their findings but created their story in the first project, meaning that they could not relate canvas nodes and links to their story. Not all material from the canvas was captured in the story, e.g., the "answer" created on the canvas for Q1. Several tweets were identified and clustered into groups to answer Q2.

B.3.2 Participant 02

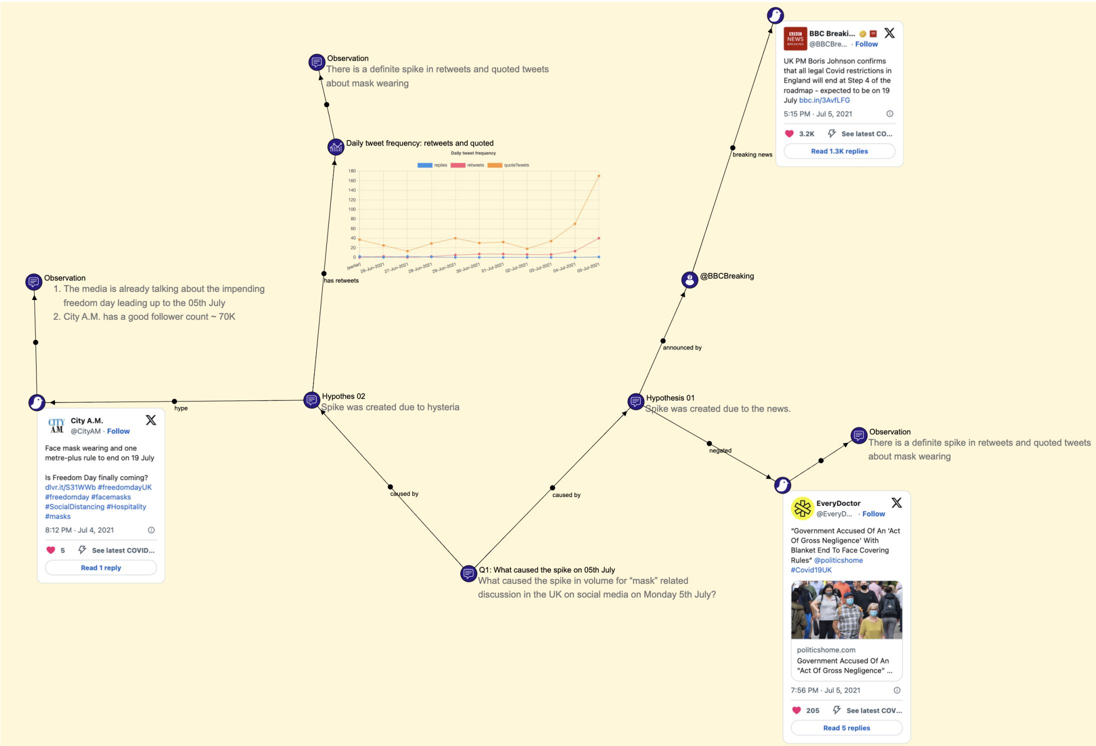


Figure B.3: Canvas for participant 02

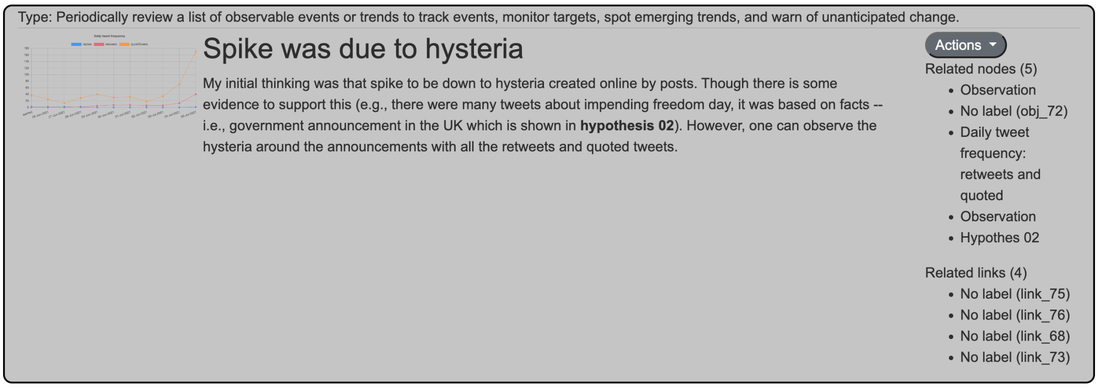


Figure B.4: Story for participant 02

Participant 02 created a relatively small number of nodes (11) and links (10) and laid them out on their canvas in a simple style. They did label each of the links and included 3 tweet nodes and 1 dynamic chart of daily tweet frequency. They created a single story element which asserted their hypothesis that the spike was perhaps due to hysteria, but they concluded in the summary that rather than hysteria it was an increased volume of social media activity based on media reporting of government facts rather than other sources. They did not attempt to answer Q2, but they did link the relevant nodes and links to their story element for Q1.

Good use of layout to follow the flow of different hypotheses from Q1. Only a single story frame created, but this links to the relevant canvas objects and summarises the thinking that led to the conclusion

B.3.3 Participant 03

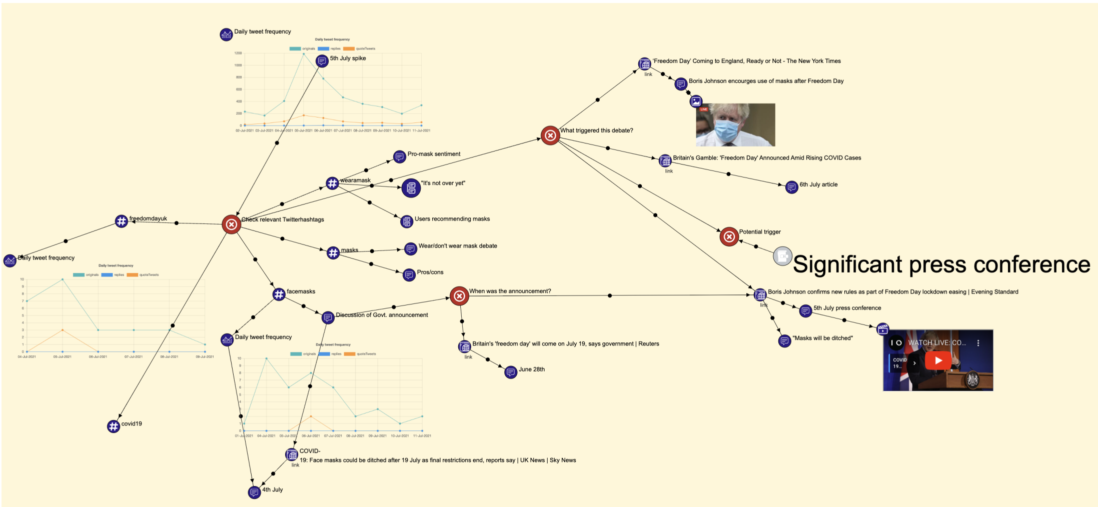


Figure B.5: Canvas for participant 03

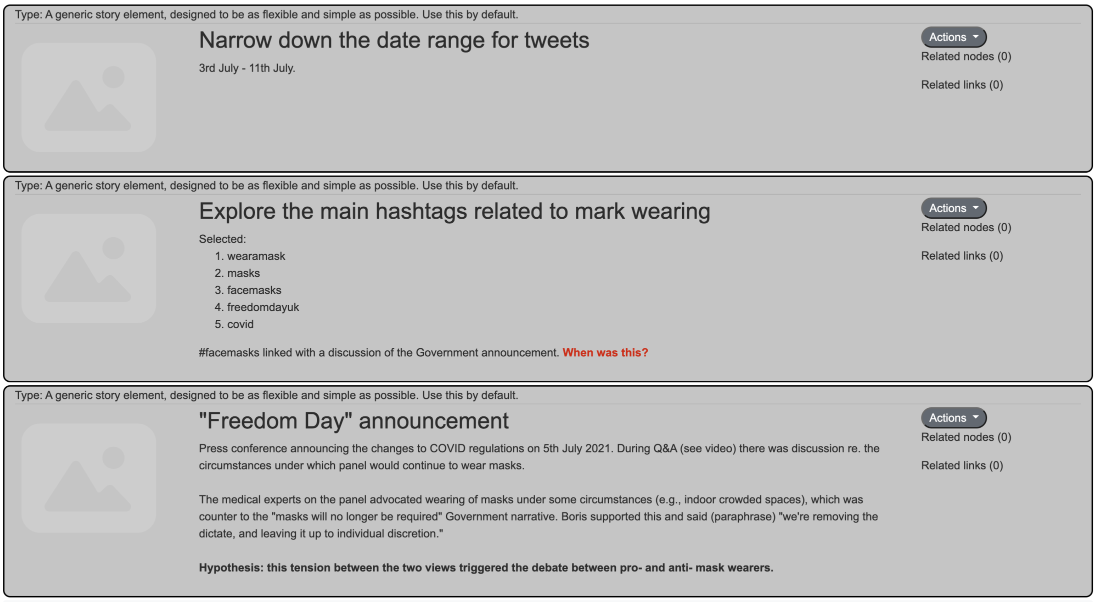


Figure B.6: Story for participant 03

Participant 03 created 33 nodes and 33 links on their canvas and laid these out as a single graph with different clusters within the overall single graph structure. They used the “unknown” node type to capture 4 different nodes that represent questions or goals for themselves in this exercise, and these are created throughout the time period of the exercise rather than clustered at the beginning, middle or end, reflecting their ongoing investigation process. This user included a number of external news articles as well as a directly embedded video of the 5th July press conference from YouTube. They also included three dynamic charts for the overall tweet frequency as well as specific hashtags. Whilst they created a lot of links between nodes these were not labelled.

This participant reported their thought process via the three story elements that they created, with the last element being their summary description of the media reporting, followed by subsequent narrative contrasting pro- and anti- mask positions. They did not attempt to answer Q1 or Q2 directly in the story, but the content of the final story element is clearly in support of Q1, albeit not explicitly stated or conclusively answered.

Very structured canvas layout with nodes and links recording the path of their investigation. Multiple chart nodes created, showing tweet frequencies on different days as well as relevant video and image content added to the canvas. They used the “unknown” palette item to record key questions and investigation points. No tweet nodes were created, suggesting that they were looking at larger patterns rather than the impact of individual tweets

B.3.4 Participant 04

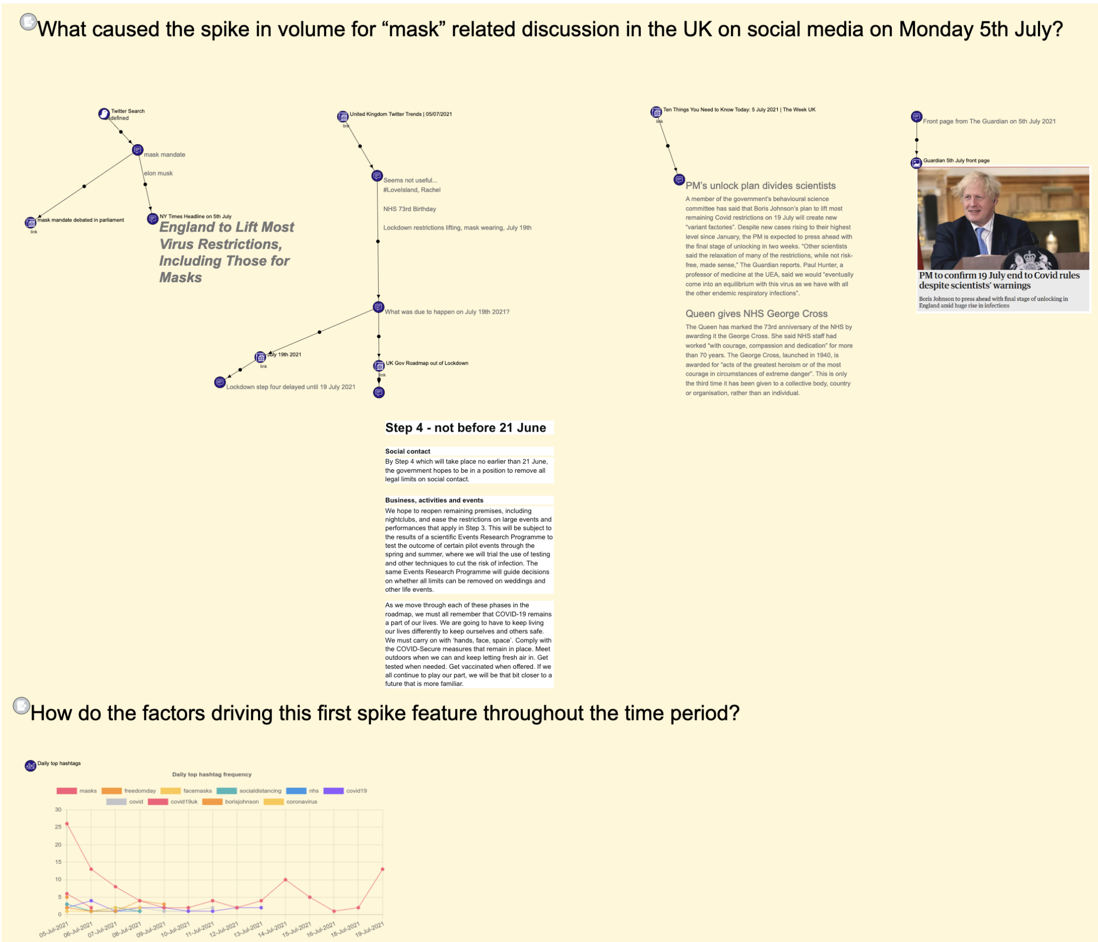


Figure B.7: Canvas for participant 04

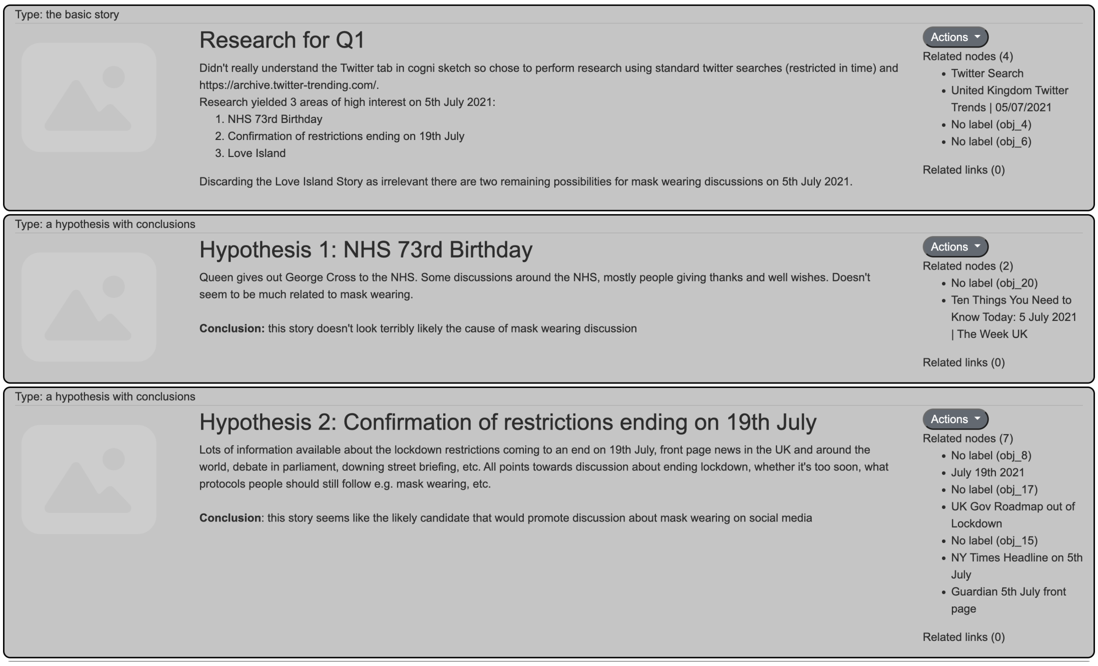


Figure B.8: Story for participant 04 (1 of 2)

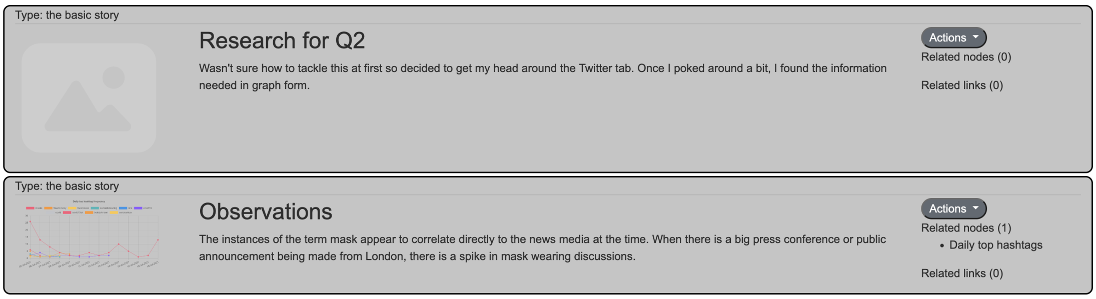


Figure B.9: Story for participant 04 (2 of 2)

Participant 04 created a strong example of a narrative story, stating the research questions and their hypotheses and then following these, linking to the relevant nodes and links from the canvas. They created 18 nodes and 11 links and laid their canvas out as a series of small graph fragments in a tabular style. Some links were labelled, and a mixture of text and image content was captured. The two target questions were captured on the canvas as header nodes, with the relevant graph fragments located under these. Only a single dynamic chart node was created (for Q2), and no tweets or other dynamic content were included[3], but some external links plus text copied from them were captured as nodes on the canvas. This participant concludes that the early spike is driven by media discussion of the government announcement and discounts the possibility of the NHS 73rd birthday being the cause (and discounts another prominent hashtag relating to Love Island[4]). In their second story element they investigate and discount the NHS 73rd birthday as being a substantial cause for the spike. In their third story element they investigate the media articles relating to the relaxing of restrictions and conclude that it is these announcements and coverage that drives the spike indicated by Q1.

In their 4th story element, they attempt to answer Q2 and fed back that at this stage they persevered in understanding how the embedded Twitter pane worked and found a graph that provides the answer, which they include on the canvas and then link to in the final story element. In this story element they feedback about the approach taken and note a direct correlation between Twitter activity volume and various press conferences or public announcements throughout that period, noting that the answer to Q2 is therefore the media articles seeding and driving the social media activity.

In summary: A simple layout chosen in the canvas pane, with small clusters of linked nodes aligned to each question. A lot of the linkages between and across the nodes are explained via the story frames instead of being explicitly linked on the canvas.

[3]This is explained in their first story node where they state that they could not understand the embedded Twitter analysis pane, so instead reverted to simple external web and Twitter searches.

[4]Since this participant was searching all of Twitter and not the subset related to “mask” tweets on the Twitter pane they were seeing additional topics like Love Island that the other participants would not have seen.

B.3.5 Participant 05

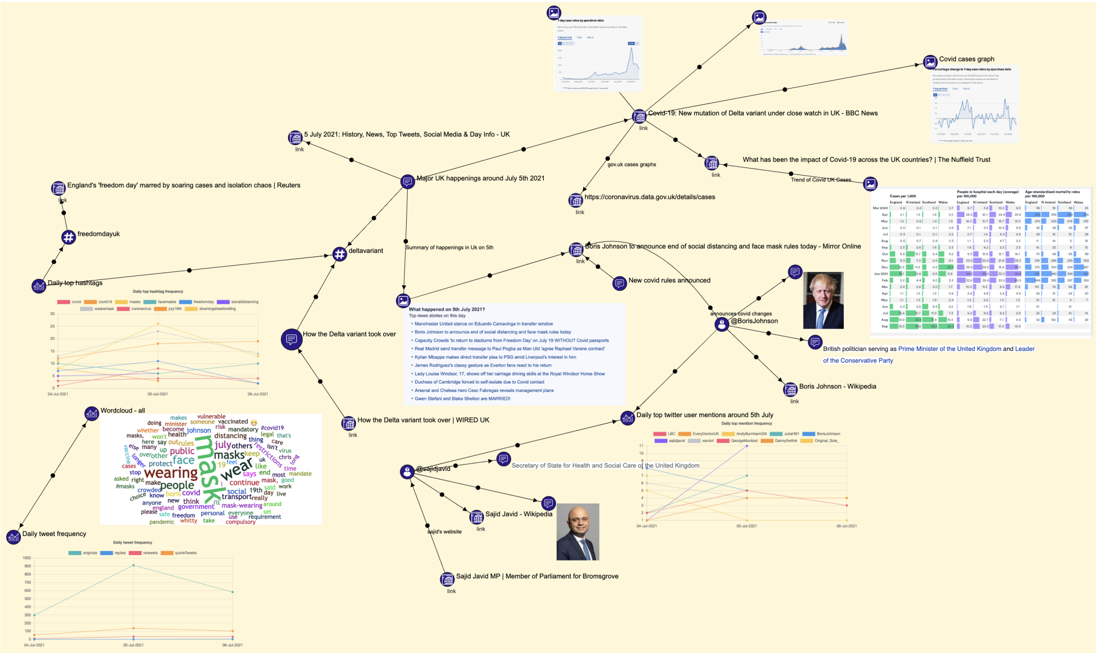


Figure B.10: Canvas for participant 05

Figure B.11: Story for participant 05

Participant 05 created a reasonable number of nodes (30) and links (28). They drew extensively on external sources such as mainstream media articles and online medical resources as well as some images from web searches which they copied onto the canvas. They created two small clusters on the canvas for two individuals (with associated images, Twitter accounts and summary descriptions) via the chat interface. They created a small number of text nodes with some descriptive text to annotate some of their findings, mainly around the Delta variant, although this could be content sourced from the internet and copied into this text node. They also extensively labelled their links between nodes rather than leaving them unlabelled.

This participant created three story elements and used images from the canvas for two of these to help illustrate the story, but they did not link any nodes or links to the story nodes. In the first story element they noted that Delta cases were rising in early July, whereas in the second story element they noted that a major press announcement was made by the Prime Minister on 5^{th} July. The third story element simply merged and summarised both points.

B.3.6 Participant 06

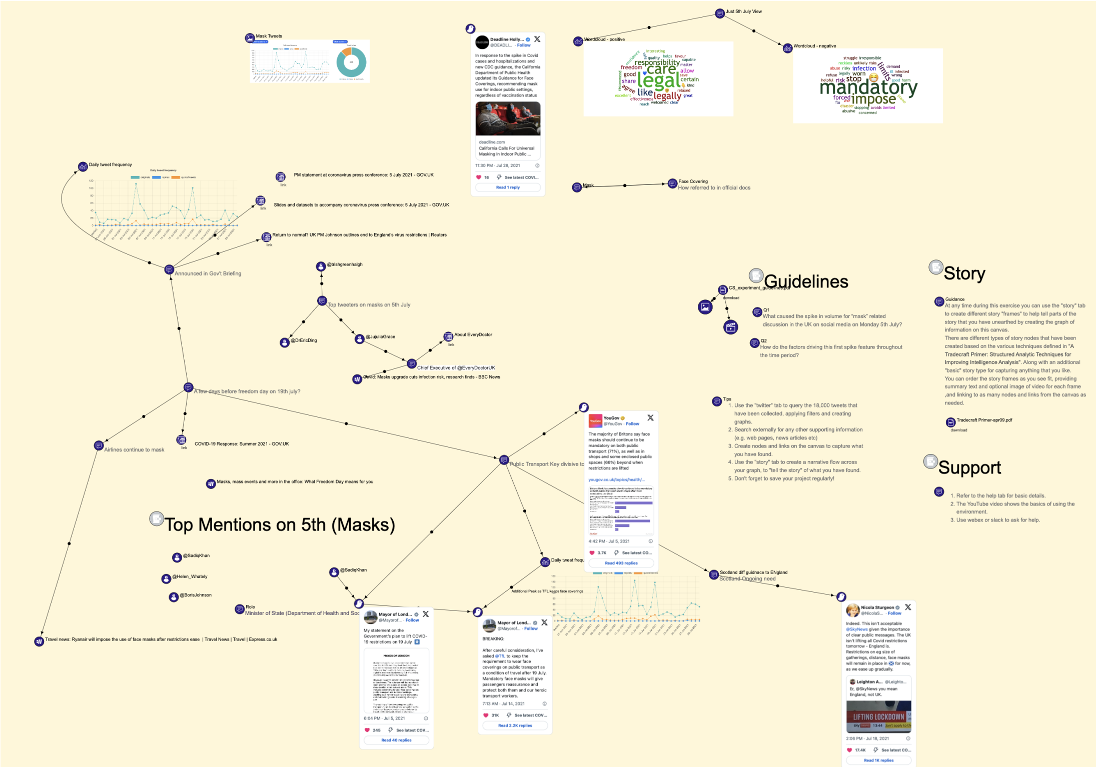


Figure B.12: Canvas for participant 06

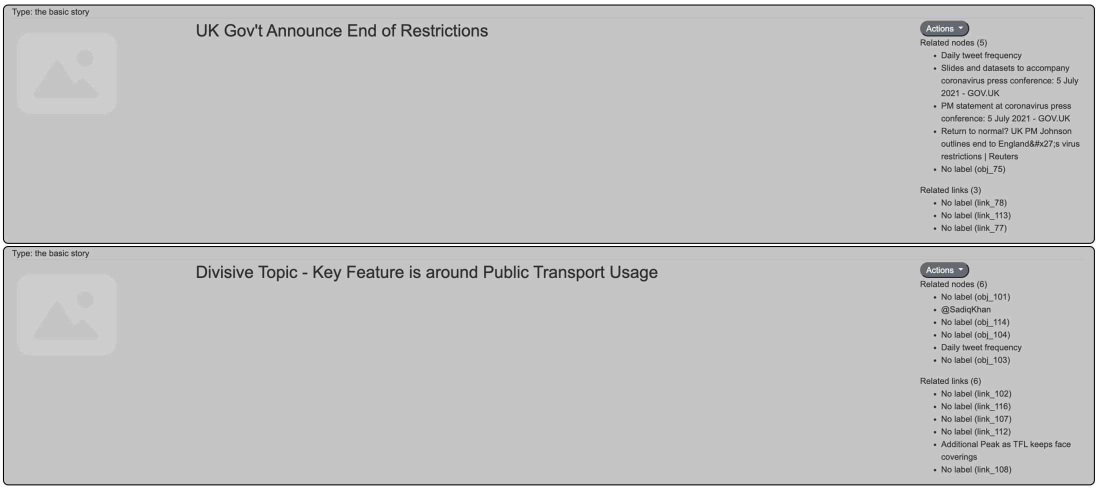


Figure B.13: Story for participant 06 (1 of 2)

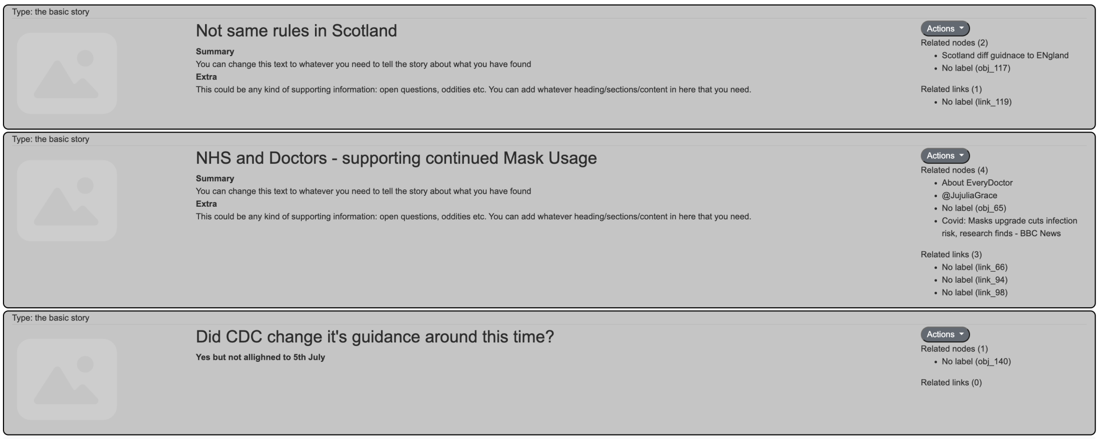


Figure B.14: Story for participant 06 (2 of 2)

Participant 06 has created their new content on the same canvas as the original guidance and has laid out one major graph to the left of the guidance, with a smaller related graph and some standalone nodes (Twitter usernames and hashtags). There are also fragments and a few single nodes relating to tweets and word clouds at the top of the page.

This participant created five story elements with nodes and links related to each of these. For the first two story elements there is no textual description provided, but instead the label and the related nodes and links are used to convey the meaning (announcement of end of restrictions, plus use of public transport). The same is the case for the third element, noting that different rules apply in Scotland, and the fourth story element which states that NHS doctors and nurses support continued mask usage. In the fifth and final story element they ask whether the CDC has changed guidance at this time? In the description they state that the guidance was changed, but this is not aligned to 5th July.

While this participant did not create many descriptions for their story elements they did create 11 separate text nodes with short descriptions on the main canvas, with each providing some commentary on the graph or a related node. They also created 5 web nodes, with URLs to external resources such as government slides, news articles and the details of organisations they had captured on the canvas. They also pasted in the text of the government press conference as well as recording the web page URL. They recorded 7 Twitter users on the canvas and linked some of these into the graph and recorded 5 tweets from official government and media sources.

B.3.7 Participant 07

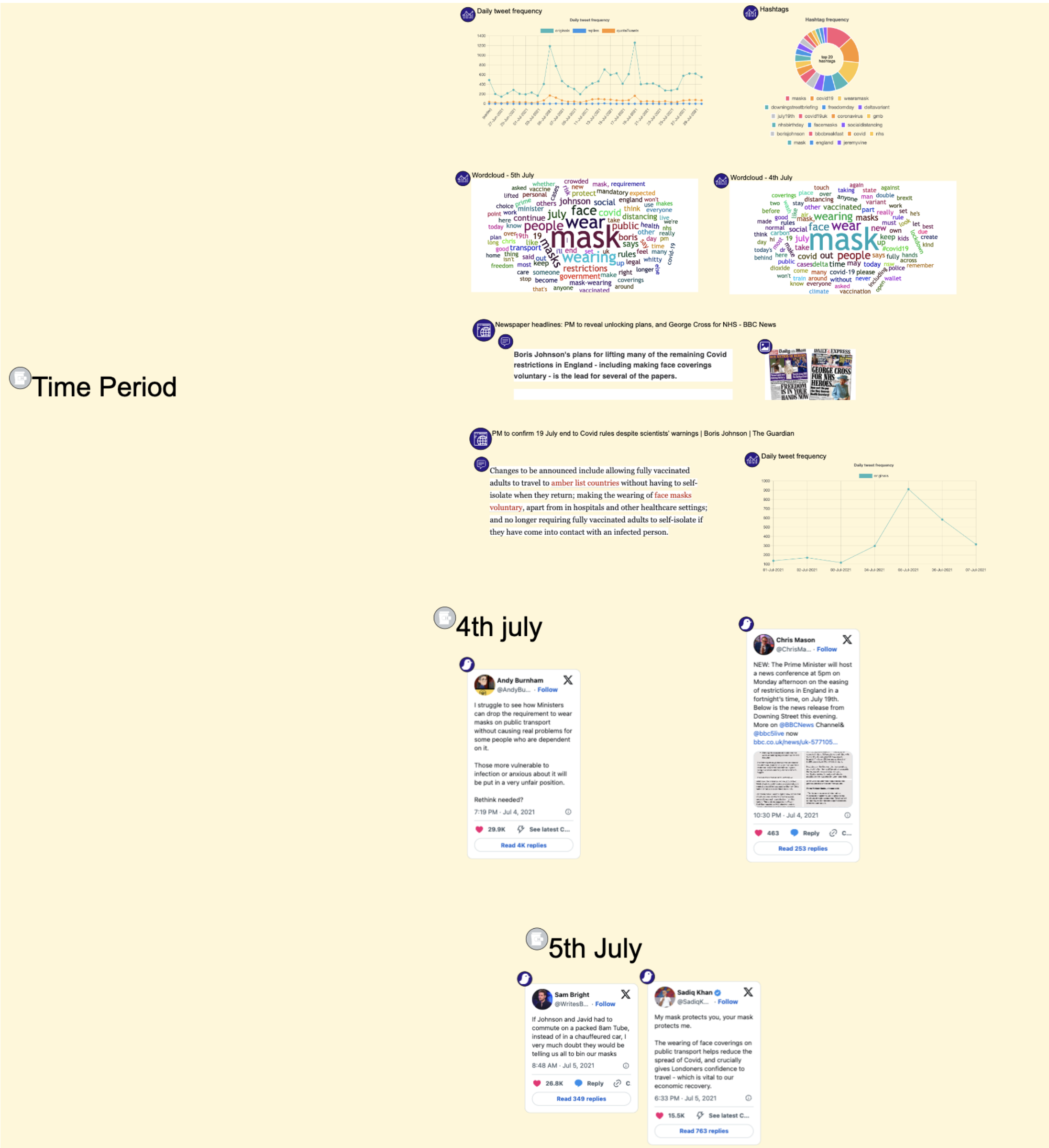


Figure B.15: Canvas for participant 07

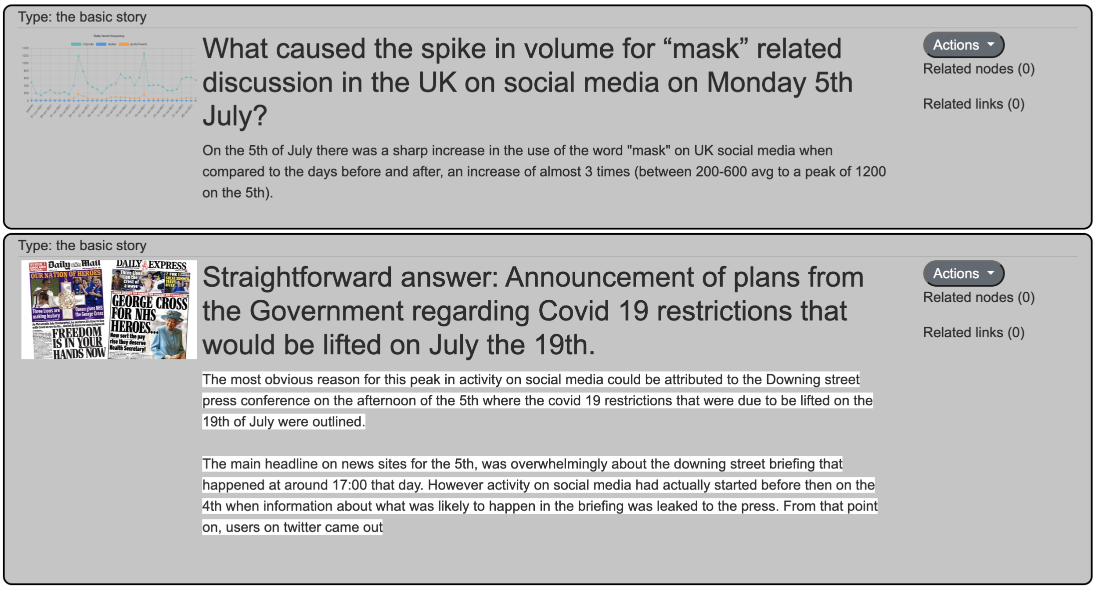


Figure B.16: Story for participant 07

Participant 07 did not use links to relate any nodes together on their canvas. They chose an overall layout for the whole graph based on a loose tabular (row and column) format and chose to keep the original guidance graph on the canvas rather than deleting it or creating a new project. They created 17 new nodes, of which 5 were dynamic charts or word clouds, 4 were tweets, plus a small number of text nodes, headers and one image node which is a copy of some newspaper headlines from the BBC website. They created two story elements, the second of which directly answers Q1, but there is no attempt to answer Q2. Whilst neither of the story nodes are related to any of the nodes on the canvas, the second story node uses the image with the newspaper headlines as the summary image. In answering Q1 they note that the spike was driven by the Downing Street press conference that day, but that there was already chatter beforehand at lower levels about the likely content of the announcement.

B.3.8 Participant 08

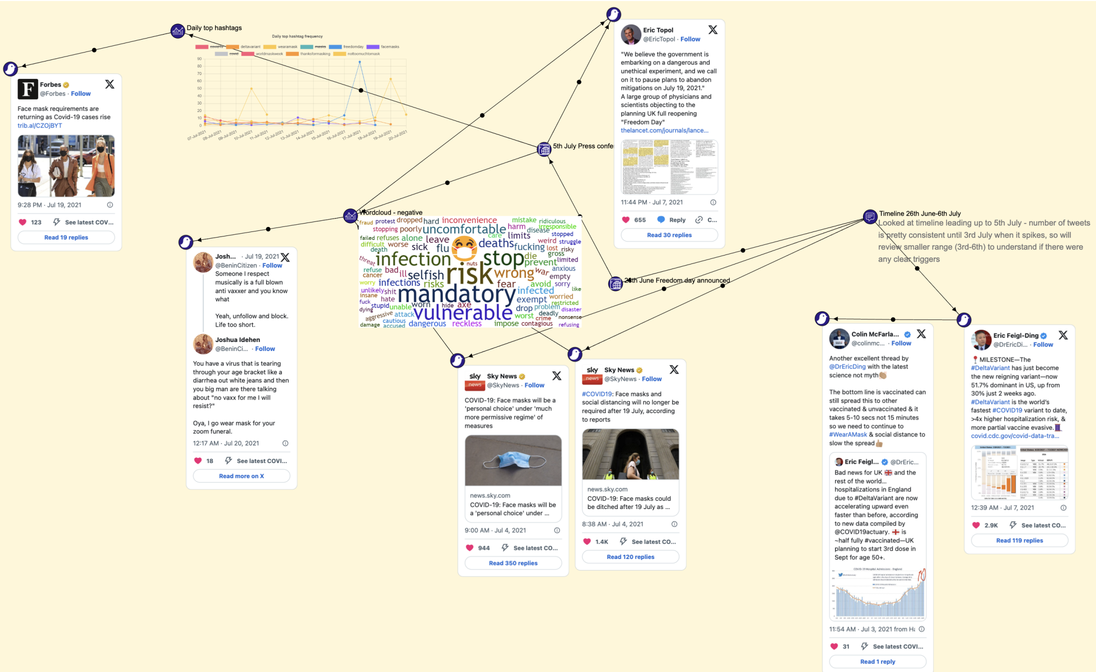


Figure B.17: Canvas for participant 08

Figure B.18: Story for participant 08 (1 of 2)

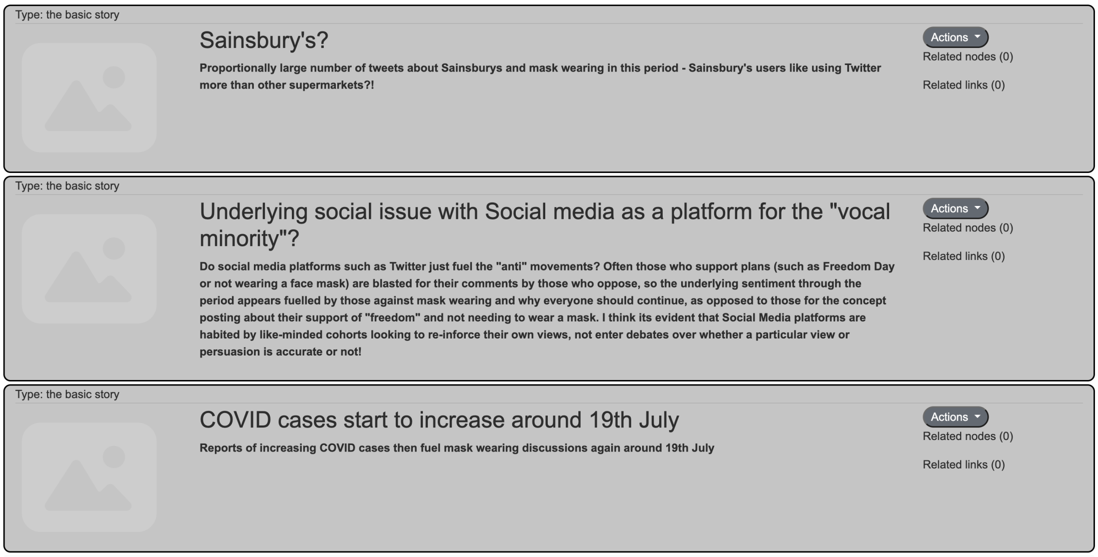


Figure B.19: Story for participant 08 (2 of 2)

Participant 08 created relatively little context (7 tweet nodes, 2 dynamic chart nodes, 1 web pages and 1 text node). However, they did create 6 story elements and provided substantial textual commentary on these, explaining their progress and their thoughts and outcomes. Also, they did annotate a number of the nodes that they created, by adding text properties to these standard nodes and providing some textual commentary in those. In the first story node they describe the timeline of government announcements relating to Freedom Day, starting on 28th June, alongside the rise of discussion about the Delta variant. In the second story node they directly assert that the spike was caused by reporting and discussion relating to the Downing Street press conference that day. This answers Q1, although they do not explicitly state this. The third story then looks at common underlying trends, finding several relevant hashtags. In the fourth story they specifically focus on one supermarket brand and question whether that is more prominent because Twitter users are more frequent shoppers there? In the fifth story node there is some opinion expressed about different communities on Twitter and the entrenchment of views rather than open discussion. The final story element is short and simply asserts that rising cases are driving increased discussion. This participant does not relate any nodes or links to any of their story elements, leaving instead the reader to infer these links. This participant does attempt to answer both questions but does not explicitly state this or relate back to the questions directly.

B.3.9 Participant 09

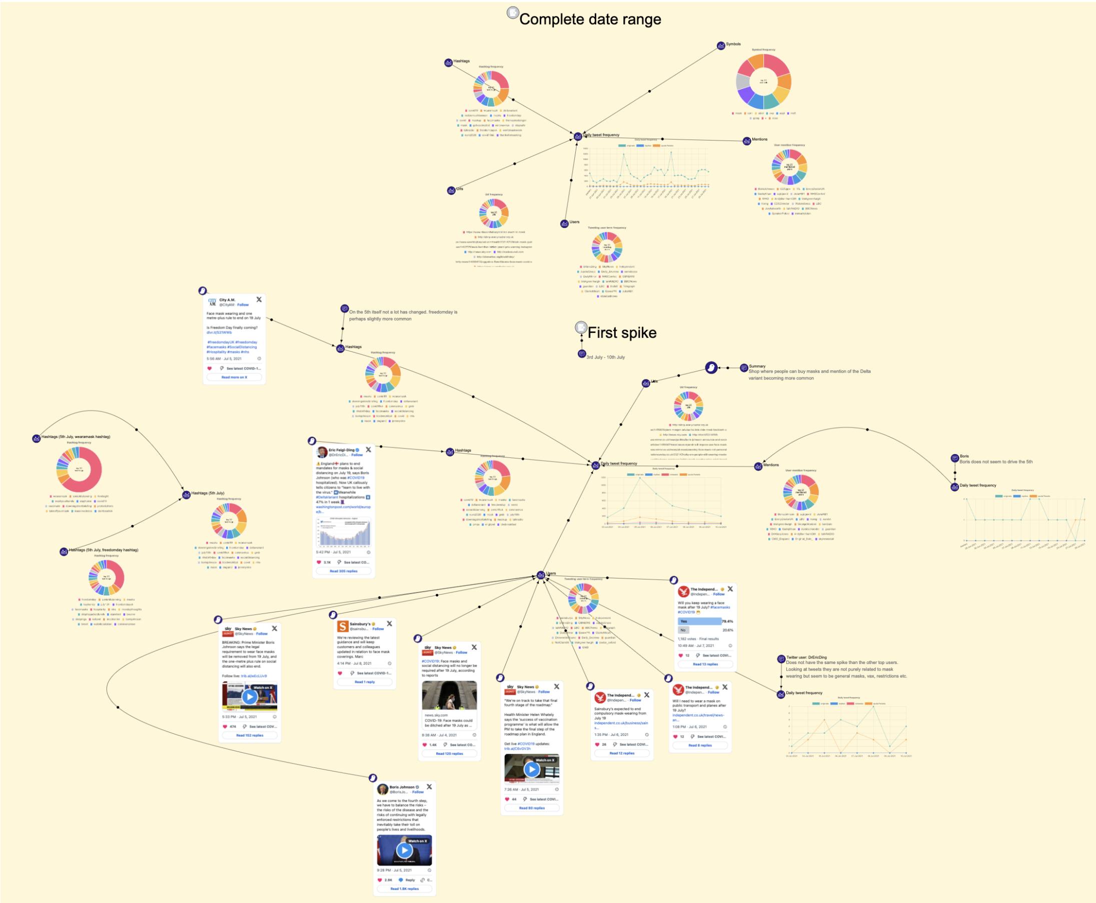


Figure B.20: Canvas for participant 09

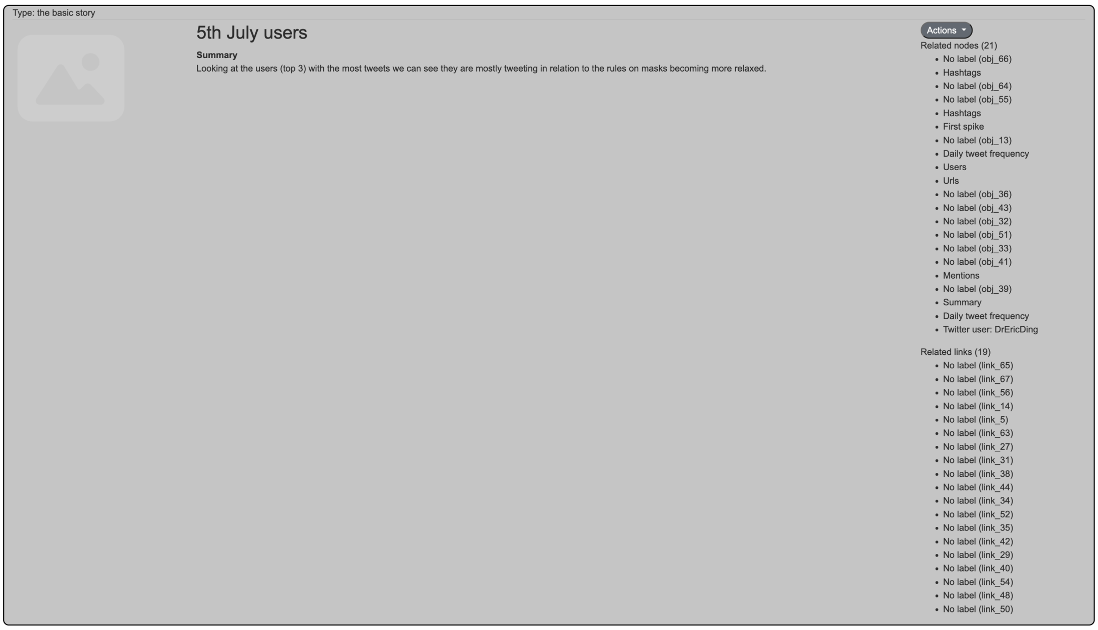


Figure B.21: Story for participant 09 (1 of 2)

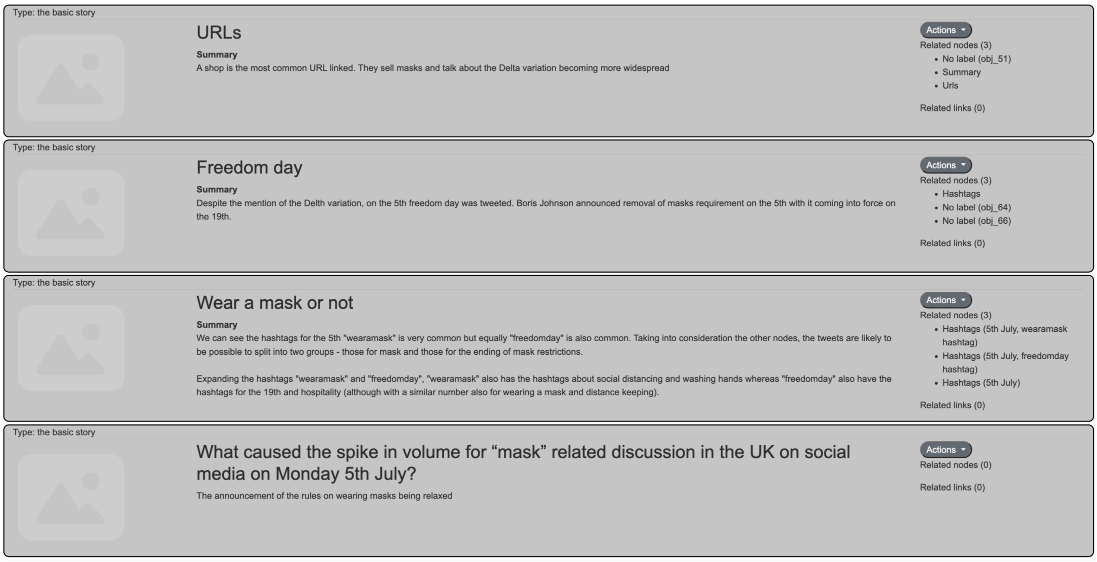


Figure B.22: Story for participant 09 (2 of 2)

Participant 09 made extensive use of dynamic pie charts (13 in total) and specific tweets (10 in total) on their canvas as well as 4 linked charts for tweet frequencies over time. They also created 5 story elements, 1 of which related to Q1 for this exercise. This participant deleted the original guidance nodes and links from the canvas rather than leaving them in place or creating a new project/canvas for their activity. They used links to relate different clusters within their graph and laid the canvas out to show one main cluster with a few satellite structures. Some links were bent to better accommodate some of the nodes for the more widely linked nodes. Little specific textual narrative is given, and the reader is left to infer the reasons why the various charts and tweets are linked or clustered. Some exceptions to this were observations relating to a specific Twitter user and that tweets relating to the selling of masks were becoming more common, along with mentions of the Delta variant and some short comments noting that some possibly expected aspects did not seem to be driving the spikes. Instead, they saved the bulk of their narrative text for the story elements. They have extensively linked their story elements to the relevant nodes and links on the canvas, and in one story element they offer an answer to Q1 in the exercise, specifically that the spike in volume of discussion was driven by the announcement of the relaxing of restrictions.

B.3.10 Participant 10

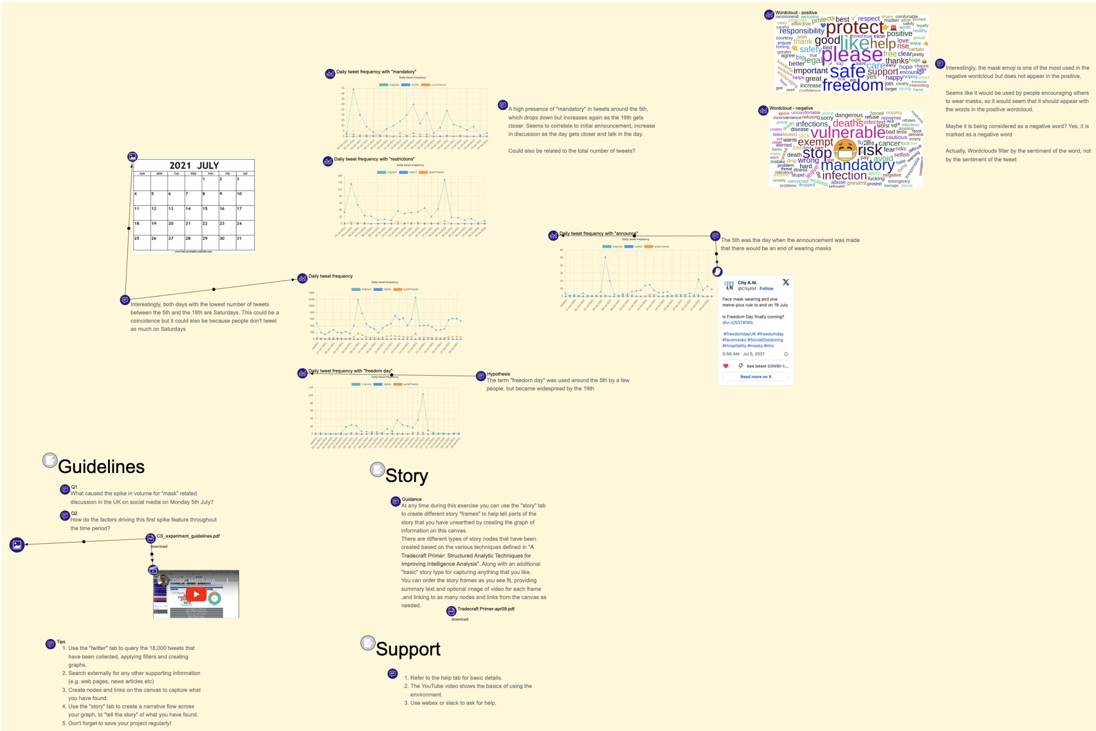


Figure B.23: Canvas for participant 10

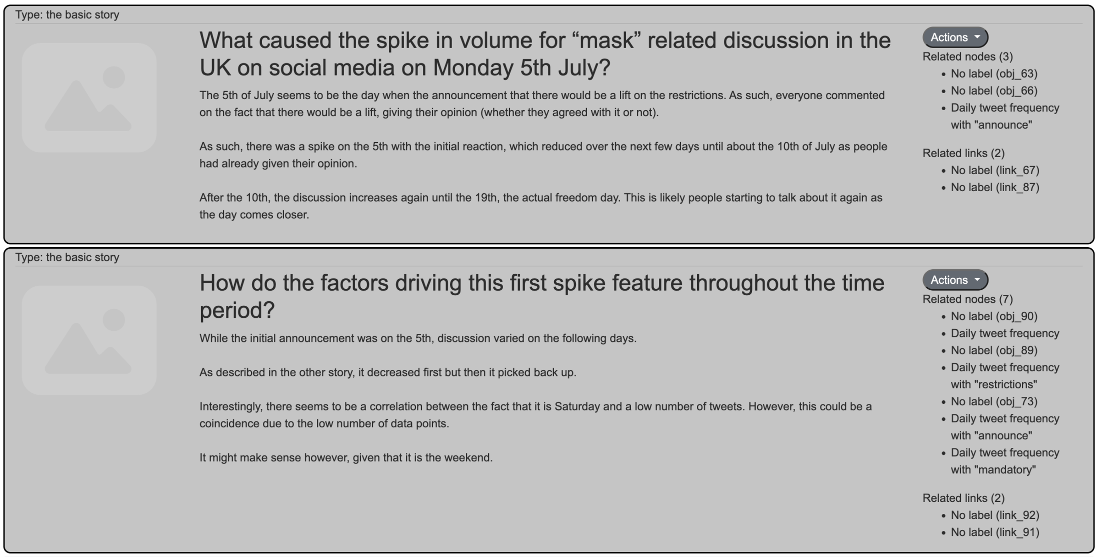


Figure B.24: Story for participant 10

Participant 10 created two story nodes and titled each with the question that was asked at the beginning of the exercise. Whilst they were not prolific in creating new nodes and links (14 and 5 respectively) they did use text nodes to track their thinking and to record a hypothesis. Their use of text nodes to provide a brief commentary of their progress throughout the exercise is useful and helps the reader of the graph to understand their motivation. This participant also notes that the mask emoji is only used in a negative context whilst many of the tweets that included the mask emoji could be considered positive[5]. However, the fact that through exploration of the data this participant noticed this issue and chose to comment on it is useful and an indication that they felt able to annotate their canvas with observations such as this.

In their answer to Q1 they correlated the initial spike to the government announcement on 5th July which drove discussions, then fading away until nearer the actual date of 19th July when the volumes began to increase again.

For Q2 they reiterated the observation from Q1, but also noted that the overall tweet frequency (for this collection) decreased on Saturdays. They offered no firm answer for Q2.

Some attention was paid to layout on the graph, in the sense that the charts and word clouds are positioned in a style that avoids overlapping but there is no obvious visual structure to the graph, and relatively little usage of links between nodes

[5]This is a known "issue" in that the sentiment analyser library considers the mask emoji to be negative, due to the fact that without a specific context the mask emoji is used to convey illness or isolation and therefore can be treated as negative. In the context of this experiment the pandemic-related setting and the question of restricting regulations means that some tweets use the mask emoji positively and others negatively.

B.3.11 Participant 11

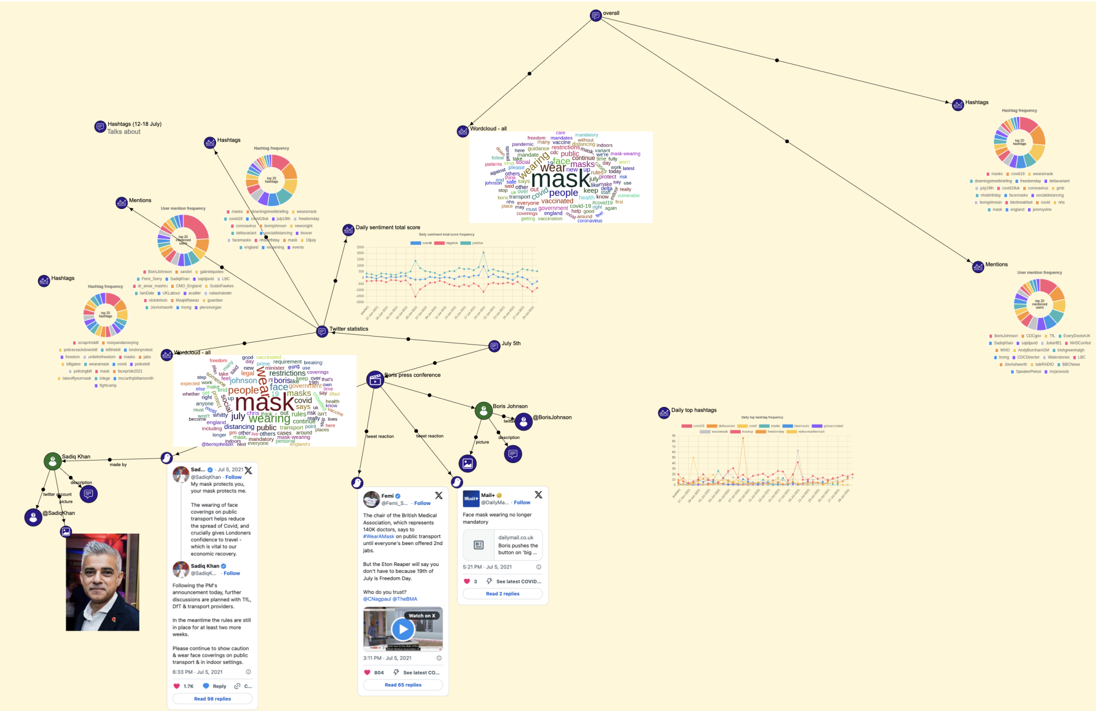


Figure B.25: Canvas for participant 11

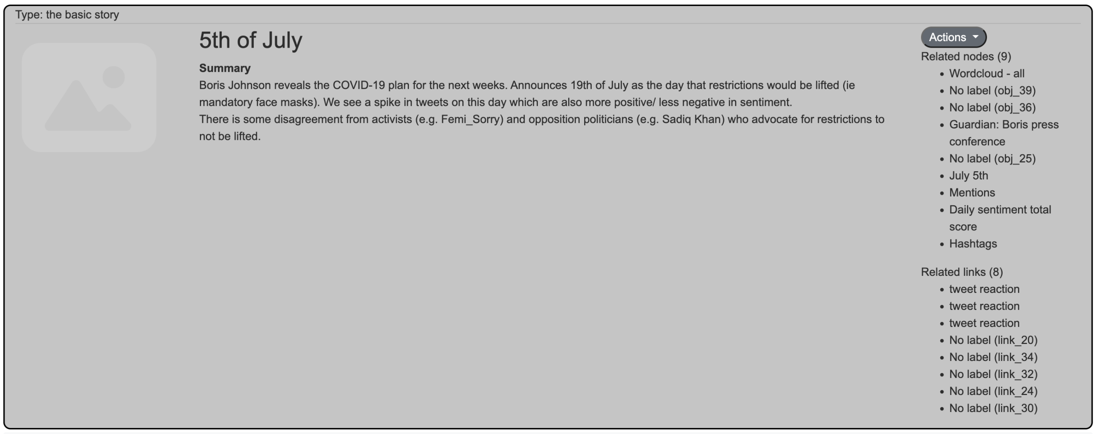


Figure B.26: Story for participant 11

Participant 11 focused on the creation of nodes and linking them together rather than providing textual commentary through text nodes or similar. A single story node is created with a short paragraph of textual summary noting that a press conference by the Prime Minister on 5th July drove discussion between different groups. The story node is linked to 9 nodes and 8 links on the canvas which are related to the textual summary provided.

There is no direct attempt to answer either of the two questions that were tasked as the goal of the exercise.

9 dynamic charts or word clouds were embedded and linked into the story and a video from the Guardian found externally and created on the canvas which is laid out as an overall single graph rather than separate clusters (two word clouds are slightly separate. This participant created a new project for the exercise, leaving the guidance on the original project

B.3.12 Participant 12

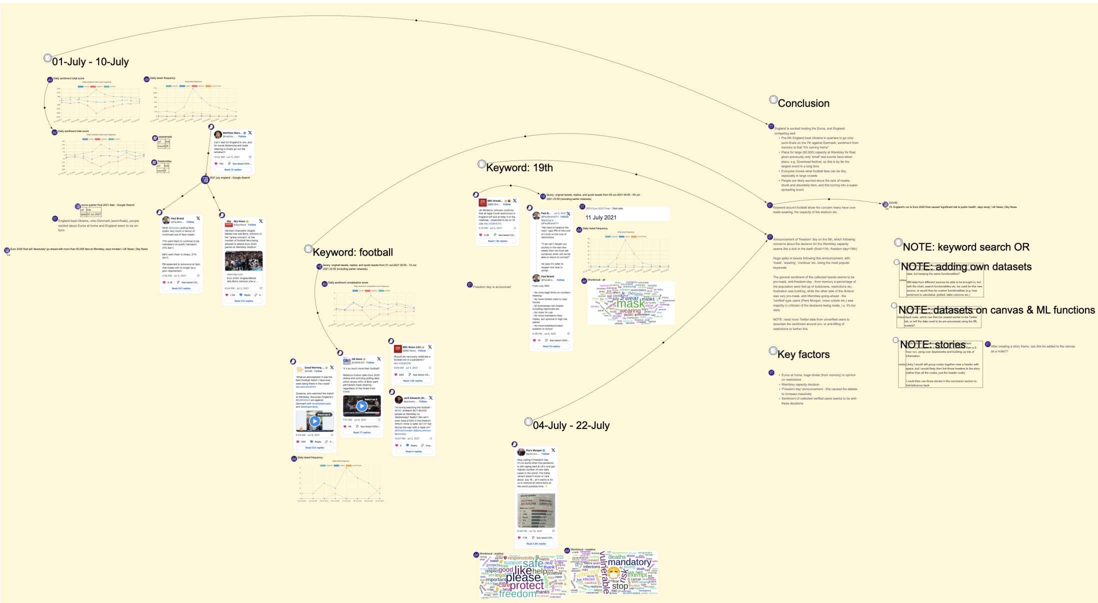


Figure B.27: Canvas for participant 12

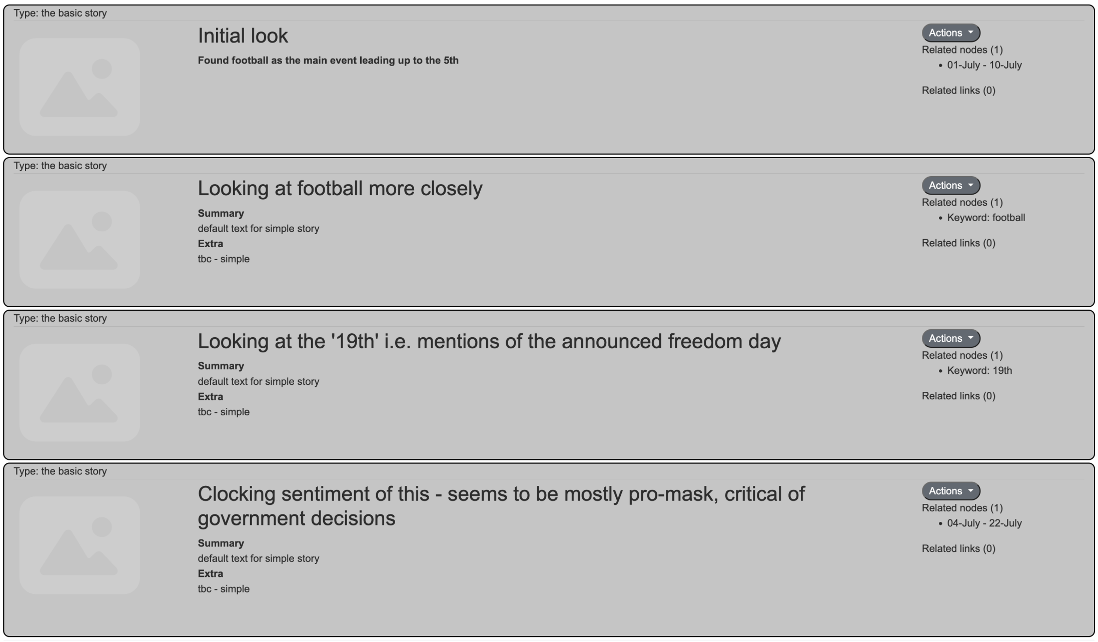


Figure B.28: Story for participant 12

Participant 12 chose to lay out their individual investigation areas as small clusters with a heading for each; usually some key individual tweets and one or more generated frequency charts or word clouds. In most cases they then linked these clusters with long links to summary or conclusion text that they wrote. They started by looking at key dates and extended this towards the end to also consider football as a particular contextual term (the only participant to do this) before moving back to investigate the 19th July as a specific date relevant to the exercise. They also created 4 separate nodes labelled "NOTE" which contained observations/feedback on the environment. This participant added properties to links within their graph, thereby annotating the link itself rather than just the nodes on the canvas. Also, they were the only participant observed to do this.

They carried out a google search (externally to the environment) but then created a node with the search details (as a URL) to enable them to link this to three other specific tweet nodes. Showing a link to the topic of football (European championships and the England team participation) for these tweets even when it is not immediately apparent from the raw tweet. Also, on the topic of football, contrasted tweets favouring large crowds and no masks against those raising concern about crowd size and distancing.

When summarising their conclusions (as text nodes on the canvas) this participant wrote a paragraph for their main clusters of research. Noting mainly that excitement about football was driving a desire for more normality to attend the event and participate contrasted against more official public health and public figure concern about the implications. They noted the tension created between the Euro final (11th July) vs Freedom Day (19th July) and the discussion this caused (citing specific tweets and showing graphs). They also suggested that access to more unverified Twitter user tweets (deliberately removed from the sample) would have helped them investigate if the unofficial view from everyday people was different to the verified Twitter user perspective.

About 30 minutes before the end of the exercise they commented about not using story nodes, but how they would do so with more time available. They then created 4 fairly simple story nodes and linked them to their main summary nodes on the canvas in the final part of the exercise. Some good items raised as nodes on the canvas as feedback/questions on the environment. This included some proposals for integrating machine learning agents into the environment and allowing them to write to the canvas

2 hashtags (#wearamask and #freedomday) were investigated and linked, 9 dynamic charts or word clouds were embedded and linked into the story along with 2 external google searches, and 2 links to Sky News articles were dropped onto the canvas. The layout seems carefully designed with tightly linked clusters of content related to specific investigations, with links to overall summary content. Curved links used when needed, layout appears visually meaningful.

Appendix C

Outputs from the DT workshop

The Design Thinking (DT) workshop is described in Section 3.2. This appendix contains additional supporting data in the form of: (1) summaries for the three personas (created after completion of the workshop by summarizing the post it notes created by each team during the relevant exercise), and (2) a summary of the top-rated ideas identified during the workshop.

C.1 Persona summaries

C.1.1 Corporal Palmer - 1st line maintainer, 26 years old

This was the persona created by Team A.

Persona summary: *As a relatively junior rank, Corporal Palmer is deeply focused on the successful delivery of their role as first-line equipment maintainer, ensuring the safe and successful provision of their equipment. They care about their reputation and skills and are always looking for opportunities to develop. They are in regular contact with their superior (second-line support) and may well supervise other technical support staff. Their job is complex and requires a good memory and a breadth of knowledge. The potential for AI assistance can be worrying as they fear their job could be replaced, as it is sometimes already with higher-paid civilian contractors. New technologies and configurations mean they must work hard to keep their skills current and their senior officers aware of operational concerns. They are motivated by technology and want to understand* *how everything works, often with a degree or similar in a technical subject. An AI assistant must be able to help them or offload some of their technical or bureaucratic workload to be seen as a benefit to them. They already use lots of technical test equipment and are familiar with the value of such items, and they spend time talking to others about the platforms and capabilities they support. They like their job and want to maximise their technical skills for when they retire from the service and seek employment outside. They feel big pressure to keep their systems up and running and can be uncomfortable with new technology that they're not familiar with. If things break then it's up to them to fix them, otherwise the force will be let down. They feel overworked, stressed by new technology now and in the future and may well be easily annoyed by poor tools, cumbersome processes or people who don't know what they are talking about. They are essential and unique but under pressure and fearful of too much change.*

C.1.2 Major Adam - Staff Officer, 33 years old

This was the persona created by Team B.

Persona summary: *Major Adam is degree educated and very motivated to deliver the very best intelligence. He's used to being in charge and drives his unit to maximum performance, drawing on extensive experience and his personal track record of constant delivery. Some of his peers refer to him as "Loud, confident and wrong". He is tasked with delivery of intelligence to his commanding officer, drawing on material gathered by his team. He knows that what he reports will have direct influence on decision-making and drives his team hard to cover as much ground as possible, checking sources, and testing credibility. He believes that he is excellent at this job and takes his responsibilities very seriously. He is unaware of his shortcomings and often forms opinions without having full evidential support. He feels overworked, with too many threads to follow, and that his team could do better, and his superior officers don't fully understand what he is briefing. He sees the value of technology to support his team but also fears dependency on it, and the* *potential cost and complexity that comes with it. His team are his biggest asset, and he must manage them strongly, not afraid to make examples if needed. He wants them to work as hard as he does, but perhaps he needs to work less hours. He's looking to the future and his next role and keen to keep his reputation as a hard-working tough delivery-focused officer and the strongest of his peers. He's always stressed, overworked, feels under-appreciated and underpaid and is on the edge of chaos but always delivering and he loves his job and wouldn't have it any other way.*

C.1.3 Commander Brian - Joint Operations, 1 star

This was the persona created by Team C.

Persona summary: *Commander Brian has seen a lot as he has risen through the ranks and finds himself working as joint operations commander in an overseas operation. His role requires close coordination with non-military groups such as Non-Government Organisation (NGO)s, the media and politicians. He is extremely confident and sees this operation as a job to be done well, and for those around him to help in that delivery. He needs top quality intelligence and other information from his team and wants them to expose any-and-all issues and concerns with their data, not hide away. He understands this command role and is keen to deliver quickly so that he can be recognised and move on to the next assignment. If technology can help him achieve his goals, he will happily add it to his team; he knows they are under-staffed and under-resourced and feels they need better technology to be able to win. Technology dependency is a big concern however and he worries about what pitfalls could arise from using it, and how adversaries may be able to exploit any kind of technology they rely on. He prides himself in his team and is very selective about who advises him and how they run their operations. He hates political interference and the higher chain of command and sees his role as providing solid evidence-based operations and intelligence. He uses a wide variety of techniques such as red-teaming and regular reviews and con-*

sultations and sits on numerous boards to ensure he is well connected. He feels personally accountable, reads a lot about techniques and approaches and drives his long working hours down into his team. He is usually over-stretched, making decisions with too little information and the relentless pace can be exhausting but this is what he signed up for and he feels like he's making a big difference but if only he could have more time and resources and be more confident.

C.2 Highest prioritised ideas

The table below lists the highest rated ideas during the prioritisation grid exercise for the previously generated big ideas.

Team	ID	Idea	Topic
A	2	AI 'help' to ease cognitive load. 'app' 'chatbot'	H1
A	7	AI feeds Cpl Palmer with lessons learnt	A2
A	8	Historical precedence AI	A2
A	14	AI help 'app'/'chatbot' to aid translation concept to action	A2
A	24	Easy remote fixing, over the network diagnosis + repair	X
A	26	Psychologist 'app' to provide reassurance / calming when your body needs it – linked to watch / phone?	H1
A	42	Why is this location intelligence important? Explanation + £s + time	H3
B	2	A metric for info provenance	H3
B	3	Data confidence tool – do I have all that is available?	H3

B	13	Train alternate thinking / contrarian to question group think	A1
B	15	Reliability ratings for data. Open source vs restricted	H3
B	20	Taxonomy of confidence – common + understood	H3
B	23	Overall AI confidence calculator	A2
B	24	Education to understand AI functions / processes = increased confidence(?)	H3
B	25	Competing AI algorithms to deliver variety / confidence	A2
B	28	A fall-back method if the AI fails	H2
B	29	Library for lesson ID from previous problems – are we repeating errors?	A1
B	43	A decision making tool for the human	H1
C	5	AI knows info at higher classification than commander	A2
C	8	Commanders virtual personal agent (bespoke)	I
C	10	AI run rapid wargaming	A2
C	12	Big data	X
C	22	Record / document gut feel	H2
C	26	Personal red team (based on Brian's record)	T
C	27	Digital conscience	A2

Table C.1: The highest rated ideas from the prioritisation exercise