Research Space
Journal article

Developing conversational agents for use in criminal investigations

Hepenstal, S., Zhang, L., Kodagoda N. and Wong B.L.W

This is the authors' accepted version of:

Sam Hepenstal, Leishi Zhang, Neesha Kodagoda, and B. l. william Wong. 2021. Developing Conversational Agents for Use in Criminal Investigations. *ACM Trans. Interact. Intell. Syst.* 11, 3–4, Article 25 (December 2021), 35 pages. DOI:https://doi.org/10.1145/3444369

Developing conversational agents for use in criminal investigations

SAM HEPENSTAL, Defence Science Technology Laboratory, UK
LEISHI ZHANG, Middlesex University London, UK
NEESHA KODAGODA, Middlesex University London, UK
B.L. WILLIAM WONG, Middlesex University London, UK

The adoption of artificial intelligence (AI) systems in environments that involve high risk and high consequence decision making is severely hampered by critical design issues. These issues include system transparency and brittleness, where transparency relates to (i) the explainability of results and (ii) the ability of a user to inspect and verify system goals and constraints, and brittleness (iii) the ability of a system to adapt to new user demands. Transparency is a particular concern for criminal intelligence analysis, where there are significant ethical and trust issues that arise when algorithmic and system processes are not adequately understood by a user. This prevents adoption of potentially useful technologies in policing environments. In this paper, we present a novel approach to designing a conversational agent (CA) AI system for intelligence analysis that tackles these issues. We discuss the results and implications of three different studies; a Cognitive Task Analysis to understand analyst thinking when retrieving information in an investigation, Emergent Themes Analysis to understand the explanation needs of different system components, and an interactive experiment with a prototype conversational agent. Our prototype conversational agent, named Pan, demonstrates transparency provision and mitigates brittleness by evolving new CA intentions. We encode interactions with the CA with human factors principles for situation recognition and use interactive visual analytics to support analyst reasoning. Our approach enables complex AI systems, such as Pan, to be used in sensitive environments and our research has broader application than the use case discussed.

CCS Concepts: • **Computer systems organization → Human-centered computing**.

Additional Key Words and Phrases: explainability, criminal intelligence analysis, conversational agents, transparency

ACM Reference Format:
Sam Hepenstal, Leishi Zhang, Neesha Kodagoda, and B.L. William Wong. 2020. Developing conversational agents for use in criminal investigations. 1, 1, Article 1 (January 2020), 35 pages.

1 INTRODUCTION

1.1 The Challenge

Artificial intelligence (AI) based conversational agent (CA) technologies are still in their infancy. However, their popularity is increasing [22, 23], because they provide more intuitive, natural, and faster access to information. A CA can significantly speed up repetitive information retrieval tasks, compared to traditional query tools or manual processes. This technology is therefore attractive for a wide range of domains that need fast reasoning and decision making informed by data analysis, for example, in a criminal investigation. The volume of data that requires processing by police today

Authors' addresses: Sam Hepenstal, Defence Science Technology Laboratory, UK; Leishi Zhang, Middlesex University London, UK; Neesha Kodagoda, Middlesex University London, UK; B.L. William Wong, Middlesex University London, UK.

© 2020 Crown copyright, Dstl. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gov.uk
XXXX-XXXX/2020/1-ART1 $15.00
https://doi.org/10.1145/3444369

is a significant barrier to bringing criminals to justice. This is a key area in which an intelligent CA that can interpret analyst questions and retrieve the appropriate information, including providing machine reasoning, can have significant impact.

There are challenges, however, with reference to high risk and high consequence decision-making environments. Algorithmic transparency is lacking in commercial conversational agents, where due to "black box" approaches it is difficult to understand how they have understood a user, explored the data, and built a response. Typical applications for conversational agents tackle concise user tasks for mundane processes that can be translated to a finite set of user intentions. The use of the CA disguises the underlying processes making interactions more natural, for example when booking a meeting. Thus, traditional CAs have not been built with algorithmic transparency in mind. The tasks these CAs are able to perform are constrained to a predefined set of 'intentions', which can lead to frustration when a user's misguided mental model leads to the expectation that the CA can behave like a human. There are dangerous implications in high risk and high consequence environments if a user misinterprets or underestimates the constraints of a system. A CA typically comprises various complex processes, including natural language processing of a user's question and intention, identification of an appropriate action that triggers additional data processing, and formulation of responses. Matching this collection of underlying processes and their relative impacts on an investigation, to a user's mental model for the system, is difficult. Different methods can be applied to achieve the same aim but, depending upon the specifics involved and their interpretation, an investigation can be guided towards different paths with consequences. The method chosen and impact is not clear within a concise natural language response. Furthermore, the CA can learn and evolve new capabilities, and become more accurate on which processes to call over time through interactions with analysts. This means that a question on one day may not trigger the same set of processes if asked again on another day. This could be problematic for allowing analysts to form expectations and trust in a system. For an analyst to trust the information retrieved by a system in an investigation, they need certainty in their understanding of the search methods and any caveats. The analyst may be required to go to court to explain their analysis, methods and conclusions, and they therefore must be able to be held accountable. There are also ethical implications of using artificial intelligence applications that are a key consideration for designing CAs. Without trust and ethical justification, underpinned by system transparency, AI systems will not be used by analysts.

1.2 Our Approach

We look to bridge the gap between theoretical transparency requirements and the design of operational visual analytics systems. Visual analytics is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [49]. We consider how to design an interactive conversational agent system to support analytic reasoning in a criminal investigation and build a working prototype, therefore contributing to the field of visual analytics. We focus upon the perspective of system transparency, where transparency is a common issue shared by other AI applications. Our approach, therefore, is not limited to intelligent agent applications.

The work described in this paper is closely informed by three separate user studies with expert users in the domain of criminal intelligence and investigations. These studies provided us with the necessary understanding to (i) model analyst questions and intentions in an investigation, (ii) develop explanation components for a conversational agent (CA) prototype, and (iii) evaluate our approach.

1.3 Our Contributions

The main contributions described in this work are as follows:

(1) A novel prototype system that:
   (a) Models, structures, and presents system behaviour in a way that mirrors human recognition. By understanding how analysts think when they perform investigations, we build an architecture for deriving conversational agent intentions that reflects human thought processes.
   (b) Can evolve to incorporate new analyst intentions. Analyst questions adapt and flow as they perform an investigation, thus there is a requirement for fluidity in system behaviours and capabilities.
   (c) Provides a platform from which to develop autonomous investigative agents, capable of both challenging investigation scope by following their own lines of inquiry and providing algorithmic opacity by explaining their reasoning at different levels of summarisation.
(2) An initial evaluation of our approach, both to identify the needs for system transparency and to design a conversational agent system that meets them.

2 RELATED LITERATURE

Analysts perform criminal intelligence analysis to derive situational awareness and they can benefit significantly from Artificial Intelligence (AI) support. Analysts play an important role in an investigation, as the results of their analysis underpins decision making by police commanders. For example, intelligence analysis directs the prioritisation of lines of inquiry and assessments of key suspects. The process of intelligence analysis involves repetitive and intellectually non-trivial information retrieval tasks where "each piece of insight leads to intense periods of manual information gathering" [16]. Criminal investigations are guided by the believed 'scope' of the situation [16], where the questions asked by analysts are framed by their expectations given the scope. Analysts cannot realistically question every possibility and therefore, in a time-pressured situation, they must focus their attention and information gathering. Throughout an investigation, their questions, and subsequent interpretations of the results, allows them to build upon or to challenge the scope. Analyst arguments are not clearly structured at the outset, and a progressive series of questions along an investigation path guides the direction of inquiries. The path is susceptible to change depending upon the results of each question. This questioning is an important process, where at each stage insights can emerge in different ways and may not be directly related to the question the analyst originally asked. For example, a question may trigger an unexpected response, or intuition, that gives a fresh perspective. Investigations do not simply involve the selection of evidence to accurately prove or disprove a hypothesis that has already been formed. In most cases there are large gaps in an analyst's understanding of a situation at the outset of an investigation. This hampers the ability of an analyst to form detailed hypotheses. There is therefore a need for iterative development of insight, as information becomes available. Klein [25] captures this picture nicely where he explains that performance improvements depend on both reducing errors and increasing insights. We propose that an intelligent system, therefore, that simply reduces error by improving the speed and accuracy of information retrieval or hypothesis testing, but does not account for insight, can actually hamper overall performance.

Wong and Kodagoda (2016) [55] present the process of repetitive questioning in an inference decomposition chart, showing pathways between different activity stages and related inferences, premises, and claims. The activities at each stage include looking for patterns, strengthening interpretation of modus operandi, questioning specific details, and predicting paths in the data. The activities are manually driven and are prone to expectation bias, given that analysts need to narrow their scope in order to ask useful questions. We identify this as an opportunity for intelligent agent systems that can provide additional reasoning capability, for example by exploring alternative lines of inquiry that fall outside the current investigation scope. The semantic nature of a semantic knowledge graph itself provides some opportunity to share reasoning with human analysts. For example, if an analyst asks a question about an individual and whether they were in an area at a particular time. An intelligent agent can reason that the individual has the class of 'person', so if the specific individual in mind was not present in the area at the right time, it may be able to suggest other people who were. The reasoning abilities of an intelligent agent could be enhanced through an understanding of investigation methods and states.

If an intelligent system can improve the investigation process, the impact could be significant. Criminal investigations involve high risk situations that require rapid decision making, such as in kidnapping events. At the beginning of an investigation there is typically little information to help make decisions, and a key requirement for an analyst is to develop understanding of the situation by finding and collating important information. For this they interact with available data and databases. Writing query syntax manually, or via traditional analysis tools, can be cumbersome and time consuming. SPARQL [40], for example, is a complex query language which requires careful thought to configure. Additionally, the volume of data that requires filtering and processing by police is significant. In June 2019 Cressida Dick, the Commissioner of the Metropolitan Police, explained that "sifting through vast amounts of phone and computer data is partly to blame (for low solved crime rates) as it slows down investigations" [47]. A more natural interaction, which removes the requirement for analysts to translate their questions into restrictive syntax or structures, could speed up this process significantly. If an analyst were able to communicate with their data in the same way as they do with their colleagues, through natural language, then they could achieve significant time savings and speed up investigations.

In a policing scenario, when an analyst is presented with a situation they immediately look to make sense of it. They apply experience to recognise aspects of the situation and construct a plausible narrative explanation with supporting evidence. Klein [26] presents the Recognition-Primed Decision (RPD) model to characterise how humans recognise and respond to situations, including their cues, expectancies, actions and goals. Situation recognition is what we desire of a CA when it attempts to understand, act and respond to analyst questions. However, it must also be able to explain how and why it has achieved its conclusions, as you would expect of a human analyst. This need for explanation relates to system trust and ethics in policing. We define trust as a measure of the strength in the expectancy that interactions with another entity will result in positive outcomes within an uncertain and risky environment [6]. Trust is not a binary nor absolute value. It varies based on the situation, personal factors (e.g., propensity to trust and disposition), historical interactions, and moderating factors such as workload [44]. There are also implications for the success of the investigation. As discussed earlier, insights and inferences are made at each stage of questioning. Without adequate visibility of the methods applied and underlying data, an investigation may be hampered.

2.1 Algorithmic Transparency of Conversational Agents

Commercial CA technologies have not so far accounted for transparency, instead focussing on designing interactions that are intuitive and natural. We use the definition for system transparency as given by Hepenstal et al. that transparency is the ease with which a user can (i) explain any results provided by a system, in addition to (ii) being able inspect and verify the goals and constraints of the system within context [15]. Popular technologies such as Google Home, Siri and Amazon Alexa present us with an easy way to access music, films, or plan our day. Many services have incorporated chatbots into existing processes to manage interactions with customers, including directing them to the right information or department. This saves companies money and can save customers time waiting in a queue. The risks of an incorrect or misleading response are low and the resulting consequences limited, particularly given the ease with which a user can validate results against an expected and desired conclusion to their interaction. As a result, traditional CAs have not been built with algorithmic transparency in mind. If you ask Google Assistant, for example, why it has provided a particular response it will not be able to tell you and instead responds with humour, such as "Let's let mysteries remain mysteries." This is not adequate for use in criminal investigations.

There has been limited research that explores the need for a CA to explain the responses it makes. Preece et al. describe the ability to ask a CA 'why' they have provided a particular response, so an analyst can obtain the agent's rationale. An explanation could be "a summary of some reasoning or provenance for facts" [39]. This understanding of explanation is consistent with research into explainable machine-learning, where the focus is placed upon the specifics of the data retrieved, or the internals of a model. Intelligence analysis is a field where analysts operate in complex, subjective, uncertain and ambiguous environments, and analysts find it difficult to trust opaque computer systems. In order to make high risk and high consequence decisions, informed by visual analytics systems, a simple explanation of the data or a model which defines a response is not enough to satisfy the needs for understanding. Our research has developed a design framework for algorithmic transparency [15]. This describes the necessity to go beyond XAI when designing intelligent systems, to include visibility of the system goals and constraints within context. Context relates to the usage and user, including a user's mental model for the ways in which the CA system works.

Users who have a different mental model to the realities of the system can encounter difficulties and are prone to error [37]. We propose that users can be confused by CA interfaces where, because the interface can understand and respond with natural language, users have higher expectations than the actual capability and do not understand how to interact with the CA. A good example is Anna, a chat bot used by IKEA between 2005 and 2015, where users were frustrated by the bot not being able to interpret them accurately enough. The human nature of interactions was recognised as a problem, "trying to make Anna 'human-like' meant that people were more likely to ask it stupid questions" [51]. In reality the chatbot can only respond if it recognises specific programmed intentions. Questions such as, "I don't want to buy it, I want you to pick it up", can trigger the contextually incorrect response: "Unfortunately IKEA does not currently offer the service of store pick-up for online orders" [33]. Issues like this occur due to both an inaccurate interpretation of the human on behalf of the chatbot, but also to a fundamental misunderstanding of the possible intentions and capabilities of the chatbot on behalf of the human. The user utterance is difficult for the chatbot to interpret, due to the ambiguity that the user wanted an item picked up from their home and not from the store, which would be typical. If an intention has not been considered by the chatbot developer then it cannot be addressed by the chatbot, this is fundamentally different to an interaction with a human. User trust in a system will degrade if it cannot quickly and accurately address their question without reason. As a minimum, the user should be able to understand why their question has not been accurately answered and what methods have been triggered. For Anna, a degradation of trust may cause a customer to find another route to contact IKEA, however in high risk and high consequence environments there could be more damaging impact from a lack of transparency. In previous work, we have demonstrated mental model and transparency issues in the context of a CA for intelligence analysis [15]. For example, when performing intelligence analysis a human analyst may ask a CA the question, "is there a link between Person A and Person B?" The CA can interpret this and give the response "No link found." However, there are subjectivities where the human analyst may have an incorrect mental model for how the system has worked out the answer. For example, what does 'link' mean, and is the analyst considering any links across multiple nodes, or only direct links? The analyst needs to know that their mental model and intended goal matches the underlying system processes and system goal. They also need to identify where there are constraints that prevent the system from achieving their goal.

To achieve effective teaming between an analyst and a CA, the analyst must be able to trust the information retrieved by the CA and the processing involved. A key aspect enabling trust engineering in systems and addressing ethical concerns is transparency, so that the analyst can predict, interpret and refute any results, acknowledging caveats where they exist [11]. While trust is a crucial element in enabling analysts to team with AI systems, it needs to be handled carefully. In intelligence analysis, where analysts should apply critical thinking and scepticism before accepting hypotheses, the impact of trust is a complex issue and it is damaging if it leads to analytical complacency. If an analyst trusts the results of an algorithm without interpreting the reasoning and understanding any constraints, then there could be harmful consequences. Too much trust without verification may lead to poor decisions, unrecognised bias, or allow algorithmic deception. Analysts should therefore rightly be wary of trusting algorithmic systems which they cannot understand nor inspect and verify i.e those which are not transparent.

There are also significant ethical requirements for AI systems to be used in policing. For example, past work undertaken as part of the Visual Analytics for Sense-making in Criminal Intelligence Analysis (VALCRI) project, developed a solution for policing which includes AI algorithms and uncovered important ethical issues. These are "accidental discrimination, the Mosaic effect, algorithmic opacity, data aggregation with mixed levels of reliability, data and reasoning provenance, and various biases" [31]. We focus on the issue of algorithmic opacity, or transparency, and propose that this can help address other ethical concerns, for example to allow inspection of the provenance of machine reasoning when responding to an analyst's question. In Leslie's [30] guide for the responsible design and implementation of AI systems in the public sector, transparency is a requirement in terms of both the product and the processes underpinning AI systems.

2.2 Brittleness of Conversational Agents

Intelligent agents come in a variety of forms and include those that react to stimulus based upon rules, or a model, to more advanced agents that can work towards a goal, and even assess how a goal can best be attained. In the case of a CA for retrieving information to support criminal intelligence analysis, we are firstly interested in a reflex agent that can interpret natural language utterances from a user and respond based upon some rules, such as by applying appropriate methods to search for information and then describing results back to the user. This type of CA is referred to as a "spoken dialogue system" [34] which uses spoken dialogue to interact with users and accomplish a task.

As we touched upon previously, typical CAs understand users by matching their input pattern to a particular task category (intention), for example using 'Artificial Intelligence Markup Language' (AIML) [42]. In terms of the Law of Requisite Variety [5] brittleness occurs when the technology fails to cope with the variety of demands that it has to cope with when in use. In many cases, categories require pre-configuration and the system is therefore brittle. Two examples are Language Understanding (LUIS) [2] and the DeepPavlov framework [1].

LUIS is used to develop chatbots as one of the Cognitive Services for Microsoft Azure. LUIS matches 'utterances', which are statements made by a user, to 'intentions'. These can then trigger a set of rules depending upon which intention is matched. For example, if a user makes the utterance "book me a flight to London", LUIS will match their utterance to the intention for booking a flight, thus triggering the various rules that allow them to do so. Other examples of typical intents used by LUIS include specific tasks such as "Make Booking", or "Get Agenda" [3], where each requires bespoke functions and thus, the approach is brittle. For the CA to respond to a new task requires the programming of a new set of instructions. There is no obvious mechanism to provide consistent transparency for the inner functions or constraints of a matched intention.

DeepPavlov provides a number of options for developing chatbots, from goal-oriented approaches to open-domain question answering and pattern matching. Without the necessary text for training a model, open-domain question answering to select the best answer, or sequence-to-sequence approaches to predict the next sequence of words from text, are not appropriate. Additionally, we may wish to trigger a more complex set of processes, than merely return text. Thus, in the domain of criminal investigations where there is not a large amount of training data, we require what DeepPavlov defines as a pattern matching skill, which matches the user's utterance to a pattern and returns an answer from a predefined set. This is similar to the LUIS approach. If an intention pattern and response has not been considered by the chatbot developer when developing the system, the chatbot cannot evolve new capabilities through interactions with the user.

We require a CA which can translate from natural language to retrieve data from a semantic knowledge graph. This involves the creation of SPARQL [40] query syntax and processing of data returned. A straightforward method to allow analysts to interact via natural language with a graph database could be to directly translate natural language into SPARQL [57]. While this allows analysts flexibility and avoids intention brittleness, we may have more complex desires from a CA, for example additional processing or reasoning with information returned. Our intentions do not necessarily translate to a single SPARQL query statement and may comprise multiple processing steps. There is also no clear means to provide transparency of the SPARQL translation. i.e. how the subjectivities of natural language are interpreted in SPARQL syntax.

Juji provides a novel model-based approach to building a CA to conduct interviews [58] that consist of multiple interactions. Conversations are modelled on GOMS (Goals - Operators - Methods - Selection rules) [8] to reflect the cognitive structure of the 'conversation space' and the threading of topics together to achieve conversation goals. The model-based approach allows for greater flexibility than traditional CAs, where various ingredients are combined to define interactions and a user can flow between a selection of topics via 'conversational acts'. Juji [58] has developed a large topic library, to allow discussion of specific subjects ranging from hobbies to challenges and achievements. Brittleness is mitigated by the flexibility with which topics can be designed for specific purposes and selected and discussed, using topic-agnostic models for different stages of dialogue. This approach is sufficient for interviews given the bounded nature of conversation topics and stages of dialogue. In intelligence analysis, however, the conversation space is unbounded, as are potential requirements and information retrieval methods. Juji also does not consider CA transparency, given that this is not necessary for interviewees.

Intelligence analysis is not a fixed process and analysts need flexibility. Gerber et al. describe the fluidity of analysis, where an analyst moves from intuition, following 'leap of faith' assertions, to identify new insights [13]. These different processes present a variety of goals for information retrieval, from purely explorative to testing specific hypotheses. The methods of information retrieval also differ depending upon the subjective interpretation of the analyst's intention, as does the response required from the CA. Wong and Kodagoda (2016) explain how analysts evolve their initial understanding of a situation, formed through an "iterative combination of abductive, inductive and deductive inferences, information searching, associations, and further sense-making" [55]. Analyst intentions throughout this process adjust as fluidly, thus a CA should be cognisant of this.

For a CA to be used in intelligence analysis it is important that it can respond to new topic intentions, learning when it has found a good response and adapting when it has not. Rather than attempt to design responses for every possible intention an analyst has, we believe a better approach is to allow the CA to learn new intentions as it is used. However, changing capabilities present additional challenges for trust in a system, where we cannot be sure that a CA will process questions in the same way if we repeat them in the future, and transparency is therefore even more critical for systems that learn and evolve their capabilities.

2.3 Transparency, Brittleness and Formal Concept Analysis

Formal Concept Analysis (FCA) is an analysis approach which is effective at knowledge discovery and provides intuitive visualisations of hidden meaning in data [4]. FCA represents the subject domain through a formal context made of objects and attributes of the subject domain [41]. It can be seen as an unsupervised machine learning technique [12], where we can extract objects and attributes from the natural concepts found in the data and organised in a partial order structure. FCA is particularly helpful for revealing distinct concepts with semantic meaning across large amounts of data, and thus is well suited for delivering transparent analysis with appropriate contextual understanding. The lattice structure formed by FCA can be seen to represent an ontology graph of interconnecting higher and sub-concepts, where "by contrast to time-consuming manual building of domain ontologies the formal concept analysis establishes concept lattices automatically" [21].

While manual processes are brittle, FCA allows an ontology to evolve. Some limitations exist, such as the "naming of complex formal concepts " and "mapping concepts to their descriptions" [21], however the significance of these depends upon the purpose of the FCA. FCA has been used to analyse crime data to identify groups of offenders committing the same crime types in the same locations [41], and the use of violence by organised crime in drug trafficking [4]. To the authors knowledge, however, FCA has not been applied to develop and package modules to build system capabilities. We describe how objects and attributes that represent analyst requirements can be accurately captured, or learned, through interactions with a CA. FCA can then provide a transparent and dynamically evolving model for creating intention concepts from associated objects and attributes. We believe FCA is particularly effective at enhancing transparency, due to its use for knowledge representation and the ability to delve into concepts to understand the objects and attributes contained within.

Evolving CA intentions are a significant advancement on current approaches, such as pattern matching, where rather than having to hardcode intentions they can be formed fluidly based upon interactions with an analyst and an understanding of the attributes that address their needs. The lattice graph, which results from FCA, also assists the CA with explaining its intention and identifying similar concepts that can be suggested as alternatives to the analyst.

3 MODELLING CA INTENTIONS (USER STUDY 1 [16])

3.1 Objectives

In order to understand how a conversational agent (CA) could interact with an investigator, in terms of the types of questions that can be asked and the intentions it can fulfil, we first needed to gather data on how analysts think through interviews. We sought to understand the cognitive strategies applied by an analyst in an investigation, so that this could inform the modelling of intentions. In this section, (i) we analyse the transcribed interview data to identify the questions asked by analysts as they conducted their analysis, together with the requirements they had for responses that allowed them to advance their investigations. (ii) We use our analysis to model conversational agent intentions appropriately.

3.2 Methodology

We performed Cognitive Task Analysis (CTA) interviews with four analysts, each with more than three years of experience. Three of the analysts have worked in policing, across various police

Table 1. Interview Statements and Extracted Conversational Agent Questions

Transcript statement capturing requirement	**Summary of information needed**
One of the neighbours had suspected he had been kidnapped, and a witness had seen him being bundled into a car and alterted the police because they knew he was vulnerable. [Study 1, Analyst 1, Timestamp: 12:10]	'Person 1' could be victim (lives in 'Location 1') and has vulnerability for learning difficulties and history of being bullied.
Questions extracted from statement	**Responses required from questions**
1. Who lives at 'Location 1' (Kidnap Location)? 2. Does 'Person 1' have vulnerability markers?	1. 'Person 1' lives at 'Location 1'. 2. Yes, 'Person 1' has marker for learning difficulty.

forces, and the other analyst had a background in defence intelligence. Each interview lasted an hour and applied the Critical Decision Method (CDM) [26] [53] to elicit analyst expertise, cues, goals and decision making on a memorable investigation they were involved with from start to end. The interviewer began by asking more general questions about the day to day role of each analyst, before asking them to focus upon describing a particular investigation.

The CDM interview technique was used to ensure important information was captured. Of particular interest were the nature and requirements of analyst questions at critical stages in an investigation, specifically, their cues, goals, expectations and actions. These stages are typically time-pressured and are therefore prime situations in which conversational agents could be of assistance. Interviews addressed the analyst's experience, such as the conditions which allowed them to use their prior knowledge and the recognition of situations which a novice analyst may have missed. A timeline of key events was sketched out by the analyst and explored in detail to identify where, why, and how the analyst recognises and responds to each situation. We have looked to identify themes across critical investigation stages, including how analysts recognise situations and make decisions.

3.3 Results

The interviews in Study 1 were transcribed and a spreadsheet was populated with the specific statements made. Throughout the interviews the analysts described many types of questions which were asked during investigations, each with different information requirements implied. Table 1 presents how specific information retrieval questions and requirements were extracted from interview transcript statements, together with the implied information needs. The questions and responses are representative, demonstrating how the required information could be retrieved through interactions with a conversational agent (CA). For each statement in each interview, we have inferred a summary of the information required. From this we have derived question-answer pairs to provide examples of interactions with a CA that deliver the necessary information i.e. the question in Table 1, "who lives at Location 1 (kidnap location)?", is paired with the answer "Person 1 lives at Location 1." Another example, taken from a kidnapping scenario, is where the analyst described that "*The first thing I did was I looked through every database for the victim's name, custody records, PNC (Police National Computer), stop and search, vehicles he drove, to see if he had*

Table 2. Example Objects and Attributes [19]

Aspects of Recognition-Primed Decision (RPD) Model (Attribute Types)	**Descriptions of attribute processes to answer question: "Has [victim name] been reported in any activity?"**
Cues	Pass specific input details (Victim Name, Activity)
Goals	Present confirmation
Expectancies	Expected that input details and pattern exist
Actions	Perform adjacent information search for entities extracted
Why?	Retrieve list for further exploration.

been stopped and searched with other people in the vehicle and if they had been named. You have to keep interrogating lots of different databases." [Study 1, A1, 15:00]. For this statement the analyst requires a summary of past activity involving the victim and we can propose a number of questions to direct towards a CA to address this, including "what people in the system are connected to the victim?", "has the victim been linked to previous incidents?", "has the victim been an offender or victim?", "does the victim own a vehicle?"Alternatively, to cover everything, an analyst may ask the CA, "what do you know about the victim?" For each question, we have also postulated the response desired by the analyst which allows them to continue with their investigations, for example, 'the victim has previously been a victim of assault in 'Event 1' and 'Event 2'.

3.4 Discussion

We found that the key aspects of each question event in the investigations could be described succinctly using the recognition aspects of the Recognition-Primed Decision (RPD) model [24]. The RPD captures the way that an analyst recognises situations and makes decisions on the best course of action. These recognition aspects are:

- Plausible Goals
- Relevant Cues
- Expectancies
- Actions 1...n

We transformed each question-answer pair into a set of RPD attribute functions that could deliver the response the analyst needed. For example, to answer the question, "how many vehicles have travelled to the victim's address?", the analyst provides the cues 'vehicles', 'travelled' and 'victim's address'. Their goal is to retrieve summary information i.e. 'how many', and they are interested in finding a specific pattern of data in the database, which connects the cues. In another interview the analyst describes a situation where they were looking for phone calls to a suspect's phone where, once they identified a number which was of interest, they "*then go and find other phone calls*" [Study 1, A4, 14:15] from that number. They could, thus, ask a CA "how many times has [phone number of interest] called the suspect's phone?" This question comprises the same RPD aspects as the first example, albeit with different cues. Table 2 provides an example of RPD aspects that address a different question. A model of human recognition is also useful for providing contextual visibility of the processes performed by a CA. When an analyst asks a question, the CA needs to first recognise their intention and then explain how it has behaved in a way that a user will recognise. We have begun to identify and consolidate these aspects, as shown in Table 3.

By breaking down question and answer pairs and structuring their components against the RPD model, we extract attributes that can be used by a CA to process a response i.e. through

Table 3. Consolidated Decision Analysis Table [16]

Aspects of Recognition-Primed Decision (RPD) Model (Attribute Types)	**Descriptions of example processes to address RPD model aspects. Each process or function can be defined as an individual attribute, of the associated type.**
Cues	Inputs for Who, What, Why, Where, When, How (person's name, vehicle registration, time span etc...) and relationships where necessary.
Goals	To retrieve summary information, or specific details
Expectancies	Expected event pattern for scope informed by past events with similar scope (experience).
Actions	For information retrieval these include: adjacent information (i.e. who is registered to phone number, or who are their associates), connected information (i.e. what associates linked to a telephone number called by an offender live in a particular location), common connections (i.e. in what locations have both phone numbers been together) amongst others.
Why?	To build on, refute, or confirm scope and associated pattern.
What For?	To advance the investigation

the combination of code modules and the creation of SPARQL query syntax to interact with a knowledge graph. By associating each attribute to an aspect of the RPD model we effectively encode human recognition into the functional processes selected by the CA, thus enhancing visibility of that processing.

In this study, we extracted specific RPD attributes that addressed over 600 analyst questions. Within these questions there were many different combinations of attributes. For example, the question "Has [victim name] been reported in any activity?", requires the action attribute, 'Perform adjacent information search'.

A complete 'intention' is formed when attributes are grouped so that each aspect of the RPD model is represented. To group attributes, we applied Formal Concept Analysis (FCA). FCA allows a hierarchy of concepts to be generated automatically based upon associations between objects and attributes. FCA requires a formal context; this is a table representing relationships where objects have certain attributes. Going back to the question, "Has [victim name] been reported in any activity?", the question is the object and the attributes are the RPD model specifics, which have a relationship with the object.

Grouping attributes with FCA presents a number of benefits. Firstly, it simplifies the classification task by providing question strings with distinct concept labels that can be used for training, so requires less data. This is significantly less complicated than training a system to consider each question individually, with various associated attributes. In order to succinctly and consistently present the behaviour of a CA to a user, we wanted to be able to explain the different attributes triggered using the RPD model, and FCA preserves the underlying attribute structure so it can be inspected. To help a user understand the system behaviour we also wanted to be able to explain alternative but similar intentions, and their associated questions. Deriving distinct intention concepts, rather than having lots of questions each with their own attributes and structure, therefore

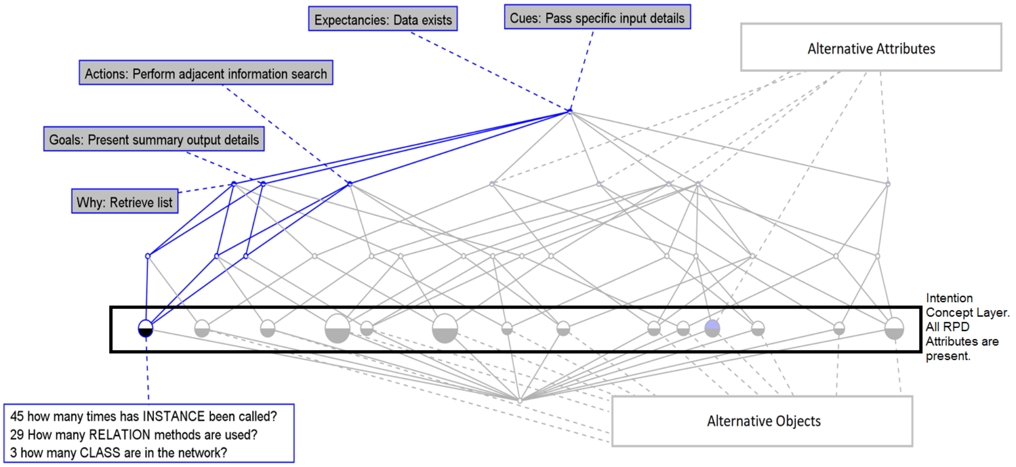

Expectancies: Data exists
Cues: Pass specific input details
Alternative Attributes
Actions: Perform adjacent information search
Goals: Present summary output details
Why: Retrieve list
Intention Concept Layer. All RPD Attributes are present.
45 how many times has INSTANCE been called?
29 How many RELATION methods are used?
3 how many CLASS are in the network?
Alternative Objects

Fig. 1. Concept Lattice for RPD Model Intentions (computed and drawn with Concept Explorer [56]) [19]. This lattice illustrates how intention concepts are formed. This visualisation is not presented to the user. At the highest level, all question objects are associated with the same attributes. At lower levels, concepts diverge and become more distinct.

aids explanation of the system behaviour as intention concepts can easily be compared. For example, a user can see what other questions match a selected concept, or any other concept. Additionally, because FCA provides a concept hierarchy, a user can explore alternative similar concepts where there are shared attributes, but some differences. The hierarchy allows an analyst to see how each concept is derived, and where gaps in attribute combinations exist. An important aspect of transparency is for a user to be able to "actively explore the states and choices of the AI, especially when the system is operating close to it's boundary conditions" [20]. FCA makes it much easier for a user to understand the boundaries of system behaviours, by presenting the possible range of intentions succinctly. This would be significantly more difficult if we did not calculate intention concepts, and associations between them, and instead knew only the attributes that address a given question. FCA can be performed automatically, and therefore allows new intentions to evolve, if a new combination of attributes is observed.

The lattice, as shown in Fig. 1, forms distinct object groupings, or concepts, based upon shared attributes, and these are represented as nodes. Each layer of the lattice captures the functional attributes that are shared by different question objects and we can see how concepts diverge as they move down levels, where the combinations of attributes become more distinct. At the top level we can see that all the questions we extracted from the interview data involve the attributes, 'Expectancies: Data exists' and 'Cues: Pass specific input details'. This is a clear constraint, for example, if a user desired an alternative function to handle the cues in a question, when no such concepts exist. The lattice helps highlight the constraints in the possible intentions and attribute functions that can be triggered. The path highlighted in blue shows the concepts that have the selected concept (containing objects 45, 29 and 3) as a sub-concept. This visualisation allows us to clearly see attributes that are more or less common across concepts, as well as semantically similar concepts. The second to final layer of concept circles represent complete concepts, where all RPD attributes are present, and are sized based upon the number of associated objects. We can see that questions linked to objects 45, 29, and 3, are answered by combining the same set of RPD attributes.

In our prototype application, each attribute triggers a module of code that appropriately processes the data or creates some SPARQL query syntax. To answer each of these three questions we can use the same core modules of code with different input cues. For example, if we are interested in the number of vehicles in our data we may ask "how many vehicles are in the network?" and we can see the relevant attributes from which to draw code by inspecting the highlighted concept and paths in the lattice (Fig. 1). We would use specific inputs extracted from our query as cues i.e. 'vehicles'. We are expecting that 'vehicles' exist and we look for adjacent information i.e. where there are instances of the class 'vehicles'. We present a summary count in response, and we retrieve a list of the vehicles, assuming that the analyst wishes to explore these further.

Given that intention concepts and attributes are generic, where the results returned by a CA are framed by the entities extracted from the question text and used as cues, attribute functions can be combined and applied agnostically to the underlying data. A typical problem with FCA is the naming of concepts, however, in our case it does not matter what a concept is called or how it is summarised. We merely need to understand what attributes, and therefore modules of code, underpin it, so that we can provide transparency of its processes. When formed in this way CA intentions are inherently explainable, because they are simply a combination of modules to address each aspect of the RPD model and we can provide descriptions for each. This approach chimes with what Lipton defines as decomposability for model interpretability, that each component part of a model "admits an intuitive explanation" [32]. We have taken a further step to frame decomposability with human recognition by consistently applying the RPD model. Traditional CA architectures, for which a user's utterance triggers an intention that cannot be broken apart, present black boxes where processing is hidden from the user. Molnar [36] explains that the future of machine learning interpretability will be model-agnostic and modular, we go some way to achieving this for the interpretability of CA intentions. We have used the concept lattice approach described to provide the 'brain' of a prototype CA that can classify an intention concept from user inputs and respond transparently. Our approach is also flexible and does not require a developer to code every intention in advance, where other common approaches do.

Study 1 provided the basis for modelling CA intentions that informed the build of a prototype application. In studies described later in this paper, we sought to test and evaluate our approach in terms of both the technical implementation and whether our RPD intention structure aids recognition.

4 UNDERSTANDING CA RESPONSES (USER STUDY 2 [19])

4.1 Objectives

We sought to understand what requirements intelligence analysts have for CA responses, including explanations of the various components of the system and any concerns. This would inform how we developed a prototype system for investigations. Again, we applied CTA, but this time to understand analyst cognitive strategies related to explanations of system behaviour, with reference to an investigation scenario. In this section, (i) we analyse questionnaire responses to identify the different components of a CA system that require explanation and (ii) we present how explanations can be tailored to specific needs.

4.2 Methodology

We interviewed four experienced intelligence analysts, each with more than 10 years working in an operational role, to explore their requirements for understanding the responses and processes of a CA in the context of a criminal investigation. These analysts were different to those interviewed in Study 1. Our focus was on using a CA for information retrieval tasks, to aid our design of response

Table 4. CA Condition 1: Data Only Explanation

Q: Are any suspects connected to James White?
A: Yes, James White is connected to Frank Howes and Paul Keen. James White was victim in an assault where Frank Howes and Paul Keen were suspects. Concept match confidence = 0.92.

Table 5. CA Condition 2: Data and System Explanation

Q: Are any suspects connected to James White?
A: Yes, James White is connected to Frank Howes and Paul Keen. James White was victim in an assault where Frank Howes and Paul Keen were suspects. Concept match confidence = 0.92. I am looking for a connecting path between James White and suspects (Descriptor). I am looking for single shortest paths and therefore, will not consider more complex connections.

explanations in a prototype application. Interviews therefore covered a scripted investigation scenario where an analyst is asking questions of a CA. Each interview lasted approximately an hour. We presented interviewees with a series of questions and corresponding CA responses, with two explanation conditions. The questions asked were:

(1) What might you need to understand about an individual output? Why?
(2) What might you need to understand about the whole system? Why?
(3) What is more important to you; explanation of individual outputs or explanation of the whole system? Why?
(4) What was most helpful about the explanations presented? Why?
(5) What was least helpful about the explanations presented? Why?
(6) What was missing / how could the explanations have been improved?
(7) Can you reflect on the key differences between the two conditions you have been shown. Do you have a preference towards one condition or the other?
(8) If you had a conversational agent, what would you ask it? What would you want it to do?

For one condition, responses described the data alone (Table 4) and in the other condition, responses described the data and the system processes (Table 5). All data was fictional and the text was predefined and created by hand. The outputs reflected those that could be produced by a system, however in this study the participants did not interact with a real system. For both conditions, we provided an example accuracy measure to reflect the confidence with which a concept match was made, in the form of a numerical value between zero and one.

We switched the order in which we presented conditions to analysts, so that two analysts saw condition 1 then 2, and two saw condition 2 then 1. We did this to remove any ordering effects, for example, when we asked analysts to identify additional understanding needs or to compare the two conditions. We were not attempting to test the differences between the two conditions, rather we used the conditions as a starting point from which we could explore additional needs with the analysts. We acknowledge that it is difficult for an analyst to provide a detailed assessment of a system without first experiencing it in a realistic scenario. However, the purpose of this study was to form a high level understanding of the broad system components that need explaining and the

Table 6. CA Explanation Area Framework and Sub-Themes [19]

Framework Area	ETA Sub-Themes
Clarification	Clarification of data attributes and structure, entity details, system input variables, metrics, question language, system processes, response methods, response language.
Continuation	Provide information to support continuation of investigation, including use of past interactions to move to next.
Exploration	Associated/additional data in responses or on periphery, intention match, system processes, source documents.
Justification	Provide information to justify selected system processes and the data defining the response.
Verification	Additional details for entities, correct intention match and impact/constraints of system processes. Check data reliability.

nature of the explanations. This approach also allowed us to identify limitations that we could rectify in our interactive evaluation study.

4.3 Results

Before analysis, we reduced the interview data to identify individual statements made by analysts in discussions as they answered questions. In total there were 114 distinct statements extracted, with counts for each analyst ranging from 24 statements to 34. To analyse the statements we used an approach called Emergent Themes Analysis (ETA), as described by Wong and Blandford [7, 54], where broad themes, which are similar ideas and concepts, are identified, indexed and collated. A single researcher coded all of the data, to ensure consistency.

ETA allows the researcher to build the theory from the data, where a top down approach feeds from broader themes to more specific and fine line concepts. It is useful for giving a feeling of what the data is about, with structure, and is fast and practical [28]. Through ETA we derived a framework for the different components of a CA system that require tailored explanations.

We identified that analyst statements could be broadly associated with the core functional components of a CA that need explaining, for example entity extraction from the user's query, or the system processes applied. This is an interesting finding, indicating that analysts have specific considerations for each function of a CA. We should not provide explanations of the data in the same way as we provide understanding of the intention classification, or the extracted entities, the response, or system processes. This perspective aids our approach to transparency where we can identify and define explanation needs for distinct system components.

From broad CA component themes, we have drawn out the detailed understanding needed by an analyst in an explanation. These are the sub-themes. The sub-themes are further categorised to form a general framework (Table 6) for explanation needs from an intelligent CA system. Table 7 provides a detailed example of how broad themes contain sub-themes, which are consolidated within a framework area, for selected statements about 'System Processes'. The sub-themes are shown where statements relate to either clarification of system input variables or clarification of

Table 7. ETA Snapshot for Clarification of System Processes [19]

Broad Theme	Sub-Theme	Framework Area	Statement
System Processes	Clarification of system inputs. Clarification of system processes	Clarification	I am concerned that info is missing because of search criteria. Understanding as a tool is also important for the whole system, such as when and where to use it. How have the results been worked out and what methods have been applied?

system processes. The needs for understanding described in the statements can be addressed by providing greater clarity, and thus we have created a framework area for 'Clarification'.

Exploring the interview data through the Emergent Themes Analysis (ETA) method and structure is helpful when we come to design CA components. For example, examining Table 7 again, we can see that to provide understanding of system processes to an analyst we need to allow for clarification of both input variables and processes. Drawing upon details in the statements, we can see that it is important to clarify any constraints related to the search inputs, the general capabilities of the system as a whole, and specific processes applied in any instance. We incorporate explanations that provide clarification of these aspects, in addition to solutions for other themes extracted through ETA, into the design of our prototype application.

An analyst's ability to have clarification, verification, and justification of system processes is crucially important, as identified by all the analysts interviewed. This finding supports the significance of system visibility, as defined in the framework for providing algorithmic transparency presented by Hepenstal et al. [15]. It reiterates the need to go beyond traditional approaches to explainable AI (XAI), which focus upon explanations of the features for a model and accuracy measures. None of the analysts raised understanding the machine learning intent classification as a concern, beyond wanting to know how to phrase questions and to see past questions from other analysts. Much greater interest was placed upon the evidence that underpins a response and the system processes applied. In fact, analysts found intention classification accuracy confusing and unhelpful, because they thought it related to "*the data itself, not the classification*" (Study 2; Analyst 3; Question 1; Condition 1). This is an interesting insight, given that much of XAI research looks to confidence scores as a method of explanation.

Analyst statements predominantly relate to a need to justify follow up questions and the underlying rationale of the system for use in court (Study 2; A1; Q2; C1). Additionally, an understanding of the system processes selected by the CA, including descriptions of the methods applied (all analysts, multiple statements), and inherent constraints, such as the questions which cannot be answered by the CA and information which has been omitted by the process (Study 2; A2; Q1; C2 | Study 2; A3; Q2; C2; | Study 2; A4; Q4; C2). Essentially, analysts need to be able to justify, clarify and verify the CA intention triggered by their query and the related functional attributes. Analysts require this understanding once their query is complete, so post hoc explanations appear to be

Table 8. System Component Explanation Framework [19]

CA Component Theme	**Framework Area (common for multiple analysts)**	**Summary of Sub-theme(s)**
CA Intention Interaction	Clarification (3), Continuation (2)	Clear language to understand classification (i.e. no confusing response metric) and information to support continuation of investigation.
Data	Clarification (3)	Clarification of data updates and source, and data structure to aid forming questions.
Extracted Entities	Clarification + Verification (3)	More information of entities extracted for clarification and verification.
Response	Clarification (4), Justification (4), Exploration (2)	Justification of response with underlying data, clarification of language (not trying to be human) and terminology, ability to explore results in more detail.
System Processes	Continuation (4), Verification (4), Clarification (3), Exploration (2), Justification (2)	User wants system understanding to support continuation of investigation, to allow them to verify processes are correct and explore them in more or less detail and justify their use/approach and constraints.

sufficient for this use case. We believe our design provides a neat mechanism to pick apart the system processes and provide, for each, the understanding required.

In Table 6, we display the framework areas and related sub-themes, which emerged from ETA. Specific areas in the explanation framework can be linked to existing models for sensemaking, such as the Data Frame Model [27] for elaborating and reframing questions, or Toulmin's model for argumentation [50] to provide justifications. Thus, to aid sensemaking with the CA application, for each framework area, we can design explanations that comply with existing sensemaking structures.

Table 8 presents the key framework areas for each component theme, where at least two analysts made associated statements, together with a summary of sub-themes specific to both CA component and framework area. Different CA components draw more heavily on particular aspects of the framework and therefore the explanations for individual components should be tailored. For example, to understand the underlying data that feeds the CA, analysts desire an explanation that covers clarification of data source, currency, and structure. However, to understand the answers given by the CA in its response, they also want an explanation that provides justification i.e. the reasons or logic for why the answer has been given, in addition to aids to explanation which allow them to explore related data in more detail.

4.4 Discussion

Our analysis, structured through ETA, helps us to design and tailor the explanations of individual CA component parts. The framework also has broader application than CAs. The themes and sub-themes could equally apply to the design of other intelligent applications for use in criminal investigations. For example, a system to search an offender database with a suspect's image and apply facial recognition to retrieve possible matches. In this case an analyst would need to be able to clarify and verify data attributes, metrics, and methods, such as what features are important to the model and whether the suspect's image provides the necessary information. If it does not, then they should be aware that an alternative image is required, otherwise results are unlikely to be satisfactory. The analyst would also need justification for the information retrieved, in terms of the features within the data that are most significant and the system considerations and constraints. Results should also allow for continuation of the investigation and more detailed exploration, for example to suggest improvements to the suspect image, which is the input, or allow exploration of other supporting data for a match.

Trust was raised on a number of occasions by analysts in relation to "*trust in the development pipeline*" (Study 2; A4; Q3; C2) for the system process which has pushed out a result, and, "*trust and confidence in the data (reliability)*" (Study 2; A1; Q1; C1) which underpins a response. Even where it was not mentioned explicitly, the need to provide explanations to reduce uncertainty in the data, in what entities are extracted, in the intention match, system processes, and response, has an implicit relationship with how much the analyst should trust the output from the CA. For example, when an analyst asks for an explanation of "*the information source and how complete is the database?*" (Study 2; A1; Q1; C1) this has a direct correlation with their trust in the answer. If the source is reputable and corroborated and the database is current, they will likely feel more confident than if it is not. Likewise, where an analyst asks a question of a CA they are "*looking for clear, logical, impersonal response, with reproducible steps, as a justification for court.*" (Study 2; A2; Q1; C2) If the response provides justification to this effect then they will be more likely to trust the answer, than if it is illogical or impossible to reproduce and validate.

In intelligence analysis, trust cannot be a one-time feature where once earned it is forgotten. We must continue to provide these explanations so an analyst can feel confident to check the workings of core functional components in a system. We propose that by providing meaningful explanations that address aspects captured in our framework, analysts can be confident in each of the core functional components of our CA and we can develop and sustain trust in the system.

5 CA PROTOTYPE

We have built a prototype CA application called Pan [17], which uses Formal Concept Analysis (FCA) to define the different intention concepts to which it can respond. Pan was developed as an experimental prototype, to probe requirements further than was possible with the static prototype used in Study 2. Pan is currently a question-answer system that can retrieve information related to network analysis in an investigation, with a system architecture as shown in Fig. 2. There is scope to develop Pan to a fully conversational system, where it can converse across entire lines of inquiry, rather than merely answer questions. However, in this initial research we were more interested in the needs for transparency of system processes related to information retrieval, rather than conversational abilities. Pan, therefore, has been designed specifically with the Algorithmic Transparency Framework in mind (Fig. 3).

5.1 Formal Concept Analysis (FCA)

As explained in Study 1, each question in the training data has been associated with the functional attributes, which write and process database queries, required to answer it. We have mapped these functional attributes against the structure of the recognition aspects of the RPD model.

FCA takes analyst questions as objects and the associated functions as attributes, forming a hierarchy of concepts where collections of objects and attributes are shared. Complete intention concepts are those that have a functional attribute for every element of the RPD model. For the example highlighted in the lattice in Fig. 1, one attribute handles specific input details, combined with another that searches for adjacent information for the extracted entity, one configures the goals for the output text, and another defines what information is stored in memory for further investigation.

Once FCA was complete, we trained a simple text classification model to match a new user question to a complete RPD concept, using associated question objects and concept identifiers as labelled training data. To ensure the classification model is generic we have removed any specifics from questions. For example, the question "what vehicles are owned by James White?" is translated as "what class are relation instance?", as these aspects are present in our database. We required the text classification model to be accurate to the degree that users could interact with the system and trigger helpful intention concepts. This would enable us to evaluate the prototype with a realistic scenario. Our classification model was based upon the pre-trained sentence embeddings method, InferSent [9] for English sentences. We used version 2 of the InferSent model, trained on fastText vectors [35]. We had only 658 labelled question objects to build our classification model. There were 10 complete concepts when our model was initially built and these defined the possible classes that could be predicted by the model. We used stratified sampling to ensure a similar proportion of different classes appeared in both training and test datasets, where the test data comprised 20% of the overall data. Our model was able to achieve over 99% test accuracy overall. Importantly, it could classify objects and retrain quickly, as required in a conversational interaction. Some concepts were overrepresented in the dataset, however given that the classification model could correctly classify and trigger all of the 10 intention concepts, this was considered sufficient for the experimentation. When a user asked a question of Pan, any entities would first be extracted from their input. These entities provided the cues that were then processed according to the functional attributes of the matched concept. In this way, FCA could combine multiple distinct combinations of attributes flexibly to meet different analyst intentions. We propose that, by combining RPD model-based attributes with FCA to define complete concepts, we provide a highly flexible and recognisable approach to form CA intentions. This design allows for CA learning, both to reinforce the classification of existing concepts and to evolve new concepts. It is also possible to model and predict paths of questioning, by drawing links and passing entities between concepts in an investigation [18].

The question objects and corresponding RPD functional attributes are critical for providing visibility to an analyst for the responses given by a CA and are akin to explainability scenarios i.e. "narratives of possible use that seek to answer the questions: who will use this system and what might they need to know to be able to make sense of its outputs?" [52] For our prototype, the CA retrieves information from a semantic knowledge graph. Therefore, the attribute functions relate to the construction of SPARQL queries, graph traversal methods, and processing of the data returned. Graph traversal methods, or 'actions' within the RPD framework, are loosely based upon the types of graph task identified by Lee et al. [29]. Natural language responses to the user are configured using an appropriate model together with the information retrieved from the knowledge graph. Therefore, the manual creation of responses and curation of a response library is not required. When matched,

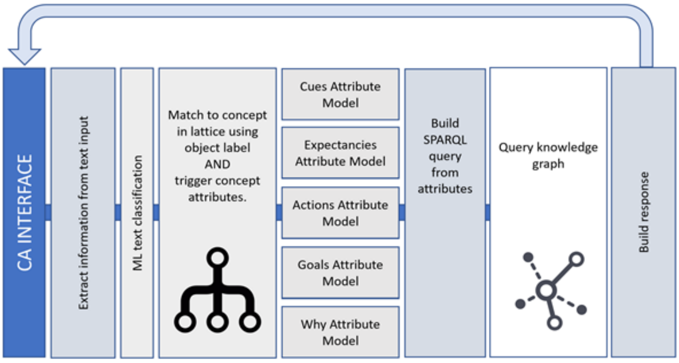

CA INTERFACE
Extract information from text input
ML text classification
Match to concept in lattice using object label AND trigger concept attributes.
Cues Attribute Model
Expectancies Attribute Model
Actions Attribute Model
Goals Attribute Model
Why Attribute Model
Build SPARQL query from attributes
Query knowledge graph
Build response

Fig. 2. CA Component Diagram: The user first enters a question or statement into the CA interface and entities are extracted as they type. When they click submit, intent classification is performed to match the user's input string to a concept in the lattice. In typical intention based CAs, the classification matches an intention which triggers a body of code designed to fulfil a specific task i.e. making a flight booking. Our approach is different where, rather than programming each intention individually, intentions are formed through FCA and the appropriate rules are defined by attributes within the matching concept.

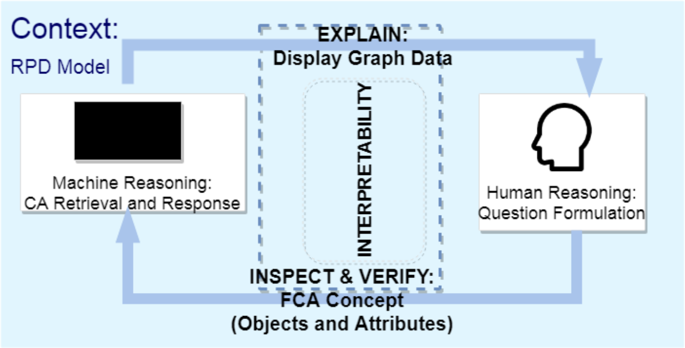


Fig. 3. Algorithmic Transparency Framework for CA Prototype

the concept builds a response from the related attributes which reflect the RPD model. In this way, Pan is designed for user interpretability with the Algorithmic Transparency Framework in mind. Fig. 3 shows the framework noting the specifics of our implementation. There are similarities between our framework and the knowledge generation model for visual analytics [43]. The knowledge generation model describes how visual analytics systems can lead to interesting user observations and findings in various ways. These include observations from visualisations, of either data or computer models, and inspection of the models or data directly. A finding related to computer models could be a conspicuous model result or identification of an unusual system behaviour. There are overlaps here with the delivery of system transparency. However, our transparency framework focuses upon two core perspectives to provide interpretability of machine reasoning. These are, (i) explanation of the behaviour, typically delivered through visualisations of the underlying data and important features, and (ii) inspection and verification of the machine reasoning behaviour. Specifically, this relates to the goals and constraints of the behaviour. Our framework emphasises the latter perspective, i.e. the need to inspect and verify the goals and constraints of system reasoning behaviour, including the models and functions triggered. As explained by Sacha et al.

[43], “visualization is often used as the primary interface between analysts and visual analytics systems whereas understanding the model often requires more cognitive efforts.” To adequately inspect and verify system behaviour we require a granular structure that reflects the context of use. We want analysts to recognise the goals and constraints of system behaviours, therefore we structure functional models in a way that aids recognition i.e. the RPD model. Interviews indicated that analysts were content with post-hoc explanations, where they seek understanding once a result has already been provided. We also address system brittleness by allowing CA intentions to evolve with FCA. We suggest that Pan is an example of shared human-machine reasoning, where a human analyst reasons to propose questions for information retrieval, and the machine interprets them to explore the data and respond.

5.2 Explainability

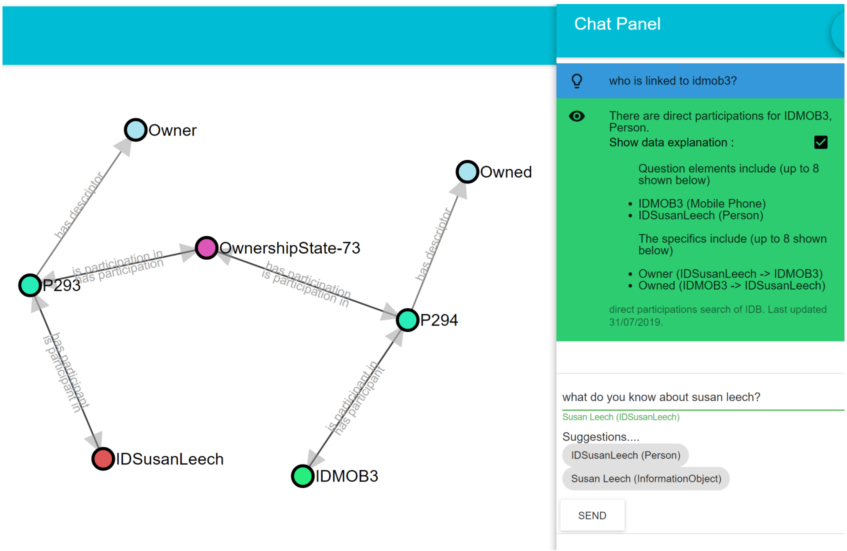
Fig. 4 is a screenshot of the CA interface which shows the result when an analyst asks a question, “who is linked to IDMOB3?” The CA matches their input to an intention concept, triggers associated attributes, and responds. Our system is designed with a particular context in mind; network analysis in criminal investigations. The design of explanations are therefore consistent with the user community, where complex network graph visualisations are familiar to most analysts.

The graph layout is force-directed, but can be rearranged by the user. Data can be added to the graph from any question asked, so that the graph can build from multiple interactions. On hovering over an answer (the eye icon), the relevant data in the graph is highlighted by manipulating the opacity of other data.

We provide the information required for an analyst to understand the CA component themes of ‘Data’, ‘Extracted Entities’, and ‘Response’. As analysts type their query and entities are extracted, they are provided with identifier information where possible. In Fig. 4, identifier information is provided for ‘Susan Leech’ where it reads ‘IDSusanLeech’. The response also gives clarification of the data, including the source, currency, and structure through a visual explanation of the underlying data. There is scope to simplify the visualisation, for example by hiding participation nodes (P293, P294), however we first wanted to explore the impact of presenting the complete underlying data on analysts using the system. The underlying data is also available for further exploration. Analysts can interact directly with the graph to retrieve details for each node, by double clicking nodes to view an information panel. They can also delve into additional data directly associated with a node by presenting it on the graph. Responses are short, impersonal, reiterate the entities that are included in the query for clarification and verification, and provide clear justifications along the lines of the argumentation model presented by Toulmin. The CA makes an initial claim, which is the response that ‘Susan Leech’ is linked to ‘IDMOB3’, and it provides the warrant for this claim where specific evidence for important links is described.

We made these design choices using the data and framework formed through ETA. By drawing out themes and sub-themes we could identify specific design features to address the original analyst statements. For example, for entity extraction, where analysts made statements, such as “*I want more details to verify entity extraction, so I can confirm the person is of interest*” (Study 2; A1; Q1; C1), or “*I want more detail of entities and relations, there is subjectivity*” (Study 2; A3; Q1; C1). We group these within the theme for entity extraction as the relevant CA component that needs explaining and sub-themes of clarification and verification of entity details, as the information required. Hence, in our design we provide the entity as a hint when it is recognised, together with a supporting identifier in brackets. Other design features have been considered and designed in a similar fashion.

The ability of the analyst to see the evidence which supports the CA response as a graph is akin to the definition of explainability given by Gilpin et al. [14], where the internals (data features) are presented in their entirety. This assists with the analyst’s need to explore data further, where



who is linked to idmob3?

There are direct participations for IDMOB3, Person.
Show data explanation : ☑

Question elements include (up to 8 shown below)

- IDMOB3 (Mobile Phone)
- IDSusanLeech (Person)

The specifics include (up to 8 shown below)

- Owner (IDSusanLeech -> IDMOB3)
- Owned (IDMOB3 -> IDSusanLeech)

direct participations search of IDB. Last updated 31/07/2019.

what do you know about susan leech?

Susan Leech (IDSusanLeech)

Suggestions....

IDSusanLeech (Person)

Susan Leech (InformationObject)

Fig. 4. Screenshot of section of prototype CA explainability: An analyst can choose to display the underlying data that informs a response as a graph in the main display area, as shown, or expand in a separate popup window. By clicking on the lightbulb icon, an analyst can see more information about the intention concept they have triggered.

it is helpful to have, "*short statements as answers, but need ability to drill down further into data*" (Study 2; A3; Q4; C1) and also to enhance "*understanding of the data structure and model*" (Study 2; A1; Q4; C2). In our design, we have allowed analysts to configure the layout to emphasise the structure of the underlying data by selecting the 'Data Structure Layout', thus assisting them to form a better understanding of how to frame their questions. It is important that analysts can interact and delve further into the data and explore the graph by opening pathways that lie on the periphery of the response. We believe that doing this helps to overcome the automation dilemma, that we will miss key insights through automation of the search process and leave ourselves open to error or deception.

In Fig. 4, we can see that a person has been found who owns 'IDMOB3'. From this information alone, we cannot understand how the CA has recognised and acted upon our input, which may be open to subjective interpretation. For example, what do we mean by 'linked'? In this instance, the CA has interpreted our intention to consider people that have direct participations in an activity with 'IDMOB3'. Thus, any more distantly connected people, for example a partner of Susan Leech, who lives in the same household, is ignored. Analysts need visibility to inspect and verify the goals and constraints of a CA in order for the analyst to trust that the CA has correctly understood their intention and queried the data appropriately. This must include the provision of methods, constraints and caveats for the response.

5.3 Visibility

Our approach to provide visibility does not improve the accuracy of intent classification itself. There will always exist subjectivities in language that cannot be resolved by more accurate machine learning. For example, two analysts may have different interpretations of what it means for entities to be 'linked' to one another. Both may be correct, as there is no clear universal interpretation, and this creates ambiguity that can lead to misunderstandings. We propose that our approach improves the accuracy with which system behaviour is recognised, interpreted, and understood. To achieve this, we model intentions with a granular structure that reflects human recognition. We use information granules derived from the RPD model to provide visibility of the system behaviour. We achieve this by presenting the analyst with the intention concept triggered, detailing each of the RPD model attribute methods applied. The analyst can inspect the cues, goals, actions and related constraints to verify them.

Fig. 5 shows the prototype display that allows a user to step into the 'brain' of the CA, when they click the lightbulb icon. We have not presented a numerical value to explain the accuracy of the intent classification, as this was described as misleading and confusing by all analysts interviewed in Study 2. Confidence and accuracy have a different meaning in the context of intelligence analysis, where the terms are typically used to describe the reliability of the information; this definition is therefore at odds with traditional research into XAI. Instead, analysts expressed a need to clarify the entities extracted from their question, so we describe these clearly.

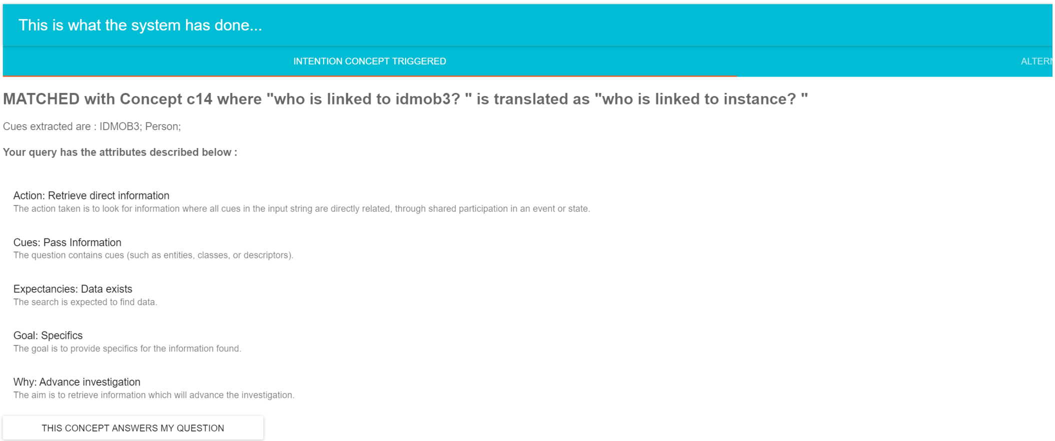
To enable analysts to understand the system processes we have provided simple descriptions to clarify each RPD attribute triggered, indicating the relative constraints. Individual descriptions are clearly defined, as they relate to specific modules of code and functions. For example, for the action 'retrieve direct information' shown in Fig. 5. An analyst can see that only data that directly relates all cues (IDMOB3 and a Person) will be returned. Thus, if they were expecting broader links to be considered, then they are mistaken. Analysts noted that such information helps "*prompt adjustments to query, due to better understanding of system processes, and leads to increased confidence in the response.*" (Study 2; A1; Q3; C2)

Our FCA approach to build intentions is helpful, either to confirm the attributes triggered are appropriate, or to identify objects from similar concepts by moving across the lattice hierarchy. These are semantically similar, given that they share attributes. A user can select to view alternative concepts (Fig. 6) that are similar based upon either machine learning classification, or distance in the concept lattice, that could be pursued to aid continuation of the investigation. We have developed specific design features based upon our ETA analysis. For example, where an analyst asks, "*what is the right language and way to ask the question?*" (Study 2; A1; Q1; C1) We associate this with the component theme for intention interaction, with the sub-theme 'clarification of question input language', because an understanding of how to ask a question will help users to match their desired intention. When analysts step into a concept they interact with alternative concepts and can see what questions to ask to trigger them. We highlight the key differences between attributes used in the matched intention compared to alternative intentions.

By allowing an analyst to interact with the intention of a CA, providing information granules that reflect the RPD model, we believe that the analyst will recognise and learn how best to speak with the CA to achieve their desired results. The CA will also learn how the analyst asks questions as it retrains.

5.4 Brittleness

We aim to mitigate the brittleness problem by allowing our CA to learn from the analyst and evolve possible intentions through FCA. Firstly, we allow the CA to retrain through supervised learning where the analyst can provide positive feedback to the CA by selecting that "This concept answers my question" and their question is added as a new object in the formal context, where the attributes are read across from the matched concept. The machine-learning model retrains and classification accuracy improves. Secondly, if the matched intention does not suit the analyst, the CA suggests others based upon both machine learning classification confidence and distance in the lattice. On


This is what the system has done...
INTENTION CONCEPT TRIGGERED
ALTERN
THIS CONCEPT ANSWERS MY QUESTION

MATCHED with Concept c14 where "who is linked to idmob3? " is translated as "who is linked to instance? "

Cues extracted are : IDMOB3; Person;

Your query has the attributes described below :

Action: Retrieve direct information
The action taken is to look for information where all cues in the input string are directly related, through shared participation in an event or state.

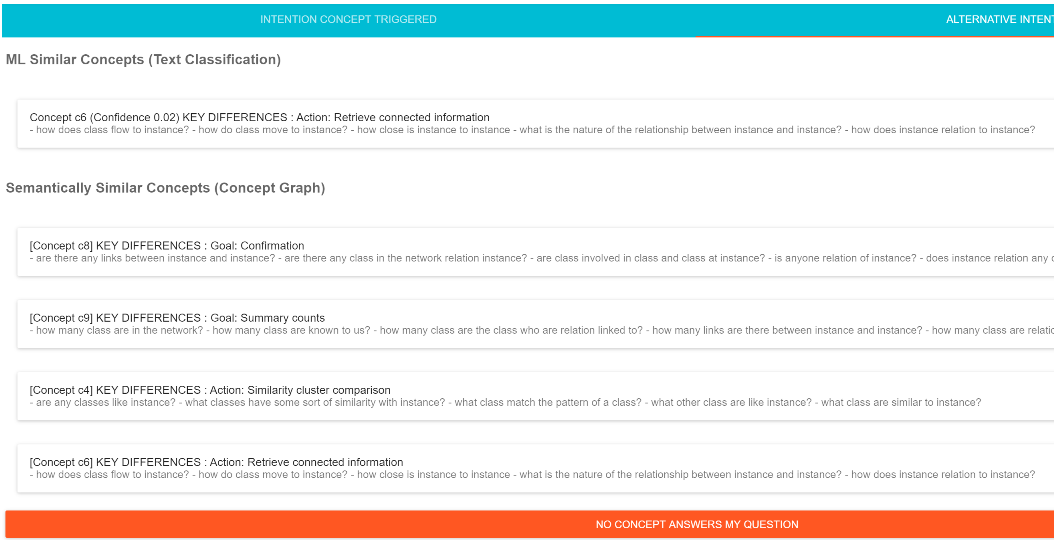
Cues: Pass Information
The question contains cues (such as entities, classes, or descriptors).

Expectancies: Data exists
The search is expected to find data.

Goal: Specifics
The goal is to provide specifics for the information found.

Why: Advance investigation
The aim is to retrieve information which will advance the investigation.

Fig. 5. Screenshot of section of prototype CA ‘brain’ display: A user can see the cues extracted from their question, in addition to descriptions for each of the functional attributes triggered. These are structured against the RPD model.



ML Similar Concepts (Text Classification)

Concept c6 (Confidence 0.02) KEY DIFFERENCES : Action: Retrieve connected information
- how does class flow to instance? - how do class move to instance? - how close is instance to instance - what is the nature of the relationship between instance and instance? - how does instance relation to instance?

Semantically Similar Concepts (Concept Graph)

[Concept c8] KEY DIFFERENCES : Goal: Confirmation
- are there any links between instance and instance? - are there any class in the network relation instance? - are class involved in class and class at instance? - is anyone relation of instance? - does instance relation any c

[Concept c9] KEY DIFFERENCES : Goal: Summary counts
- how many class are in the network? - how many class are known to us? - how many class are the class who are relation linked to? - how many links are there between instance and instance? - how many class are relatio

[Concept c4] KEY DIFFERENCES : Action: Similarity cluster comparison
- are any classes like instance? - what classes have some sort of similarity with instance? - what class match the pattern of a class? - what other class are like instance? - what class are similar to instance?

[Concept c6] KEY DIFFERENCES : Action: Retrieve connected information
- how does class flow to instance? - how do class move to instance? - how close is instance to instance - what is the nature of the relationship between instance and instance? - how does instance relation to instance?

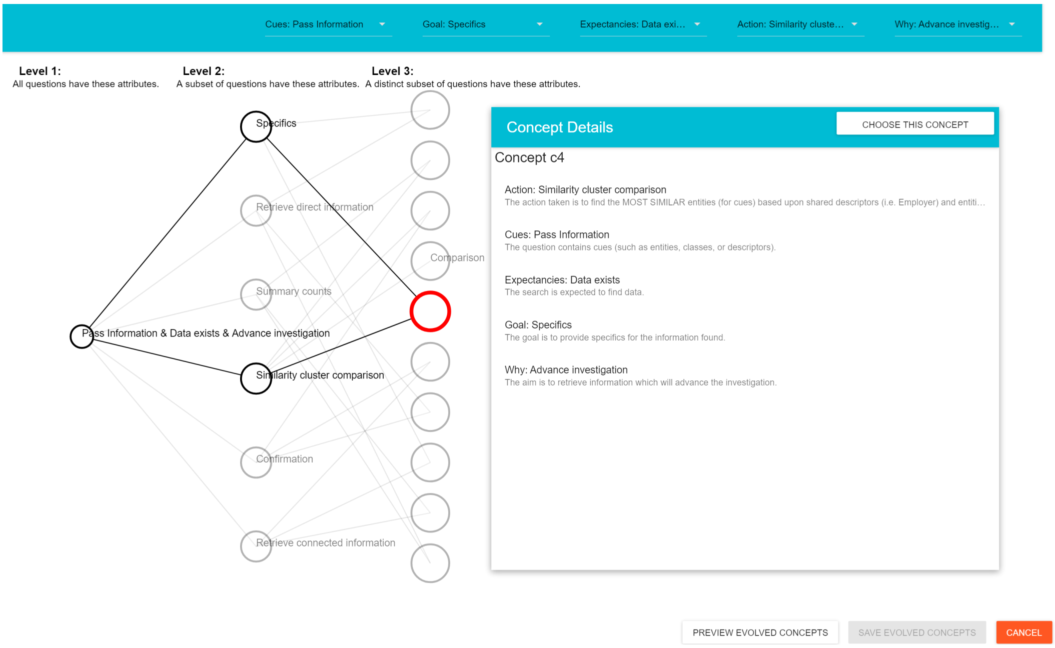
Fig. 6. Screenshot of section of prototype CA ‘brain’ display: Alternative concepts can be explored and attention is drawn to the key differences and questions that trigger them

selecting an alternative concept, the classification model also retrains and adds the object to the formal context with the attributes of the selected concept.

A similar approach to learn intentions is feasible with other intention classification systems, such as LUIS. However, if no appropriate concept exists then the intention of an analyst’s question may not have been previously considered. In typical applications, where topics are pre-defined in a library, a new intention topic would need to be created and this would involve coding. This is not something a user could typically do. By using FCA to automate the formation of intentions through concept generation, we simply need to identify the appropriate attributes that will address the user’s query even if they have never been combined previously in a complete intention. FCA therefore allows the possible intentions for a CA to evolve by combining attributes that are functional modules of code.

When an analyst selects "No concept answers my question" (Fig. 6), they are presented with the concept lattice displayed in Fig. 7. This visualisation is a working proof-of-concept to demonstrate that intention concepts can evolve dynamically through interactions with a user and the use of FCA. The visualisation in Fig. 7 describes all the possible combinations of attributes that are available to a user. The visualisation is designed to provide a quick understanding of the possible intention concepts available, and any constraints or boundaries. With this information we propose that a user can make more intelligent choices about how to interact with the system. The concept hierarchy is defined by levels, as a user moves from left to right. At the first level, the attributes are common to all questions that can be posed to the CA, so these will always be triggered. This is an important capability boundary for the system that needs explaining to a user, so they can manage their interactions or develop new capabilities. As the user moves to the second and third level, the associated attributes apply to a subset of the questions in the training data. The height of each level emphasises the diversity of concepts at that level. The concept lattice is labelled to show where attributes first appear. This is designed to help a user to identify concepts and sub-concepts that combine their desired attribute functions, as well as to highlight limitations. Not all attribute combinations exist, for example, the 'Comparison' goal attribute can only be used in combination with the 'Similarity cluster comparison' action attribute. A user can see this in Fig. 7, where the 'Comparison' attribute appears at level 3 and thus is not triggered by any other concept. A user can interact with the concepts by double clicking any of the nodes (when selected, the node is highlighted red and the attributes are described in the panel on the right) to see the functional RPD attributes triggered by each and the links between different levels i.e. how complete intention concepts are formed. A user can also evolve new intentions themselves. We capture the analyst's question (as an object) and allow them to select the appropriate RPD model attributes to answer their question from drop down lists (Fig. 7). For example, in Fig. 7, where the user asks the question, "what people are similar to James White?" (translated as, "what class are similar to instance?"), they require attributes that return specific details of people and compare similarities across clusters of instances. The analyst can select these attributes and choose to preview the evolved concept lattice. FCA recalculates with an updated formal context to include the analyst's question, as an object, and the selected attributes, to create a new concept. The concept lattice is redrawn and it presents how the newly created concept slots within other concepts, providing visibility of alternative associated capabilities. We propose that undertaking this process of concept evolution in itself gives the user a better feel for the constraints and capabilities of the CA. It helps provide an understanding of "*when and where to use*" (Study 2; A1; Q3; C1) the tool, awareness of which was noted as important by analysts. In this way, the CA intentions learn and evolve through interactions with an analyst. The approach is explainable by design, given that the combination of attributes will always reflect the RPD model. It is important that systems that evolve and change their capabilities over time are transparent, particularly in the area of intelligence analysis where it has been noted in interviews that the ability to retrace your steps and accurately audit your work is crucial. We propose that, while the system capabilities will not remain static, so a question in the future may trigger a different response to the same question today, we mitigate issues by providing sufficient explanation consistently structured through the RPD. We believe this consistency will help engender trust in the CA.

Our approach partially mitigates the brittleness problem, because whilst new intentions can evolve, they still require the creation of functional attributes. It may, therefore, be the case that an intention is desired by a user that does not exist and cannot be evolved, without writing a new



Cues: Pass Information
Goal: Specifics
Expectancies: Data exi…
Action: Similarity cluste…
Why: Advance investig…

Level 1: All questions have these attributes.
Level 2: A subset of questions have these attributes.
Level 3: A distinct subset of questions have these attributes.

Concept Details

Concept c4

Action: Similarity cluster comparison
The action taken is to find the MOST SIMILAR entities (for cues) based upon shared descriptors (i.e. Employer) and entiti…

Cues: Pass Information
The question contains cues (such as entities, classes, or descriptors).

Expectancies: Data exists
The search is expected to find data.

Goal: Specifics
The goal is to provide specifics for the information found.

Why: Advance investigation
The aim is to retrieve information which will advance the investigation.

Specifics
Retrieve direct information
Summary counts
Pass Information & Data exists & Advance investigation
Similarity cluster comparison
Confirmation
Retrieve connected information
Comparison
CHOOSE THIS CONCEPT
PREVIEW EVOLVED CONCEPTS
SAVE EVOLVED CONCEPTS
CANCEL

Fig. 7. Screenshot of prototype concept display and evolution. A user can select attributes that meet their needs and can see what concepts exist. They can explore the graph and select a concept. They can also evolve the possible concepts, through Formal Concept Analysis, if one does not exist that meets their needs.

function. However, it is easier to write a new function, or model, that can slot between others using FCA, than to build an entire intention. In our prototype, we are interested purely in delivering information retrieval for network analysis, therefore it is possible to structure the functional attributes required with the RPD model. In a different domain, or context, this may not be possible nor appropriate. We believe our approach may have wider usage in other contexts, however this has not yet been explored.

The prototype display for presenting and evolving intention concepts (Fig. 7) requires further development, user testing and evaluation. So far, we have not considered in detail how to represent this evolution to the user, including the appropriate layout and interactivity of the lattice, so evaluating and iterating upon this is a key area for future work.

6 EVALUATION (USER STUDY 3)

6.1 Objectives

Studies 1 and 2 resulted in the development of a prototype conversational agent, called Pan [17]. In Study 3, we evaluated Pan, by allowing analysts to ask questions of it to investigate a fictional scenario. In Study 2, analysts were simply responding to what they were shown. They speculated on their concerns and provided assessments about what helped them to understand the responses from the CA. While interesting, these findings were limited given that the analysts did not have the opportunity to experience the CA for real, and their self-reported levels of understanding may not have reflected true understanding. They were not required to direct the investigation themselves, therefore were less invested in interrogating the scenario so that they could draw a conclusion. In Study 3, we sought to rectify these issues.

6.2 Methodology

10 operational intelligence analysts were recruited from a range of organisations, with backgrounds that included the police, National Crime Agency (NCA), military, and prison service. These analysts were not involved in the previous study and all were required to have held a role involving network analysis. The participants had a minimum of 3 years full time operational experience. Participants were paired by closest similarities in their background, organisation, and length of experience. Each pair was split, with one assigned the condition with transparency and the other the condition without, to ensure a degree of similarity between groups. As with the predefined questions and responses, for one condition analysts were given the ability to view textual responses from Pan, including details of the data that underpinned a response. They could visualise this data with a network graph (Fig. 4). For the other condition, analysts could access and visualise the data in the same way, however, they were also encouraged to step into the triggered intention to inspect and verify the functional processes (Fig. 5). The focus of our study was on understanding the provision of system transparency, therefore we did not provide the analysts with the ability to evolve intention concepts. After the investigation exercise was complete, analysts were asked a series of interview questions. These were adapted from Study 2. Analysts were remote, thus they interacted with Pan via a researcher who shared their screen via video conferencing software. A single researcher performed all of the experiments and followed a checklist of activities, ensuring consistency. The audio from experiments was recorded and transcribed, each interview lasting approximately an hour. Analysts posed 14 questions to Pan, on average, during the interactive exercise. Our interactive experiment allowed us to validate our earlier findings on the components that require explaining and the nature of the explanations. We also evaluated Pan as a visual analytics tool, specifically how the interactive visualisations and transparency provision helped analysts to reason about hypotheses and draw insights.

In order for AI systems to be used for high risk and high consequence decision making they must provide transparency of their reasoning. As put by one analyst, "*[the principal analyst] said none of my analysts would stand up in court where the beginning point of their evidence is an algorithm.*" [Study 1, A4, 32:30] and that, "*you have to be able to trace it (your reasoning) all the way back to evidentially explain why you did each part... an analyst always has to justify what they have done, so does a system.*" [Study 1, A4, 35:00] We believe that Pan addresses these issues by providing algorithmic transparency of its reasoning, using an architecture that aids recognition and explanations that match our explanation framework. One use case for Pan is police intelligence analysis, where analysts are exploring large and complex knowledge graphs. In Study 3, we asked analysts to explore a fictional investigation scenario using Pan. While fictional, the scenario was based on a real investigation described by an analyst in Study 1. Fig. 8 presents the initial briefing information provided to analysts. This was given with a briefing document, describing the scenario and the types of questions that could be asked of Pan.

The scenario involved a text message from a mobile phone (IDMOB1) that had been sent to IDMOB2. The task was for the analyst to form a hypothesis about who could be the owner of the phone, IDMOB1. The researcher started the investigations by asking the question, "what mobiles have been involved in call events with IDMOB1?" There was no definite right answer to the investigation, however analysts were expected to provide evidence to support their chosen hypothesis. One individual in the data was a more likely suspect than any other, due to their connection to both the phone and the organised crime group (OCG). It was therefore expected that this person would at least be mentioned as a possible suspect, even if the analyst eventually opted for an alternative hypothesis.

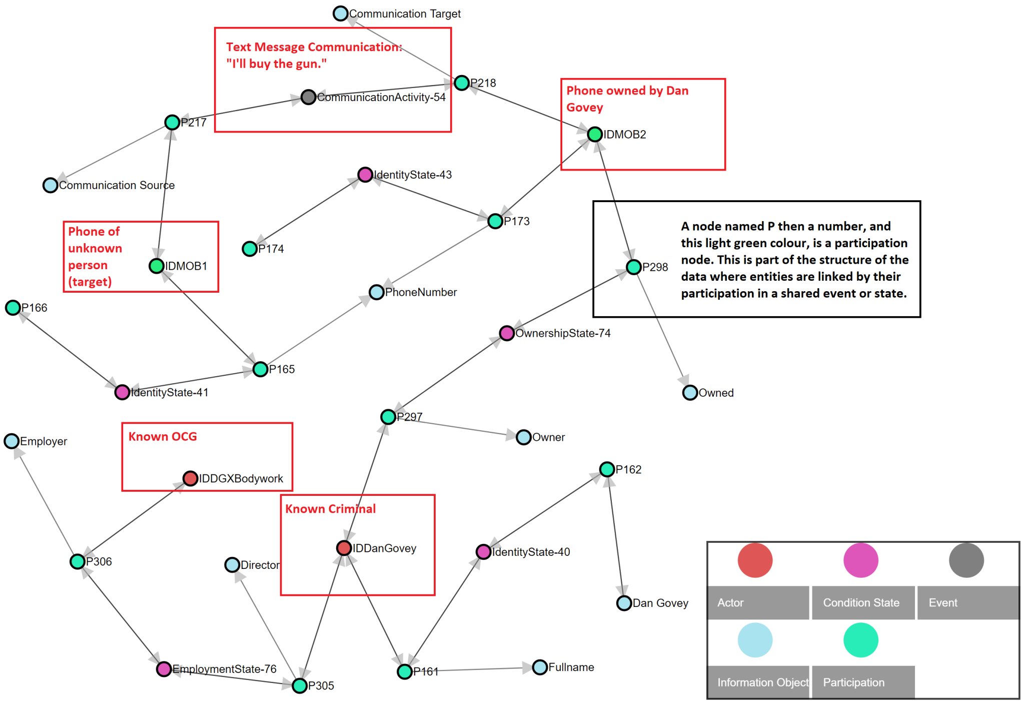


Fig. 8. Initial Briefing Scenario Information: Analysts were provided with this visualisation to familiarise themselves with the scenario, at least two weeks prior to the experiment

6.3 Results

Throughout the scenario, analysts made utterances which were transcribed and coded as to whether each concerned the system, the scenario, or an interaction with Pan i.e. a question. The types of questions asked by analysts included *"what people are associated to DGX Bodywork?"* [Study 3; A2; 7:30], *"what is the relationship between Susan Leech and Paul Richards?"* [Study 3; A3; 26:20], and *"are there any connections between IDSusanLeech and IDDarrenSmith?"* [Study 3; A8; 15:10]. The cues in a question were often linked to the results of a previous question. System and scenario related utterances were also coded to describe their nature i.e. if it was (i) a question about the system or scenario, (ii) understanding of the system or scenario, or (iii) misunderstanding about the system or scenario. There were a total of 227 utterances made by all analysts about the system or scenario, ranging from between 12 and 35 for each analyst. We compared utterances against the component themes in Table 8. All utterances about the scenario and system could be described by an aspect within the System Component Explanation Framework, with the exception of utterances about the initial briefing material. For example, to understand the data and source one analyst explained, *"I want to know what the source is of the connection between the number and Susan. Is it from her calling the police? Is it from our previous call data? Is it from us connecting her via previous investigations?"* [Study 3; A1; 10:05] By interacting with the graph visualisation the analyst could find the source of the data. We have therefore validated our findings from Study 2, that the framework describes explanation needs for the different components of a CA.

All 10 analysts, with or without transparency, were able to describe a hypothesis that was supported by evidence. They demonstrated explicit understanding of the scenario in their utterances.

For example, where one analyst explained their reasoning for why Paul Richards was a suspect, that *"IDMOB1 has been used once between MOB4 and once between MOB2 and 6 times with IDMOB3, that could indicate a closer relationship, given that Paul Richards has been identified as being involved in a domestic incident with Susan Leech. So that could indicate that they have, or have had, a domestic relationship."* [Study 3; A5; 35:00] Their understanding of the scenario and reasoning around hypotheses was driven by their interactions with Pan, specifically the data explanations provided via text responses and graph visualisations. Thus, Pan was helpful as a visual analytics tool to support their investigations.

Analysts made use of the interactive features of Pan by choosing to explore the visualisation, for example one analyst stated, *"I prefer visual, so I'd immediately visualise it."* [Study 3; Analyst 4; 7:35] Analysts also asked questions like, *"can you zoom in on this (the graph) so I can see the associations"* [Study 3; A2; 22:30], *"can I make that a bit clearer (rearrange the visualisation) so I can understand the flow of the data"* [Study 3; A6; 6:55], and, *"can I click on that and see what that data is, where that data has come from?"* [Study 3; A1; 11.05] A number of analysts explicitly stated that the visualisations were helpful, for example one analyst said they found *"being able to see that visualisation compared to just the information on the side there (the text response) is quite interesting."* [Study 3; A2; 19:25]. Another said that, *"I can start to see the links between things and how it is pulling it back, so that is quite helpful actually."* [Study 3; A6; 12:50] In addition, the graph visualisation approach was familiar and is commonly used by network analysis tools.

While it was identified that the structure of the underlying data was of interest to analysts for transparency reasons, presenting the exact structure had drawbacks for sensemaking. A number of analysts suggested that the participation nodes were not necessary to show, for example one analyst said *"now I have my eyes tuned in to ignore the P participation nodes it's actually not too bad reading it"* [Study 3; A4; 16:40], and another suggested that they could be *"articulated in the link between"* entity nodes [Study 3; A5; 11:40]. In a future iteration of the prototype a simple adjustment will be made to hide superfluous aspects of the data structure from the graph visualisation.

The focus for Study 3 was to evaluate our approach to provide transparency of system processes, notably the goals and constraints. Those who were provided transparency could view a triggered intention concept, with descriptions of functional attributes structured by the RPD model, and this helped them to recognise and understand system behaviour. In later stages of the exercise the related constraints helped frame their reasoning, and ultimately derive insight. For example, one analyst changed their investigation strategy due to their understanding of the system behaviour. At first, when they asked a question about any connections between two entities (Paul Richards and DGX Bodywork), they assumed that because *"it has not found a shortest path then there are no paths."* [Study 3; A4; 14:15]. They then viewed the intention concept and said *"but it is only looking at three."* [Study 3; A4; 14:25] This enabled them to adjust their interaction strategy, *"if I can't do it that way, I would probably then want to expand off Paul Richards just to get a bit more of an idea."* [Study 3; A4; 15:20] In another example, one analyst was reasoning about potential hypotheses and said *"...and Paul does have a shortest path connection to bodyworks"* [Study 3; A1; 21:30]. The constraints are present in their reasoning, because it is an important caveat that there may be other longer paths between the two. The analysts without transparency had no opportunity to understand the system, and therefore could not adjust their investigation strategies nor identify caveats.

After the interactive exercises were complete, analysts were asked a series of interview questions including "what might you need to know about the whole system?" The responses from analysts reiterated the areas described in our Transparency Framework (Fig. 3), those with transparency focused upon the need to inspect and verify the system goals and constraints, whilst those who were not provided with transparency described the need to explain underlying data (Table 9).

Table 9. Interview Responses: What do analysts want to know about a system?

Q	**Comments With System Transparency**	**Trans. Need**	**Comments Without System Transparency**	**Trans. Need**
What might you need to know about the whole system?	*"What is the actual chain of events that is happening in the background in order to give me an answer?" [A1; 33:00] "... to reverse engineer it… to see 'oh yeah' that's the query I would have written." [A1; 33.30] "what hasn't been picked up." [A4; 26:45] "what the limitations are behind it." [A6; 33:50] "… what does it sit on to work and what are the known limitations it bounces off." [A6; 33:55] "does it prioritise any information?" [A7; 29:45] "the rules on what it is pulling back. So it explains exactly what it has done" [A8; 26:00] "To make sure it is taking my question and then actually doing with it what I hoped it would do." [A8; 26:20] "I would read this up and say 'ok, what have I missed', if it is not pulling back what I think it should be pulling back." [A8; 26:40]*	Inspect and verify system goals and constraints	*"where the data we are being given is coming from… " [A2; 40:30] "what the entities are and the keys" [A3; 44:20] "Those linkages will be based upon a report, or there will be something behind those associations." [A5; 44:40] "what the system sits on. So whether that sits on PNC, RMS, so you understand the data you have access to." [A9; 39:45] "what are the categories where it would ping up as a result. So I presume it searches on names, locations, does it do finance? Mobile entities?" [A10; 30:55]*	Explain data

6.4 Discussion

During the investigations, analysts took many different paths and revisited previous questions fluidly. There is a need for flexibility in visual analytics systems, including the ability to evolve new capabilities. In feedback to date analysts have expressed the desire for a CA which can provide more advanced assistance than simply retrieving information from a single question. For example, where an analyst explained they would "*use it to search for thematic situations*" (Study 2; A3; Q8; C2), which would involve an understanding of thematic patterns of behaviour or activities in the data by the agent. Analysts also indicated that they could use an assistant which can provide recommendations for alternative questions, and to challenge the investigation scope. Recommendations and semi-autonomous processes performed by a CA must be designed with care in the context of intelligence analysis, given the risks of bias and deception and the need for clear explanation of lines of inquiry, including evidence, goals and caveats.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we describe how we have used FCA to form combinations of attributes that define the functional processes for a CA system, called Pan. The architecture is underpinned by CTA of expert users and is transparent by design. Our system is a CA, however the functions and features of other AI applications could also be designed in this way. This approach represents advancement in the state of the art. In Pan, we also provide a partial solution to the brittleness problem by allowing capabilities to evolve, harnessing FCA.

There are many areas of further research. We would like to conduct empirical analysis to evaluate the impact of a CA in investigation environments, with and without transparency, including on speed, accuracy and understanding of information retrieval. In doing so, we hope to capture more detail on the level of trust required in CA systems and considerations for designing transparency which aids critique rather than mere acceptance of responses. Springer and Whittaker find that in cases “transparency is distracting and undermines simple heuristics users form about system operation” [48]. We propose that, by designing transparency around the RPD model and our explanation framework, we match user recognition and can present clear and bespoke component explanations. We intend to test this further in trials with operational intelligence analysts. We will also look to refine the FCA, including the objects and attributes available, with input from analysts to consider questions with more complex datasets and requirements. Objects can be captured by further interviews and observations of analysts as they work with the CA system, while attributes can be developed with a greater understanding of possible graph traversal methods.

The current approach considers single information retrieval requests; however, information gathering in investigations typically comprises a series of related tasks where one piece of information leads to a follow up question or a new line of inquiry. For example, one analyst described a situation where they were trying to identify a person of interest and they could filter out information based on characteristics such as "*he lived in roughly the right area*" [CTA, A4, 13.45] or other "*layers of background understanding*" [CTA, A4, 15.55]. This background understanding develops in an iterative fashion as the investigation progresses and analysts use it to inform future questions. To provide transparency of the overall investigation it is insufficient to provide explanation and visibility of individual responses. Instead we need to present the user with understanding across the entire investigation pathway, including, for example, system constraints where they impact critical stages of the investigation. In future work, we will consider the design of an interface to provide analysts with system transparency across multiple CA responses and investigation pathways. We propose that our RPD design for concept attributes in individual interactions with a CA will also aid us in achieving aggregated transparency for multiple interactions.

As analysts interact with the CA and related intention concepts, we can capture interesting insights reflecting analyst thought processes when retrieving information. We propose these interactions could be used to help understand multi-stage actions performed by expert analysts, which can be used to train an agent to conduct information retrieval tasks autonomously. Our interviews with analysts have shown that, at the outset of an investigation, it is difficult to clearly define claims and hypotheses about the situation because of uncertainties and information gaps. Instead, an analyst’s understanding evolves iteratively over time as questions lead to further questions and a narrowing of the overall investigation scope, this allows the forming of inferences and reasoning. Analysts do not think in rigid structures, therefore information retrieval tasks are not tightly constrained or structured against a particular framework such as argumentation [50], for example. Rather the process is expected to be flexible. We propose that our approach to capture interactions between an analyst and a CA will provide a platform from which to develop flexible information retrieval aids. The role of investigation scope was prominent in CTA interviews with analysts, where the questions asked by analysts were framed by the initial scope, thus introducing the risk that important information beyond the scope is missed. We will consider how an autonomous agent could help mitigate constraints of investigation scope, including more complex machine reasoning to pursue alternative lines of inquiry. Analysts expressed the desire to avoid obvious follow up questions, so it would be helpful for a CA to predict what additional questions may arise and provide the information without the analyst needing to ask for it. One potential approach to develop CAs that can explore information autonomously is to use chain event graphs, where different FCA concepts are used to describe states in an investigation [18].

For autonomous systems, transparency is a critical issue and an RPD functional structure for state positions could help understanding by providing observable system states and transitions. Traditional machine learning approaches to learn and predict future events are based upon observations of real world data and are at serious risk of bias, for example through data poisoning or collection biases, and this has dangerous implications in policing as identified by Couchman [10]. We should therefore be wary of applying such approaches to develop investigations autonomously. Additionally, the insights that can arise from analyst intuition throughout the investigation process could be hampered by an autonomous approach, which hides the reasoning over decision points in an investigation pathway. Klein [25] presents a variety of ways in which insights can arise, from a flash of illumination, to making connections, finding coincidences, contradictions and through creative desperation. Fundamentally, it is difficult to capture exactly what information will lead to insight, however, by hiding key reasoning at decision points in an investigation we remove the opportunity for it to occur. It is, therefore, important that we provide information and explanations across the reasoning pathway of an autonomous agent to enable insight and transparency. Our work has various implications for the development of autonomous agents that we intend to explore. RPD intention modelling provides an effective architecture to move from a reflex based agent that responds to user queries, to a goal based agent. As analysts interact with the CA, we can capture their investigation processes including the concepts triggered and how they move from one intention to another [18]. With this knowledge, an autonomous agent can learn to flexibly move between concepts to reach a predicted goal. Traditional approaches to model autonomous agent behaviours can involve careful and brittle configuration of rule based procedures, for example to test some hypotheses. We propose that in the early stages of an investigation, when much of the information required for understanding a situation is missing, analysts cannot clearly articulate a set of hypotheses and first need to explore various routes to develop situational awareness. By allowing an agent to learn, evolve capabilities, and predict domain agnostic investigation pathways in a less structured manner, our architecture helps avoid brittleness and is more in tune with analyst thought processes. We believe our architecture, underpinned by the RPD model at each state, will allow for transparency of agent processes at different levels of inspection. By learning how analysts move between intention concepts, rather than from patterns in real world data, we can help mitigate algorithmic bias and brittleness (we do not need large training data for a particular domain) in predicting lines of inquiry.

In intelligence analysis, it is important that the CA can reply objectively, without introducing bias or ambiguity. There has been research into developing CA's that can respond with emotive language [58]. In a similar way, some narrative style responses are more persuasive than others [38]. Research by Schuetzler et. al. even indicates that people respond differently to a CA which has more human-like conversational skill [45]. For example, it is suggested that the human users show strategic behaviour to conceal their deception when conversing with a more human-like CA (in the study this was to lie when describing an image to the CA). It is unclear what effect the conversational skill would have on a CA for investigations, or whether a more skilful CA could cause analysts to interact with it differently. However, we should look to design conversations carefully. It would be damaging if a CA were able to influence the course of an investigation in a way that is not supported by evidence, for example by changing analyst behaviours or questioning strategies. Perhaps with greater transparency analysts will be better able to identify bias and challenge a narrow investigation scope more effectively. The impact of bias, however, will likely be subtle and stretch over multiple interactions. We will therefore need to consider how to explain the entirety of the conversation and selected investigation path in the context of alternatives.

Our interviews indicated that analysts did not want humanness in the response from a CA for intelligence analysis. For example, one analyst explicitly stated that they wanted to know when they were talking to a human and when to a machine. Instead, logical and clear responses were preferred. This finding is at odds with traditional work on dialogue for CA's, for example where See et al. [46] uses 'Humanness' and 'Engagingness' as measures of overall quality of a conversation with a CA. Perhaps when intelligence analysts are communicating with a conversational agent we should apply a different set of evaluation metrics where understanding and bias mitigation are more important than humanness. These metrics could mirror areas identified through our ETA. We have only interviewed a small number of analysts thus far, so future work will look further into approaches to conversation and what metrics should be used for evaluating CA systems.

ACKNOWLEDGMENTS

This research was assisted by experienced military and criminal intelligence analysts who work for the UK Government.

REFERENCES

[1] [n.d.]. DeepPavlov: An open source conversational AI framework. http://deeppavlov.ai/

[2] [n.d.]. Language Understanding (LUIS). https://www.luis.ai/home

[3] [n.d.]. Learn AI-designing and architecting intelligent agents. https://azure.github.io/LearnAI-DesigningandArchitectingIntelligentAgents/

[4] Simon Andrews, Babak Akhgar, Simeon Yates, Alex Stedmon, and Laurence Hirsch. 2014. Using Formal Concept Analysis to detect and monitor organised crime. *Lecture Notes in Computer Science* 8132. https://doi.org/10.1007/978-3-642-40769-7_11

[5] W. Ross Ashby. 1991. Requisite variety and its implications for the control of complex systems. In *Facets of Systems Science*. Springer US, Boston, MA, 405–417. https://doi.org/10.1007/978-1-4899-0718-9_28

[6] Rajeev Bhattacharya, Timothy M. Devinney, and Madan M. Pillutla. 1998. A formal model of trust based on outcomes. *The Academy of Management Review* 23, 3 (1998), 459–472. http://www.jstor.org/stable/259289

[7] Ann Blandford and B. L. William Wong. 2004. Situation awareness in emergency medical dispatch. *Int. J. Hum.-Comput. Stud.* 61, 4, 421–452. https://doi.org/10.1016/j.ijhcs.2003.12.012

[8] Stuart K. Card, Allen Newell, and Thomas P. Moran. 1983. *The psychology of Human-Computer Interaction.* L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

[9] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 670–680. https://www.aclweb.org/anthology/D17-1070

[10] Hannah Couchman. 2019. Policing by machine: Predictive policing and the threat to our rights. *Liberty* (2019).

[11] Neta Ezer, Sylvain Bruni, Yang Cai, Sam Hepenstal, Chris Miller, and Dylan Schmorrow. 2019. Trust engineering for human-AI teams. *Human Factors and Ergonomics Society Annual Meeting Proceedings.*

[12] Gemma C. Garriga. 2017. *Formal Concept Analysis.* Springer US, Boston, MA, 522–523. https://doi.org/10.1007/978-1-4899-7687-1_316

[13] Matylda Gerber, B. L. William Wong, and Neesha Kodagoda. 2016. How Analysts Think: Intuition, leap of faith and insight. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60 (09 2016), 173–177. https://doi.org/10.1177/1541931213601039

[14] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2018. Explaining Explanations: An approach to evaluating interpretability of Machine Learning. arXiv:1806.00069 http://arxiv.org/abs/1806.00069

[15] Sam Hepenstal, Neesha Kodagoda, Leishi Zhang, Pragya Paudyal, and B. L. William Wong. 2019. Algorithmic transparency of conversational agents. In *Proceedings of the Workshop on Algorithmic Transparency in Emerging Technologies co-located with 24th International Conference on Intelligent User Interfaces (IUI 2019), Los Angeles, March 20, 2019.*

[16] Sam Hepenstal, B. L. William Wong, Leishi Zhang, and Neesha Kodagoda. 2019. How analysts think: A preliminary study of human needs and demands for AI- based conversational agents. In *Proceedings of the 63rd Human Factors and Ergonomics Society Annual Meeting.*

[17] Sam Hepenstal, Leishi Zhang, Neesha Kodagoda, and B. L. William Wong. 2020. Pan: Conversational agent for criminal investigations. In *IUI '20: 25th International Conference on Intelligent User Interfaces, Cagliari, Italy, March 17-20, 2020, Companion.* ACM, 134–135. https://doi.org/10.1145/3379336.3381463

[18] Sam Hepenstal, Leishi Zhang, Neesha Kodagoda, and B. L. William Wong. 2020. Providing a foundation for interpretable autonomous agents through elicitation and modelling of criminal investigation pathways. In *Proceedings of the 64th Human Factors and Ergonomics Society Annual Meeting.*

[19] Sam Hepenstal, Leishi Zhang, Neesha Kodagoda, and B. L. William Wong. 2020. What are you Thinking? Explaining conversation agent responses for criminal investigations. In *Proceedings of the Workshop on Explainable Smart Systems for Algorithmic Transparency in Emerging Technologies co-located with 25th International Conference on Intelligent User Interfaces (IUI 2020), Cagliari, Italy, March 17, 2020 (CEUR Workshop Proceedings)*, Alison Smith-Renner, Styliani Kleanthous, Brian Lim, Tsvi Kuflik, Simone Stumpf, Jahna Otterbacher, Advait Sarkar, Casey Dugan, and Avital Shulner Tal (Eds.), Vol. 2582. CEUR-WS.org. http://ceur-ws.org/Vol-2582/paper3.pdf

[20] Robert R. Hoffman, Gary Klein, and Shane T. Mueller. 2018. Explaining explanation For "Explainable Ai". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 1 (2018), 197–201. https://doi.org/10.1177/1541931218621047 arXiv:https://doi.org/10.1177/1541931218621047

[21] Suraiya Jabin. 2015. Machine Learning methods and applications using Formal Concept Analysis. *International Journal of New Technologies in Science and Engineering* 2, 3 (2015).

[22] Bret Kinsella. 2018. Amazon Echo device sales break new records, Alexa tops free app downloads for iOS and Android, and Alexa down in Europe on Christmas morning. https://voicebot.ai/2018/12/26/amazon-echo-device-sales-break-new-records-alexa-tops-free-app-downloads-for-ios-and-android-and-alexa-down-in-europe-on-christmas-morning/

[23] Bret Kinsella. 2019. NPR study says 118 million smart speakers owned by U.S. adults. https://voicebot.ai/2019/01/07/npr-study-says-118-million-smart-speakers-owned-by-u-s-adults/

[24] Gary Klein. 1993. A Recognition Primed Decision (RPD) model of rapid decision making. *Decision making in action: models and methods* (01 1993).

[25] Gary Klein. 2017. *Seeing what others don't.* Nicholas Brearley Publishing.

[26] Gary Klein, Roberta Calderwood, and Donald MacGregor. 1989. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 3 (May 1989), 462–472. https://doi.org/10.1109/21.31053

[27] Gary Klein, Brian Moon, and Robert Hoffman. 2006. Making sense of sensemaking 2: A macrocognitive model. *Intelligent Systems, IEEE* 21 (10 2006), 88 – 92. https://doi.org/10.1109/MIS.2006.100

[28] Neesha Kodagoda, B. L. William Wong, and Nawaz Khan. 2009. Cognitive Task Analysis of low and high literacy users: Experiences in using grounded theory and Emergent Themes Analysis. *Human Factors and Ergonomics Society Annual Meeting Proceedings* 53 (10 2009), 319–323. https://doi.org/10.1518/107118109X12524441080821

[29] Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. 2006. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel evaluation methods for information visualization* (Venice, Italy) *(BELIV '06).* ACM, New York, NY, USA, 1–5. https://doi.org/10.1145/1168149.1168168

[30] David Leslie. 2019. Understanding artificial intelligence ethics and safety. *CoRR* abs/1906.05684 (2019). arXiv:1906.05684 http://arxiv.org/abs/1906.05684

[31] Georgios Leventakis and M. R. Haberfeld. 2018. *Societal implications of community-oriented policing and technology.* Springer. https://doi.org/10.1007/978-3-319-89297-9

[32] Zachary Chase Lipton. 2016. The mythos of model interpretability. *CoRR* abs/1606.03490 (2016). arXiv:1606.03490 http://arxiv.org/abs/1606.03490

[33] Bernard Marr. 2014. Dear IKEA: Your customer service is terrible. *LinkedIn* (2014). www.linkedin.com/pulse/20140325060328-64875646-dear-ikea-your-customer-service-is-terrible

[34] Michael F. McTear. 2002. Spoken Dialogue Technology: Enabling the conversational user interface. *ACM Comput. Surv.* 34, 1 (March 2002), 90–169. https://doi.org/10.1145/505282.505285

[35] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. 2018. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018).*

[36] Christoph Molnar. 2019. *Interpretable Machine Learning. A guide for making black box models explainable.* Lulu, 1st edition, March 24, 2019; eBook (GitHub, 2020-04-27).

[37] Donald Norman. 1983. Design rules based on analyses of human error. *Commun. ACM* 26 (04 1983), 254–258. https://doi.org/10.1145/2163.358092

[38] Nancy Pennington and Reid Hastie. 1992. Explaining the Evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology* 62 (02 1992), 189–206. https://doi.org/10.1037/0022-3514.62.2.189

[39] Alun Preece, William Webberley, David Braines, Erin G. Zaroukian, and Jonathan Z. Bakdash. 2017. Sherlock: Experimental evaluation of a conversational agent for mobile information tasks. *IEEE Transactions on Human-Machine Systems* 47, 6 (Dec 2017), 1017–1028. https://doi.org/10.1109/THMS.2017.2700625

[40] Eric Prud'hommeaux and Andy Seaborne. 2007. SPARQL query language for RDF. https://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/

[41] Nadeem Qazi, B. L. William Wong, Neesha Kodagoda, and Rick Adderley. 2016. Associative search through Formal Concept Analysis in criminal intelligence analysis. In *IEEE International Conference on Systems, Man, and Cybernetics*. 001917–001922. https://doi.org/10.1109/SMC.2016.7844519

[42] Nicole M. Radziwill and Morgan C. Benton. 2017. Evaluating quality of chatbots and intelligent conversational agents. *CoRR* abs/1704.04579 (2017). arXiv:1704.04579 http://arxiv.org/abs/1704.04579

[43] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum C. Kwon, Geoffrey Ellis, and Daniel A. Keim. 2014. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1604–1613.

[44] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A. Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors* 58, 3 (2016), 377–400. https://doi.org/10.1177/0018720816634228 arXiv:https://doi.org/10.1177/0018720816634228 PMID: 27005902.

[45] Ryan Schuetzler, Mark Grimes, and Justin Giboney. 2019. The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior* 97 (2019), 250–259.

[46] Abigail See, Stephen Roller, Douwe Kiela, and Jason Weston. 2019. What makes a good conversation? How controllable attributes affect human judgments. *CoRR* abs/1902.08654 (2019). arXiv:1902.08654 http://arxiv.org/abs/1902.08654

[47] Danny Shaw. 2019. Crime solving rates 'woefully low', Met Police Commissioner says. *BBC* (2019). https://www.bbc.co.uk/news/uk-48780585

[48] Aaron Springer and Steve Whittaker. 2019. Progressive Disclosure: Empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. ACM, New York, NY, USA, 107–120. https://doi.org/10.1145/3301275.3302322

[49] James J. Thomas and Kristin A. Cook eds. 2005. Illuminating the Path: The research and development agenda for visual analytics. *IEEE CS Press* (2005).

[50] Stephen E. Toulmin. 1958. *The uses of argument.* Cambridge University Press.

[51] Jane Wakefield. 2016. Would you want to talk to a machine? *BBC* (2016). https://www.bbc.co.uk/news/technology-36225980

[52] Christine T. Wolf. 2019. Explainability Scenarios: Towards scenario-based XAI design. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. ACM, New York, NY, USA, 252–257. https://doi.org/10.1145/3301275.3302317

[53] B. L. William Wong. 2003. *Critical Decision Method data analysis.* Lawrence Erlbaum Associates, 327–346.

[54] B. L. William Wong and Ann Blandford. 2002. Analysing ambulance dispatcher decision making: Trialing Emergent Themes Analysis. *In: Proceedings of the HF2002 human factors conference design for the whole person - integrating physical, cognitive and social aspects: a joint conference of the Ergonomics society of Australia (ESA) and the Computer human interaction special interest group.* (11 2002).

[55] B. L. William Wong and Neesha Kodagoda. 2016. How Analysts Think: Anchoring, laddering and associations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60 (09 2016), 178–182. https://doi.org/10.1177/1541931213601040

[56] Serhiy A. Yevtushenko. 2000. System of data analysis "Concept Explorer". (In Russian). *Proceedings of the 7th national conference on Artificial Intelligence KII-2000*, 127–134.

[57] Xiaoyu Yin, Dagmar Gromann, and Sebastian Rudolph. 2019. Neural machine translating from natural language to SPARQL. arXiv:1906.09302 http://arxiv.org/abs/1906.09302

[58] Michelle X. Zhou. 2019. Getting Virtually Personal: Making responsible and empathetic "Her" for everyone. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. ACM, New York, NY, USA, i–i. https://doi.org/10.1145/3301275.3308445