Transparent conversational agent systems for intelligence analysis

Sam Hepenstal
Defence Science Technology Laboratory (Dstl)
UK

shepenstal@dstl.gov.uk

ABSTRACT

Big data presents a problem to intelligence analysts who need to extract insights from the data and advise decision makers. Traditional methods to explore and filter large datasets are complex and require manual configuration of search parameters or query syntax. A more natural approach to retrieve interesting data could be provided via a conversational agent (CA), where analysts can interact with available information through natural language. This can speed up processing of big datasets; however there are critical flaws in existing CA systems which must be resolved before they can provide good situation awareness. In this paper we discuss two such flaws, system transparency and brittleness. We present our work to date, including the development of an 'algorithmic transparency framework', published at IUI 2019. This framework identifies the need for analysts to be able to inspect and verify system goals and constraints, in addition to explaining specific outputs. We also discuss cognitive task analysis interviews with analysts which underpin the framework, published at HFES 2019. Our research has led to the development of a prototype transparent CA system for intelligence analysis. In this paper we define the next steps, including experimenting and improving the prototype with operational intelligence analysts.

1.0 PROBLEM OUTLINE

Intelligent systems for performing intelligence analysis provide a high-level capability which is a combination of multiple cognitive computing capabilities. Braines et. al. define this high-level capability as coalition situational understanding (CSU). Specific cognitive computing capabilities include "human-computer collaboration, knowledge representation and reasoning, multi-agent systems, machine learning, natural language processing, and vision/speech processing." (Braines et. al. 2019) One example of a system for CSU is a conversational agent (CA) which can perform intelligence analysis. A CA typically comprises various complex processes, including natural language processing of a user's question and intention, identification of an appropriate action which triggers additional data processing, and formulation of responses. Each of these aspects can be trained and can learn through interactions with an analyst.

Artificial intelligence (AI) based CA technologies are still in their infancy. However, their popularity is increasing (Kinsella 2018, Kinsella 2019) because they provide a more intuitive, natural, and quicker way to access information. Typical uses of CA technologies involve low risk and low consequence scenarios. For example, when someone uses Amazon Alexa to select and play music tracks. It does not matter to the user exactly how Alexa has interpreted their question and reasoned what track is most appropriate, so long as a satisfactory track is played. If Alexa gets the selection wrong, the user will easily recognise the error and can ask a different question. So long as Alexa is not repeatedly incorrect, the user does not require a detailed understanding of the underlying algorithms which process the data. While the potential benefits of using CAs in high risk and high consequence decision making environments are large, there are critical issues of *transparency* and *brittleness* which must first be addressed.

In this report we discuss our work to date, including the development of an 'Algorithmic Transparency Framework' which was presented in a paper at IUI 2019. In our framework we define *transparency* as the ease with which a user can (i) explain any results provided by a system, in addition to (ii) being able to inspect and verify the goals and constraints of the system within context (Hepenstal et al. 2019). *Brittleness* (iii) occurs when the technology fails to cope with the variety of demands that it has to cope with when in use. We look to address these three important aspects of CA design. Here we outline the problem faced regarding CA technologies and their use for intelligence analysis. We propose that algorithmic transparency is a crucial requirement of CA systems and that this, underpinned by research into areas such as human cognition and sensemaking, should run through their design. Our framework has been informed and developed by interacting with operational intelligence analysts, including through Cognitive Task Analysis (CTA) interviews, for which we are presenting a paper at HFES 2019. Other intelligence analysis related aspects of using algorithmic systems, such as trust, have also been explored, and we are part of a panel paper on this topic at HFES 2019. Our research has led to the early development of a prototype CA system. In this report, we also discuss the next steps for our research, including experimentation with operational intelligence analysts. Our aim for this research is to understand how algorithmic transparency can be designed into advanced CA systems, and to demonstrate this in an evaluation of a prototype for intelligence analysis.

2.0 INTRODUCTION

CA 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 to direct them to the right information or department. This saves companies money and can save customers time waiting in a queue. Typical applications for conversational agents tackle concise user tasks for mundane processes which can be translated to a finite set of user intentions. Here 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. Traditional CAs have, therefore, 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*." The nature of CA technologies and the ability to interact in a natural human way can actually amplify misunderstandings. The tasks which traditional CAs can perform are constrained to a predefined set of 'intentions', which can lead to frustration where users expect the CA to behave like a human.

Typically, when a user asks a question the CA attempts to match their input to an intention. For example, if we develop a CA with Artificial Intelligence Markup Language (AIML), a commonly used approach to develop chat interactions which supports many chatbots platforms and services (Radziwill and Benton 2017). A pattern is described for a task category (intention) together with a template response if the users' text matches with the pattern. There is a need to define or learn the intentions (and subsequent actions) which the CA can fulfil, and to ensure that they are consistent and distinct so as not to confuse either the pattern matching algorithm or the user. The CA can only respond in accordance with the intention, additionally the intention will consist of processes with their own limitations. The possible intentions and the various constraints are not visible in commercial CA systems, so are not transparent. The predefined nature of intentions and related tasks also creates the critical issue of brittleness in CA systems.

CAs can significantly speed up repetitive information retrieval tasks, when compared to traditional query methods or manual processes, and are therefore also attractive for a wide range of applications which need fast responses in critical situations. The domain of criminal intelligence analysis is one example of an environment requiring critical decision making which could benefit from CA technologies. The volume of data which requires processing by police today is a significant barrier to bringing criminals to justice. 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". (https://www.bbc.co.uk/news/uk-48780585) This is a key area in which an intelligent CA which can interpret an analysts 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 a major flaw in commercial conversational agents, where due to "black box" approaches it is difficult to understand how they have interpreted the user, explored the data, and built a response. Clearly, intelligence analysts need a much greater understanding of the underlying algorithmic processes of a CA than someone selecting music. During analysis the information retrieval tasks are complex and different methods can be applied to achieve the same aim, but depending upon the methods involved and their interpretation an investigation can be guided towards different paths with consequences. Trust in methods to gather information is important for intelligence analysts, who may be required to go to court to explain their analysis and advice. The ethical implication of using artificial intelligence applications is also a key consideration for CAs. In Leslie's (2019) 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. We look to bridge the gap between theoretical transparency requirements and guidelines and the design of actual operational systems.

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 (Ezer et. al. 2019). 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 where it is damaging if it leads to analytical complacently. If an analyst trusts the results of an algorithm without interpreting the reasoning then their trust could lead to adverse decisions, due to unrecognised bias or 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. Without trust, underpinned by transparency, AI systems will not be used by analysts. One analyst described succinctly that "*an analyst always has to justify (in court) what they have done, and so should the system.*" (Hepenstal et. al. 2019) There is therefore a critical requirement for providing algorithmic transparency in applications for intelligence analysis.

Intelligence analysis requires a fluid system which can adapt to new lines of inquiry and traditional approaches to CA technologies encounter issues of 'brittleness'. Brittleness occurs when the technology fails to cope with the variety of demands that it has to cope with when in use. Developing a flexible system which can evolve to meet new analyst intentions could help counter issues of brittleness, however such a system also needs to balance algorithmic transparency.

We believe that system transparency is a crucial factor in analyst adoption of AI technologies and propose that by delivering effective transparency of AI reasoning, with a close consideration of human needs, we can design systems which can be used in high risk and high consequence environments. Whilst there is existing work in the areas of human sensemaking, trust and ethics, there is a significant research gap where it comes to translating how these principles can be tangibly captured in the design of algorithms and systems. This research will consider and experiment with specific design requirements and approaches for providing algorithmic transparency for flexible intelligence analysis applications.

3.0 OUR RESEARCH TO DATE

An important focus of our research is to understand what the requirements are for algorithmic transparency of intelligent applications for intelligence analysis. We have considered 'intelligent applications' as those which allow for shared human-machine processes when retrieving and analysing information, for example where we use a CA to query a semantic knowledge graph. An analyst will be reasoning about the situation they are faced with and this directs the questions they ask, the CA then interprets their questions and decides upon an appropriate set of processes to provide responses. These processes could include graph traversal methods and clustering algorithms which use understanding and inferencing from the semantic structure of the underlying knowledge graph data. For example, if an analyst asked a question about "*what vehicles are linked to known offenders in the database?*" The agent could interpret from the question that the analyst is looking for connections between instances of vehicles and instances of people who have had a role as an offender. Whilst it is quicker and easier for an analyst to state questions in natural language, their interpretation is subjective and thus, raises the importance of transparency for CA responses. In the example given, what do we mean by 'linked'? Different people will have different interpretations and will therefore address answering the question with a variety of approaches, each with its own caveats and purpose. This is also true of a CA and the respective goals and constraints of any process must be clearly revealed so that an analyst can inspect and verify the system.

To date much research in the area of eXplainable Artificial Intelligence (XAI) has focussed upon the underlying model features of a specific classification result, for example explaining the relative importance of different features in an image to its classification (i.e. deep Taylor decomposition). Gilpin et al. (2018), gives the definition that XAI is a combination of interpretability and completeness, where interpretability is linked to explaining the internals of a system and completeness is to describe it as accurately as possible. An explanation of the result of an algorithm, for example the features which are important in providing an image classification, are important to provide understanding to an analyst for why a classification was given. However, this is not the complete picture when we consider the need for analysts to be able to inspect and verify a system of multiple processes. We propose that, in the field of intelligence analysis, the explanation for a result is important, but so is the approach taken to reach it including the related goals and the constraints.

3.1 Algorithmic Transparency Framework: Published IUI ATEC March 2019

We have developed an initial research framework for providing algorithmic transparency, informed by former analysts, which highlights areas that we believe need consideration when designing CA systems.

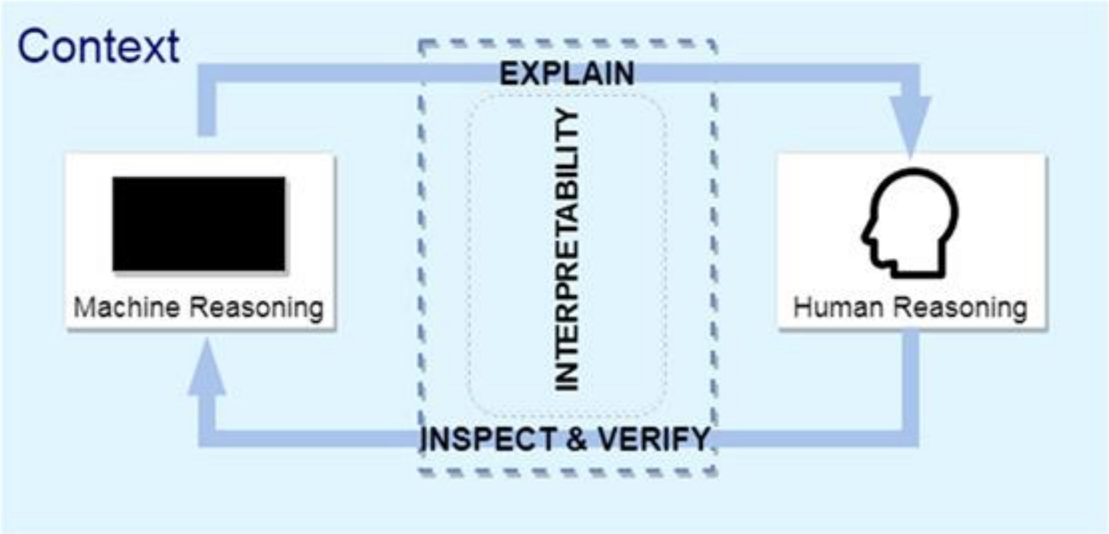

Figure 1: Algorithmic Transparency Framework (Hepenstal et. al. 2019)

Algorithmic Transparency Framework: What the user needs from black box algorithms: (i) explanations of how results from algorithms are arrived at (ii) explanations that are interpretable by the user in a manner that makes sense to them (e.g. the internals of the algorithm, including important features, an indication of accuracy or confidence, and an understanding of the data used and uncertainties, all presented in a manner which enables the user to assess if the results are sensible), (iii) visibility of the functional relationships mapped against the goals and constraints of the system, and (iv) context in which to interpret the explanations. NB: by showing goals and constraints, we include some key elements of *context, e.g. goals include some notion of the priorities and therefore some understanding of the problem, hence the context.*

The Algorithmic Transparency Framework shows how transparency relates to all aspects of the application beyond the user interface. This includes modelling the data, for example the classes and properties of the knowledge graph, and processes involved in machine reasoning such as graph traversal algorithms, so they are open to inspection and verification. Transparency cannot simply be provided as a post facto explanation or visualisation of results. Context is an important aspect of the framework and to design appropriate transparency for intelligence analysis applications we must first gain a clear understanding of the context, including specific goals and needs of analysts in different situations.

3.2 Analyst CTA Interviews: Awaiting Publication HFES October 2019

We have developed an initial research framework for providing algorithmic transparency, informed by former analysts, which highlights areas that we believe need consideration when designing CA systems.

To gain an understanding of analyst priorities, needs, and context, and to verify our framework, we have conducted four in-depth interviews with experienced intelligence analysts applying Cognitive Task Analysis (CTA). Three analysts have worked in policing, across various police forces, and the other analyst has a background in defence intelligence. Each analyst has over 3 years of experience working on a range of operational investigations, with a focus upon major crime. We chose CTA because our aim is to understand intentions which underpin how an analyst thinks and reasons. Each interview lasted an hour and applied the Critical Decision Method (CDM) (Klein 1989; Wong 2003) to elicit analyst expertise, cues, goals and decision making on a memorable investigation they were involved with from start to end. 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 investigations, specifically, their cues, goals, expectations and actions. These stages are typically time-pressured and are therefore prime situations in which CAs could assist analysts. 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.

Through our interviews we learned key insights into analyst thought processes and tasking which have helped us to consider how CAs can be designed transparently. One interesting finding relates to the scope of an investigation where this is crucial to direct intelligence analysis and enable situation recognition. Scope creates a basis from which expectancies can be drawn. In all interviews it was a key priority for analysts to identify a scope so that investigations could advance. Narrowing scope, however, inevitably leads to constrained investigations and potential bias. AI systems could provide significant benefit if they are able to expand the scope without causing extra burden upon analysts, for example by autonomously checking alternative hypotheses throughout an investigation. The investigation process described in interviews involved repetitive searching for information, where once a search returned a result many additional searches would be triggered. This is a process which could be eased significantly through the introduction of a CA.

Our interviews confirmed the need for an AI system to be open to inspection and verification. We also validated the recognition-primed decision (RPD) model (Klein 1993) as a way to understand how analysts recognise situations in an investigation and then respond. Klein describes “four important aspects of situation assessment (a) understanding the types of goals that can be reasonably accomplished in the situation, (b) increasing the salience of cues that are important within the context of the situation, (c) forming expectations which can serve as a check on the accuracy of the situation assessment (i.e., if the expectancies are violated, it suggests that the situation has been misunderstood), and (d) identifying the typical actions to take.” (Klein 1993) We propose that the factors which feed into the RPD model to aid analyst recognition could also aid a CA’s recognition of a situation. Furthermore, if the CA acts upon components which are designed around the RPD model we propose that this will significantly aid transparency because the CA can present the information it has recognised, and associated goals and constraints, in a way which mirrors analyst recognition. Table 1 presents a generic approach to answer analyst questions using the RPD model, which could help define transparent CA intentions.

Table 1: Consolidated Decision Analysis Table (Hepenstal et. al. 2019)

Hypothesis Scope	**Assess 5WH**
CUES	Inputs for 5WH (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

4.0 OUR RESEARCH NEXT STEPS

We propose that algorithmic transparency, underpinned by human factors research, must run throughout the CA system design and cannot be added as an afterthought. We seek to capture design requirements, applying and developing our algorithmic transparency framework, in order to demonstrate the impact of transparency provision upon analyst trust and use of the system. We intend to capture data from users to identify needs for the specific context of a CA for intelligence analysis. We will perform additional scenario based interviews with operational analysts, with a focus on answering the question; what do analysts need to understand about the output, the processes, and the system to have confidence in using a CA? Initially we focus upon the transparency needs for CAs to support information retrieval tasks in investigations. However, we also wish to investigate CAs which can perform additional capabilities, such as prompting and recommending courses of action, or performing elements of investigation autonomously.

There are three specific areas for research which are aligned to important technical aspects of CAs. Firstly, a CA must be able to interpret a user's intentions from their speech or text. Work to date has considered how CAs can dynamically model, learn and recognise analyst intentions, and provide transparency for their thinking and responses (Hepenstal et al. 2019). However, this is yet to be demonstrated and tested in a working prototype. We will seek to achieve this. Secondly, a CA needs to perform actions once it understands an analyst's intended goals, for example, to trigger an algorithm to retrieve data and follow investigative lines of inquiry. Our proposal is to assess what these possibilities are and how they can be captured, explained, and evolved when an analyst interacts with the CA. Finally, a CA needs to respond to the analyst. In intelligence analysis it is important that responses mitigate bias and provide clear understanding.

4.1 CA Prototype Application

We are first developing a CA system for information retrieval which is underpinned by human factors principles, specifically in the first instance the RPD model. We have applied a novel approach to design possible intentions for the CA which can evolve through interactions with an analyst. In our prototype the tasks which the CA performs are defined by a collection of functional modules, each linked to a specific component of the RPD, which are combined within distinct intention concepts. This approach allows us to effectively design capabilities which can evolve through analyst interaction, and can be inspected according to human factors principles for situation recognition. We have translated human factors research principles into functional components of an application. By developing our CA with the RPD principles at the core of intention design we propose that we can provide transparency so that analysts will be confident to use the application and to trust its responses.

For the next phase of research we intend to run experiments with the CA prototype application and operational intelligence analysts, as they use the system to complete an analysis task requiring repeated information retrieval steps. This will allow a variety of approaches to transparency, including different methods for output explanation and for visibility to inspect and verify system processes, to be tested and compared qualitatively. Of particular interest is the impact of transparency provision upon an analysts trust in a system and the subsequent decisions which are made, including whether the analyst feels comfortable to accept the reasoning and response of the CA and if they are aware of any caveats to the information provided. We plan to experiment with different algorithmic features, or action modules within the RPD structure, for example a deep neural network method to find similarities between instances in the data and a more transparent approach to do the same.

The design of our CA allows the possible tasks it can perform to evolve over time. This is a particularly critical issue for trust and transparency, where an analyst needs to be cognisant of the fact that the CA is learning and may respond differently in the future, even if this may be contrary to the analyst's expectations. While we have chosen the RPD model in the first instance because it suits the recognition nature of the information retrieval tasks described in interviews, we believe that our CA design could make use of alternative models for human cognition. An interesting model we are planning to experiment with is argumentation (Toulmin 1958), where you may task a CA with checking your arguments and looking to find refuting or supporting information. Additionally, we plan to experiment with more advanced CAs, such as those that can conduct investigations with greater autonomy, for example to predict paths of inquiry. Such applications must be designed 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, or deception, 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.

5.0 REFERENCES

[1] Braines, D., Tomsett, R., Preece, A., (2019). Supporting user fusion of AI services through conversational applications

[2] Ezer, N., Bruni, S., Cai, Y., Hepenstal, S., Miller, C., & Schmorrow, D. (2019). Trust engineering for human-AI teams. Paper presented at HFES 2019,

[3] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An approach to evaluating interpretability of machine learning

[4] Hepenstal, S., Kodagoda, N., Zhang, L., Paudyal, P., & Wong, B. L. W. (2019). Algorithmic transparency of conversational agents. Paper at IUI ATEC,

[5] Hepenstal, S., Wong, B. L. W., Zhang, L., & Kodagoda, N. (2019). How analysts think: A preliminary study of human needs and demands for AI- based conversational agents. Paper at HFES 2019,

[6] Kinsella, B. (2018, 26 Dec). 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/

[7] Kinsella, B. (2019, 7 Jan). 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/

[8] Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge.

[9] Klein, G. (1993). A recognition primed decision (RPD) model of rapid decision making.

[10] Leslie, D. (2019). Understanding artificial intelligence ethics and safety.

[11] Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents.

[12] Toulmin, S. E. (1958). The uses of argument, Cambridge University Press.

[13] Wong, B. L. W. (2003). Critical decision method data analysis, Lawrence Erlbaum Associates.