The Impact of System Transparency on Analytical Reasoning

Sam Hepenstal
Defence Science and Technology Laboratory
UK

Leishi Zhang
Canterbury Christ Church University
UK

B. L. William Wong
Auckland University of Technology
New Zealand

ABSTRACT

In this paper, we present the hypothesis that system transparency is critical for tasks that involve expert sensemaking. Artificial Intelligence (AI) systems can aid criminal intelligence analysts, however, they are typically opaque, obscuring the underlying processes that inform outputs, and this has implications for sensemaking. We report on an initial study with 10 intelligence analysts who performed a realistic investigation exercise using the Pan natural language system [10, 11], in which only half were provided with system transparency. Differences between conditions are analysed and the results demonstrate that transparency improved the ability of analysts to reason about the data and form hypotheses.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**; **User studies**.

KEYWORDS

Artificial Intelligence, Decision Support, Intelligence Analysis, Expert Decision Making

ACM Reference Format:
Sam Hepenstal, Leishi Zhang, and B. L. William Wong. 2023. The Impact of System Transparency on Analytical Reasoning. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3544549.3585786

1 INTRODUCTION

This paper describes an initial study to test the hypothesis that transparency is critical for tasks that require expert sensemaking. The study involved expert intelligence analysts who interacted with an Artificial Intelligence (AI) decision support system, named Pan [10, 11], to perform an investigation. Two conditions were compared where one group of analysts were given system transparency and another group were not. Pan was chosen for this study because it is designed for criminal investigations and delivers transparency by providing descriptions of the goals and constraints of the various processes it performs. The findings indicate that system transparency enhances the ability of experts to construct explanatory hypotheses and direct their inquiries. This study differs from typical evaluations of XAI systems, such as those that use relatively simple or unrealistic scenarios [2] with non-expert users. In this study, the role of transparency is assessed in terms of what is needed to perform expert reasoning about a real-world situation.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585786

2 STUDY BACKGROUND AND CONTEXT

In this study, analysts interacted with the Pan natural language information retrieval system [10, 11]. Pan is designed to support criminal intelligence analysts and investigators to easily retrieve information as they pursue lines of inquiry in investigations. The system aims to provide transparency of the functional processes it performs, where the design for transparency is defined by the Algorithmic Transparency Framework [8]. Chen et al. [3] present a situation awareness-based agent transparency model (SAT) that reflects the requirements to understand an agent's task parameters, logic and predicted outcomes, across three levels. The SAT model provides a useful guide to assess the degree to which transparency requirements for Situation Awareness are met. We revisit the SAT model in the results and discussion section of this paper, to consider whether the Pan system is effective at delivering transparency of the system processes.

Unsurprisingly, given the critical nature of decision making in risky domains such as intelligence analysis, guidance on the needs for system transparency across ethical [4, 6, 16], legal [7], and commercial [5, 18] domains, as well as for the intelligence community [12], focus upon providing auditability, accountability, and fair justifications for the system behaviour and resulting decisions. Shneiderman [19] introduces the concept of Human-Centered AI (HCAI), involving a range of research areas including fairness and explainability and also emphasises the core goal of producing designs for systems that are reliable, safe and trustworthy. Intelligence analysis involves intellectually challenging tasks where an analyst is trying to make sense of a situation by reasoning and drawing inferences from potentially large amounts of available data [22] with uncertainties and information gaps [23]. Analysts perform abductive reasoning to handle the uncertainties, by constructing and exploring explanatory hypotheses. The process of constructing explanatory structures, or 'cognitive frames', based upon the data available is captured by Klein et al. [15] in the Data-Frame model of sensemaking. It is important that an analyst understands the caveats that surround the information they use to construct a frame so they can accurately seek and infer data to elaborate upon a frame or detect inconsistencies and question a frame. Without an accurate appreciation of the information provided by a system, including the goals and the constraints of underlying processes, an analyst cannot effectively perform this sensemaking activity.

Whilst we agree with other research that trust, ethics, fairness, and accountability do necessitate system transparency, in this paper we propose that transparency is also necessary to support user expertise and performance. This is particularly important in domains such as intelligence analysis that involve expert-led sensemaking with partial and unreliable data. Our hypothesis is at odds with previous research expressing concerns that transparency may harm insight, for example, by increasing cognitive load [1, 17, 20]. In this paper, we test a series of hypotheses to assess the impact of transparency provision by the Pan system: **H1**: Analysts with transparency have greater system understanding than those with no transparency. **H2**: Analysts with transparency have less misunderstanding that those with no transparency. **H3**: Analysts with transparency have fewer 'meaningful delays' than those with no transparency. **H4**: Analysts with transparency have more scenario understanding than those with no transparency. **H5**: Analysts with transparency have less scenario misunderstanding, than those with no transparency.

3 THE PAN NATURAL LANGUAGE SYSTEM

The Pan system [10, 11] was chosen to be used in the experiments because it provides transparency to support expert decision making. Pan matches a user's natural language input with their 'intent' to trigger different functional processes and complete information retrieval tasks. Transparency is delivered in accordance with the Transparency Framework [8], by providing explanations of the data returned by the system together with the ability to inspect and verify the goals and constraints of functional processes applied. These processes reflect the recognition aspects of the Recognition-Primed Decision (RPD) model [13] and descriptions are given for the goals and constraints of each. The key aspects of the RPD model relevant to recognition are: **Plausible Goals:** Understanding the types of goals that can be reasonably accomplished within the context of the situation. **Relevant Cues:** The cues that are important within the context of the situation. **Expectancies:** Expectancies that can serve as a check on the accuracy of the situation assessment. **Actions 1…n:** Identifying the typical actions to take. The RPD model [13] represents how decisions are made in natural environments involving factors such as time pressure, high stakes, and imperfect information. Information retrieval in intelligence analysis involves these factors and analysts make decisions based upon their experience and recognition of the most appropriate option to progress an investigation.

In a previous study, 4 experienced intelligence analysts performed Cognitive Task Analysis (CTA) interviews [9] following the Critical Decision Method (CDM) [14]. These interviews described analyst expertise, cues, goals and decision making from start to end for a memorable investigation. It was found that specific questions reflecting the intent of individual retrieval tasks could be inferred from analyst statements about the decisions they made. As described by Hepenstal et al. [11], a total of 658 specific questions were inferred from analyst statements across all four investigations. In their investigations, analysts sought answers to these questions as quickly as possible and performed a series of processes for each. The decisions made when conducting lines of inquiry could be characterised as recognitional decisions and defined by the Recognition-Primed Decision (RPD) model [13]. To build Pan, computational processes were identified and developed

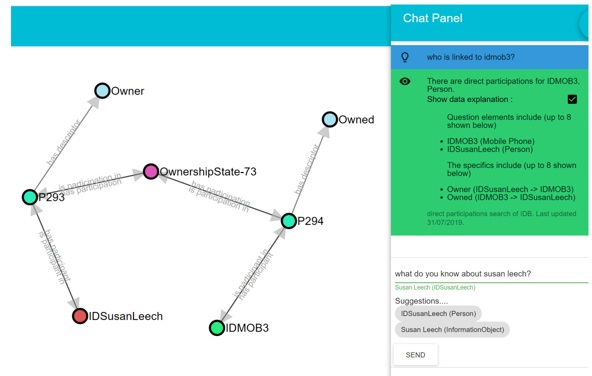

Figure 1: Screenshot of network graph explanation [11]

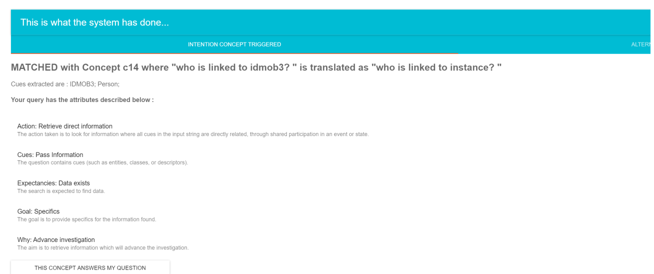

Figure 2: Screenshot of RPD functional process descriptions [11]

that could provide answers to the 658 questions where the different types were aligned thematically with the recognition aspects of the RPD model. The RPD processes, when combined, meet analyst intent for the information retrieval task underlying each question and deliver appropriate responses when triggered in sequence. Pan allows a user to explore explanations of responses through a network graph (Figure 1) and also see descriptions of the goals and constraints of each functional process triggered to answer their questions, utilising the structure of the RPD model (Figure 2). It was hypothesised that an analyst would be able to recognise the way in which their intent had been classified by the system and assess the significance of the goals and constraints of individual processes.

4 METHOD

This study compared two conditions to assess whether system transparency improved an analyst's ability to reason and form hypotheses about data retrieved in an investigation.

4.1 Participants

10 operational intelligence analysts were recruited from the police, National Crime Agency (NCA), military, and prison service in the UK to take part individually in an investigation exercise. Each analyst had a minimum of 3 years full time experience in a role involving network analysis. 5 analysts were provided with access to the transparency information and 5 were not. In this study, all

analysts were provided with explanations of the data (Figure 1). Those with transparency could also see descriptions of the system behaviour aligned to the RPD model (Figure 2), while those with no transparency could not. Participants provided their consent to take part in the study, which lasted approximately between 45 minutes and 1 hour in total.

4.2 Materials and Experimental Setup

Due to the Covid-19 pandemic it was not possible to run the study in person. Instead, investigations were performed individually through virtual conferencing software. Analysts could see Pan through screen sharing software and performed interactions with the system by requesting them verbally to a researcher.

4.3 Experimental Design and Procedure

In the experiment, analysts were asked to pose questions to the Pan system. Each question represented an information retrieval task that involved a sub-condition, matching different levels of search complexity. These levels of complexity were necessary to help understand the circumstances requiring transparency. The system was deliberately restricted to a small number of functional processes (10), so as not to overwhelm users given the short time available for each experiment. In these experiments, the search method was defined by the action function - for this there were three levels of complexity. Greater complexity existed when there was less consistency in the underlying processes and thus it was harder for a user to recognise the goals and constraints of a method. The simplest search method looked for directly associated information, with a consistent search pattern and rule applied e.g. 'what vehicles are owned by Person A?' Higher levels of complexity introduced ambiguity, for example, a user could look for further connections between entities by asking, 'what vehicles are connected to Person A?' The search pattern varied depending upon the length of the shortest path, but the search rules were consistent i.e. a shortest path will be returned up to three degrees. A user could also find similar entities where there was inconsistency in both the search pattern and clustering rules, which depended upon the input.

Participants sought to form a plausible hypothesis about the owner of a mobile phone - IDMOB1 - as part of an investigation into the sale of an illegal firearm. The researcher followed a script throughout the investigations to gather information from participants consistently about each question posed to the system, including asking what they were considering, any concerns they had, and what they wanted to ask next. So that investigations started at the same point, the first question to the system was posed by the researcher. The results from this question were used to introduce the system, including the transparency information for those with access to this. We were concerned that the results could be influenced by the general usability of the system, rather than the transparency condition we sought to test. This risk was mitigated by the involvement of a single researcher, who understood the system, across all investigations. The researcher ensured consistent usability. Once an analyst formed a hypothesis they were comfortable with, their investigation was complete. A further interview, not reported in this paper, provided a post-study briefing.

4.4 Investigation Scenario

The analysts were all provided with the same briefing material at least a week prior to the exercise. This introduced the scenario (with a network graph), task, and some limited information about the capabilities of Pan. Analysts were informed that a text message (communication activity) had taken place between IDMOB1 and IDMOB2, that said "I'll buy the gun." IDMOB2 was known to be owned by a criminal, Dan Govey, with links to an organised crime group called DGX Bodywork. The scenario was based on a real investigation. There was not enough evidence to confirm any single hypothesis, however, it was expected that analysts would consider a person, called 'Paul Richards', as the most likely owner of IDMOB1, given his connections to both IDMOB1 and to DGX Bodywork. Once the analyst found these connections, the researcher started to ask them if they had a hypothesis.

5 DATA ANALYSIS

The investigations were recorded and audio was transcribed and time stamped, comprising a total of 371 utterances (6838 words). For each utterance, timestamps were recorded for the overall time since the start of the exercise and the running time since the opening investigation question. The time since the previous question was calculated for all questions and those that were longer than 3 minutes were defined as 'meaningful delays'. Over 90% (132) of questions involved a delay of less than 3 minutes since the previous question, therefore by considering the most extreme cases we mitigated the influence of factors other than transparency. The questions following meaningful delays are circled with a red border in Figure 3. Emergent Themes Analysis (ETA) [21] was performed to delve into broad themes and identify specific themes. Careful rules were followed to encode themes so that comparisons could be drawn between the two groups and differences were explored further through qualitative analysis.

In a previous trial we found that inexperience in network analysis tasks caused misinterpretation of the graph, even when a user understood the system processes. It was therefore important to determine whether the ability of an analyst to reason about the data reflected their understanding of the scenario and task or their actual understanding of the system. To make this distinction, utterances were encoded with broad themes to reflect whether they concerned either the system or scenario. Differences could then be identified between conditions with a focus upon understanding of the system. An investigation timeline was plotted to compare analysts under the two conditions, shown in Figure 3. Quantitative analysis highlighted performance differences between the transparency conditions, while qualitative ETA enabled a deeper dive into the analysts' thinking to understand the reasons for differences.

6 RESULTS

6.1 Thematic encoding of scenario and system related understanding

Utterance encodings that reflected the broad theme of the **scenario** included sub-themes for *understanding*, *misunderstanding*, and *questions*. These sub-themes emerged as the data was explored. Analyst utterances were encoded as *understanding* of the **scenario** when

Figure 3: Visualisation Comparing Analyst Investigation Timelines Transparency vs. No Transparency: Analysts working with "No Transparency" took more time to complete their tasks. The visualisation shows a timeline of the different utterances made by analysts during their investigative exercise. The questions asked of Pan are also displayed as circle nodes and links show where results from a question were used in a subsequent question. Questions following meaningful delays, of more than 3 minutes, have a thick red border.

analysts viewed the information displayed in responses to their questions and correctly surmised the data as presented, for example: `A10: Ok, so Susan owns the phone. [21:15] (No Transparency)` *Misunderstandings* were also identified when analysts incorrectly interpreted the information shown to them, for example: `A9: but we know Susan Leech was IDMOB3 and has been in contact with Dan Govey [16.50] (No Transparency)` This was incorrect; no contact had been retrieved or was shown on the network graph. Therefore, if the analyst made an incorrect statement about the scenario information returned by the system, such as stating that a link did not exist when it was presented on the network graph, then this was encoded as scenario misunderstanding. *Questions* about the scenario or task were also encoded. Similarly, encodings that related to the broad theme of **system** included sub-themes for *understanding*, *misunderstanding*, and *questions*. If an analyst made an utterance that indicated *understanding* about the nature of the underlying **system** processes, even when describing the scenario, then this was encoded as being **system** *understanding*, for example: `A1: Paul does have a shortest path connection to bodyworks. [21:30] (Transparency)` Here, Analyst A1 correctly interpreted the graph and accurately commented on the constraints of the retrieval process. Thus, this statement was encoded as **system** *understanding*. If an analyst misinterpreted the system processes then this was encoded as **system** *misunderstanding*, even if they correctly understood the information in the graph.

6.2 Investigation timelines by transparency condition

Sequences of interactions with Pan were plotted on a timeline for each investigation, presented by Figure 3. The upper half of the figure shows a timeline of interaction events where analysts were provided with transparency; while the lower half of the figure shows a timeline of interaction events where analysts had no transparency. Each chart comprises (1) a series of horizontal lines and shapes for each analyst with the time taken to complete the exercise and (2) a network of nodes and edges. The shapes and colours reflect thematic encodings.

(1) The horizontal lines represent a timeline for each analyst. The shapes are coloured to show when the analysts demonstrated understanding (green), misunderstanding (red), or asked a question related to the scenario or the system (grey). System related utterances are shown as stars and scenario related utterances are diamonds.
(2) The network graph comprises all the questions asked by analysts in each condition, where each node is a question. This

shows the distribution and clustering of questions over time and linked questions (where the results of one are used as inputs in another). The similar heights show that analysts from each condition asked a similar number of questions overall, however those with transparency did so more quickly with fewer long pauses between questions. The question in each investigation that defined the point at which the analysts found links between 'Paul Richards' and both the mobile phone of interest ('IDMOB1') and the organisation 'DGX Bodywork', is highlighted with a green circle around the node.

All analysts – both with transparency and with no transparency – understood the scenario and explanations provided by the system, and were able to complete the investigation exercise by posing a plausible hypothesis that was accurately supported by data. However, Figure 3 shows that those analysts with transparency of the system processes decided upon a hypothesis faster than analysts with no transparency. Thematic codes are common across analysts in each condition, rather than being focused on just one or two analysts. For example, there are more green stars (system understanding) across all the analysts with the transparency condition, and more red stars (system misunderstanding) across the analysts without transparency. Figure 3 helped to identify areas for closer qualitative analysis. A surrogate measure was calculated to reflect relative understanding and misunderstanding for quantitative comparisons: by counting the number of utterances made, or delays, for each theme and calculating the average rate for a 10-minute period for the two conditions (to account for investigations of different durations). The utterance counts provided a measure to compare the two groups and to indicate reasons for differences observed in terms of investigation performance. A series of one-tailed t-tests were performed, testing one-directional null hypotheses with p-value less than 0.05. Those with transparency did not demonstrate a significantly higher rate of understanding of the scenario (**H4**, p=0.485), or a significantly lower rate of misunderstanding of the scenario (**H5**, p=0.364). However, they did demonstrate a significantly higher rate of system understanding (**H1**, p=0.042) combined with a significantly lower rate of system misunderstanding (**H2**, p=0.005) and meaningful delays (**H3**, p=0.009). We infer therefore, that the differences observed across investigation timelines was due to transparency and the impact this had on system understanding. We explore this further with qualitative analysis.

6.3 Qualitative analysis of the influence of system transparency on analyst reasoning

From qualitative analysis of analyst utterances throughout the investigations we have inferred that analysts with transparency could form expectations about what the system would do, given their knowledge of its limitations, while those with no transparency demonstrated less understanding and greater misunderstanding. Analysts with transparency indicated that they had learned what functional processes would be triggered by their questions, including the constraints, and used this knowledge to inform their interactions. Analysts with transparency demonstrated multiple examples of understanding that reflected all levels of the SAT model. There were two cases where analysts with transparency misunderstood the system prior to them seeing the transparency information. In both cases, they rectified their misunderstanding once they viewed the transparency information. For example, Analyst 4 initially misunderstood the search constraints applied by the system, `A4: I'm assuming if it has not found a shortest path then there are no paths [14:15] (Transparency)` On inspecting descriptions of the goals and constraints of each of the RPD functional processes, they recognised their error and demonstrated understanding about what the system was trying to achieve, `A4: but it is only looking for three. Oh right. [14:25] (Transparency)` Analyst 4 later demonstrated awareness of the constraints of the search method i.e. that there were potentially other paths that had not been explored, `A4: So that is the shortest path between those two [Watson Custom Metalwork and DGX Bodywork]. . . I'd want to do any path between those two. [18:00] (Transparency)` Their understanding informed subsequent interactions with the system, `A4: I'd probably now be intrigued to see what other links there are between Watson Custom Metalwork and DGX Bodywork. I think the easiest way to do that would be to ask two questions [18:20] (Transparency)`

Analysts with no transparency could understand the lowest level of complexity from the graph explanations Pan provided (defined in section 4.3). However, they did not demonstrate understanding of higher levels of complexity. Analysts with transparency were able to inspect the goals and constraints of the system processes and found answers that helped them to understand more complex functions, addressing inherent ambiguities. For example, that the shortest path search function was constrained to only three degrees. The understanding gained by analysts from the transparency information helped in future interactions. For example, Analyst 8 posed a question that retrieved only directly linked entities and then viewed the transparency information. They then said, `A8: So, are there any indirect links? . . . Are there any entities that link both to Susan Leech and Darren Smith? [14.40] (Transparency)` Due to the descriptions of the goals and constraints of the search function, provided by the transparency information, the analyst understood the constraints with the original search method, specifically that it would not capture indirect links, and rephrased their question to look for common entities that may sit between the entities. In contrast, every analyst without transparency misunderstood the system processes by forming an incomplete or incorrect interpretation of the goals and constraints. For example, `A2: And the only connection is through him [Dan Govey] having to know Paul Richards. From this I would look at it as a red herring and disregard Susan Leech in the wider context. [18:55] (No Transparency)` Analyst A2 ignored additional connections between Paul Richards and Dan Govey, due to their misunderstanding of the search constraints. Analysts with transparency made informed choices about how to proceed with their investigations and interact with the system by adjusting their language, for example, `A1: Can we go looser than association, can we say 'connection' (instead of association)? [15:00] (Transparency)` Analyst A1 knew that their question about 'associated' entities would consider only direct links in the graph, so instead by asking for 'connections' they expanded their search. With no transparency, however, analysts could not re-evaluate their assumptions about the system processes or learn how to interact with it effectively. Analysts with no transparency were significantly more likely to delay for more than 3 minutes between questions, being responsible for every such delay after 6 minutes of the exercise. Utterances during these later delays showed confusion about the system processes, including about how the information had been retrieved and a desire to explore additional lines of inquiry beyond the scope of the system.

7 DISCUSSION AND CONCLUSIONS

Analysts with transparency were able to reach a conclusion to their investigations faster than those who had no transparency (Figure 3) and with significantly fewer meaningful delays. This was not due to differences in the analysts' ability to interpret the network graph visualisation, given that all analysts demonstrated that they understood the scenario and the task and formed plausible hypotheses supported by evidence. We have inferred, therefore, that transparency as provided by the Pan system, effectively supported analysts to reason about the data retrieved, particularly with regards to higher levels of search complexity. With no transparency, analysts did not understand more complex system processes with less obvious constraints. They could not interpret the goals and constraints from explanations of the data alone and as a result could not effectively reason about the data they retrieved. With no transparency, analysts instead attempted to guess what the system had done and did not appreciate the true constraints, for example, `A3: Ok, so [when I say] 'connections' are indirect (links) effectively [13:25] (No Transparency)` In this example case, Analyst A3 had assumed that all indirect links would be retrieved by their 'connections' question, however this was not the case. As a result, Analyst A3 could not reason with a true understanding of the available data and this led to a misinformed explanatory hypothesis that discounted a key potential suspect. With no transparency, analysts could not appreciate what information was missing when they were reasoning towards explanatory hypotheses. In contrast, analysts with transparency recognised and understood the goals and constraints of all the capabilities experienced, even for higher complexity. This understanding had a direct impact on their ability to reason about the information retrieved and to conduct subsequent system interactions, which ultimately led to a more informed decision on their hypothesis. We propose these findings are generalisable to other situations where reasoning about data informs complex decisions.

This initial study has a number of limitations to address in future work, for example, the dataset was small and considered a single scenario, capabilities were restricted, and participants had a limited time to interact with the system so could not learn the kinds of responses provided to them by different questions. Additionally, analysts were not able to interact directly with the system due to Covid-19 restrictions, so we could not assess the true usability.

REFERENCES

[1] Ashraf Abdul, Christian von der Weth, Mohan Kankanhalli, and Brian Y. Lim. 2020. *COGAM: Measuring and Moderating Cognitive Load in Machine Learning Model Explanations.* Association for Computing Machinery, New York, NY, USA, 1–14.

[2] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary Främling. 2019. Explainable Agents and Robots: Results from a Systematic Literature Review. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems* (Montreal QC, Canada) *(AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1078–1088.

[3] Jessie Chen, Katelyn Procci, Michael Boyce, Julia Wright, Andre Garcia, and Michael Barnes. 2014. Situation Awareness–Based Agent Transparency.

[4] Karl de Fine Licht and Jenny de Fine Licht. 2020. Artificial Intelligence, Transparency, and Public Decision-Making. *AI and Society* 35, 4 (2020), 917–926. https://doi.org/10.1007/s00146-020-00960-w

[5] D. Delmolino and M. Whitehouse. 2018. Responsible AI: A framework for building trust in your AI solutions.

[6] Penny Duquenoy, Donald Gotterbarn, Kai Kimppa, Norberto Patrignani, and B.L.William Wong. 2018. *Addressing Ethical Challenges of Creating New Technology for Criminal Investigation: The VALCRI Project.* 31–38. https://doi.org/10.1007/978-3-319-89297-9_4

[7] Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society* 6, 1 (2019), 2053951719860542.

[8] Sam Hepenstal, Neesha Kodagoda, Leishi Zhang, Pragya Paudyal, and B. L. William Wong. 2019. Algorithmic Transparency of Conversational Agents. In *IUI Workshops*.

[9] Sam Hepenstal, B.L. William Wong, Leishi Zhang, and Neesha Kodogoda. 2019. How analysts think: A preliminary study of human needs and demands for AI-based conversational agents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (2019), 178–182.

[10] Sam Hepenstal, Leishi Zhang, Neesha Kodagoda, and B. L. William Wong. 2020. Pan: Conversational Agent for Criminal Investigations. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 134–135.

[11] 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 (aug 2021), 35 pages.

[12] Intel.gov. 2022. Principles of artificial intelligence ethics for the intelligence community. https://www.dni.gov/index.php/features/2763-principles-of-artificialintelligence-ethics-for-the-intelligence-community. Accessed: 2022-06-16.

[13] G.A. Klein. 1993. A recognition-primed decision (RPD) model of rapid decision making. In *Decision Making in Action: Models and Methods*, G.A. Klein, Judith Orasanu, R. Calderwood, and Caroline E. Zsambok (Eds.). Norwood: Ablex Publishing Corporation, 138–147.

[14] G.A. Klein, R Calderwood, and D MacGregor. 1989. Critical decision method for eliciting knowledge. *Transactions on Systems, Man, and Cybernetics* 19, 3 (1989), 462–472.

[15] G. Klein, J. K. Phillips, E. L. Rall, and D. A. Peluso. 2007. A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, R. R. Hoffman (Ed.). Lawrence Erlbaum Associates Publishers, 113–155.

[16] David Leslie. 2019. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *The Alan Turing Institute* (2019). https://doi.org/10.5281/zenodo.3240529

[17] Arun Rai. 2020. Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science* 48, 1 (January 2020), 137–141.

[18] R Roovers. 2019. Transparency and responsibility in artificial intelligence. A call for explainable AI.

[19] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504. https://doi.org/10.1080/10447318.2020.1741118 arXiv:https://doi.org/10.1080/10447318.2020.1741118

[20] 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)*. Association for Computing Machinery, New York, NY, USA, 107–120.

[21] B.L.William Wong. 2004. Data analysis for the Critical Decision Method. *The Handbook of Task Analysis for Humancomputer Interaction* (01 2004).

[22] 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, 1 (2016), 178–182.

[23] B.L.William Wong and Margaret Varga. 2012. Black Holes, Keyholes And Brown Worms: Challenges In Sense Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56 (10 2012), 287–291. https://doi.org/10.1177/1071181312561067