Formalizing and evaluating knowledge-based AI tools for intelligence analysis

Roos Scheffers

Utrecht University

Abstract

In the police context, intelligence analysis involves collecting, evaluating, and combining uncertain, incomplete, and often conflicting information to help understand patterns, assess risks, and support decision-making. This process is cognitively complex and prone to error. Artificial intelligence (AI) tools have the potential to assist analysts in this process, but only if they are sufficiently transparent to meet legal requirements. In my PhD research, I investigate how knowledge-based AI, consisting of knowledge provided by experts and rules to reason over this knowledge, can be used to aid intelligence analysis at the Dutch police and judiciary. This paper first outlines the problem and motivation for this research, and then describes the studies conducted so far as well as plans for future work.

1 Introduction

In intelligence analysis, analysts collect, evaluate, and combine evidence to evaluate hypotheses and draw conclusions [Heuer, 1999]. This information is often conflicting and incomplete, making the process laborious and complex, which can lead to costly errors. [Pirolli and Card, 2005]. At the Dutch police, intelligence analysis can have several goals, such as assessing risks, understanding criminal patterns, and supporting decision-making. In this context, intelligence analysis is influenced by domain-specific challenges and the organizational structure, increasing complexity [Visser et al., 2023]. AI assistance could save time and effort in intelligence analysis, but must comply with legal and organizational requirements. For evidence to be admissible in court and to support law enforcement actions (i.e., getting a search warrant), it must be clear why certain conclusions were drawn and on what data they were based. Therefore, it is vital to create transparent and understandable AI systems [Dechesne et al., 2019]. To investigate how AI can be used to aid intelligence analysis, my PhD will be an interdisciplinary project focusing on three fields: explainable artificial intelligence, cognitive science, and AI & law.

1.1 Explainable artificial intelligence

Research in the field of explainable artificial intelligence (XAI) aims to make transparent and understandable AI models and to extend existing opaque models to be transparent and understandable [Longo et al., 2020]. In a sensitive domain such as intelligence analysis, these aims are crucial. Machine learning models are very proficient at processing large and complex data sets, but their black-box nature is problematic for use in intelligence analysis. One approach to mitigate this problem is to combine machine learning models with knowledge-based AI, as is done in neurosymbolic approaches. Knowledge-based AI, which consists of knowledge created by domain experts and rules to reason over this knowledge, is inherently transparent because all decision rules can be accessed; however, it is not necessarily understandable since there can be many rules with complex interactions. Explanations can aid in un-

derstanding these transparent systems and the reasoning behind individual decisions. One specific approach to knowledge-based AI that can provide such explanations is formal argumentation. Formal argumentation uses transparent argumentation frameworks, which consist of arguments and relations between them. Argumentation frameworks and decisions made based on them can be explained using a variety of methods to make them understandable [Borg and Bex, 2024, Fan and Toni, 2015].

1.2 Cognitive science

Cognitive science highlights the importance of understanding the cognitive and social biases people use when analyzing evidence, drawing conclusions, and providing explanations. This leads people to prefer decisions and explanations with specific characteristics. XAI can build on this and make use of it to make sure systems and explanations are understandable to people [Miller, 2019]. Argumentation, which resembles elements of human conversation and reasoning, could be valuable for providing explanations [Mercier and Sperber, 2011]. For this purpose, several argumentation-based explanation approaches have been defined [Borg and Bex, 2024, Fan and Toni, 2015]. While various argumentation-based explanation methods aim to provide socially desirable and understandable explanations, it remains uncertain whether these formalisations are successful in representing human cognition. Despite experimental validation of other elements of argumentation [Rahwan et al., 2010, Guillaume et al., 2022], explanations have remained overlooked. This is unfortunate, as social and cognitive factors strongly influence explanations, making experimental validation essential.

1.3 Computational argumentation

The field of computational argumentation is important to this research for two main reasons. First, AI tools used by law enforcement need to comply with legal requirements by being transparent about their workings and keeping track of the processing of pieces of evidence. Transparent and deterministic systems such as computational argumentation are well-suited for this purpose. Second, in intelligence analysis, where evidence can be produced for court, evidence identified by AI and explanations of this evidence should be understandable by judges and lawyers. Argumentation offers opportunities for this since it shares similarities with legal reasoning. Both involve reasoning with facts and rules [Crombag et al., 1994]. Law is also connected to argumentation because of its adversarial, justified, defeasible, and open-textured aspects [Bench-Capon, 2020, Prakken and Sartor, 2015].

1.4 Research questions

In summary, in intelligence analysis, there is a need for transparent and understandable AI tools, which knowledge-based AI and specifically argumentation could provide. However, to identify where and how these tools could best be used and to evaluate such tools, a multidisciplinary investigation is needed, considering XAI, cognitive science, and AI & law. The overarching research question of my PhD is: *How can knowledge-based AI be used to provide explanations for intelligence analysis?* This research question is specified using several sub-questions.

SQ1a *How can existing theories of explanations in formal argumentation best be used to provide explanations for people?*

SQ1b *Which features of argumentation-based explanations apply specifically to law enforcement and judicial personnel?*

SQ2 *Can structures from knowledge-based AI be used to describe and gather knowledge from the work of intelligence analysts?*

SQ3 *How can decisions and explanations by neural and symbolic AI be compared?*

2 Methods

This section describes the different methods that will be used to investigate the sub-questions outlined

above. These questions will be tackled approximately in order; therefore, more details will be provided on the first question and fewer on the last.

2.1 Experiments

Most validation of argumentation theories is based on theory or examples, but can also be done with experiments. Experimental validation is important to ground argumentation theories in the real world [Rahwan et al., 2010]. This is especially true for explanation methods based on argumentation since, for good explanations that can be practically applied, social and cognitive factors are vital to take into account [Miller, 2019]. These social cognitive factors can only be validated experimentally.

In prior work [Scheffers et al., 2024], I compared explanations generated by participants to explanations based on relatedness in formal argumentation. These explanations using relatedness were found to be cognitively plausible, meaning that they accurately described human behavior. Currently, I am working on a follow-up study to this paper. This follow-up study aims to provide an empirical basis for different types of explanations based on formal argumentation. Specifically, I will consider sufficiency, compactness, and minimality [Borg and Bex, 2024, Fan and Toni, 2015]. For this experiment, 20 natural language instantiations of argumentation frameworks were created. These instantiations are stories written in normal language that follow the structure of an argumentation framework (Example 2.1 corresponds to the argumentation framework in Figure 1). Participants will be presented with these instantiated argumentation frameworks and asked to explain a topic argument (A in Figure 1). The explanations given by participants will be compared with theoretical explanations.

After this study, I plan to conduct a similar experiment with domain experts such as law enforcement employees, judges, and lawyers. To investigate whether their preferences for explanations differ from those of the general public. These experts reason with evidence in specific manners in specialized tasks, which can differ from the reasoning used by the general population [Heuer, 1999]. Since in intelligence analysis both stories and arguments are used to reason with evidence [Bex, 2011], argumentation theories might align more closely to their reasoning patterns than to those of the general population. Therefore, they might benefit more from argumentation-based explanations.

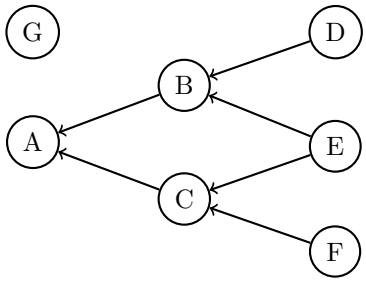

Figure 1: Argumentation framework $\mathcal{AF}_1$. A is the topic argument which is attacked by arguments B and C. In the experiment, participants are instructed to use any combination of arguments D, E, F, and G to explain the topic argument.

Example 2.1. A natural language example for $\mathcal{AF}_1$.

(A) Doctor A says the patient has Disease X based on their symptoms. So, they have Disease X.

(B) Doctor B says the symptoms show the patient has Disease Y. So, the diagnosis of Disease X is wrong.

(C) The patient's family has a history of Disease Y. So, the patient has Disease Y.

(D) Doctor B is an expert in rare diseases and is focusing too much on the chance of Disease Y. So, their opinion is biased.

(E) Lab tests do not show any signs of Disease Y. So, the patient does not have Disease Y.

(F) Disease Y is not passed down in families. So, the patient's family history of Disease Y does not matter.

(G) Patient surveys show that people are unhappy with Doctor C's diagnosis. So, Doctor C's methods should be checked.

2.2 Case study

To investigate SQ2, I plan to conduct a case study to bridge the gap between theoretical applications of knowledge-based AI and the practice of intelligence analysis. This study will collect knowledge that can serve as input for a knowledge-based AI system, which can ground future research in the real processes of intelligence analysis. Another reason for conducting an in-depth qualitative study is to look in detail at the intelligence analysis process to be able to identify at what points in the process XAI tools can be beneficial.

Little is known of the specific knowledge and inference rules used in the intelligence analysis process. Studies have investigated the structure of reports [Kruger et al., 2022] and conversations by intelligence analysts on a digital platform [Saletta et al., 2020], but these do not provide the detail required for knowledge-based AI. Several studies have annotated conversations and debates with argumentation. For example, bipolar argumentation has been used to map online debates [Cabrio and Villata, 2013], tweets [Bosc et al., 2016], and conversations in the Penn Treebank data set [Rosenfeld and Kraus, 2014]. The methods used in these studies could be adapted for use in the intended case study.

The case study will be conducted with police and justice personnel. They will be presented with a collection of digital evidence and asked to think out loud while forming hypotheses and evaluating these to draw a conclusion. The observed behavior and utterances of participants will be annotated using formal argumentation. Based on these annotations, conclusions will be drawn about the comparability of argumentation and the process of intelligence analysis.

2.3 Evaluation

There are two distinct approaches in the field of AI: machine learning-based and knowledge-based approaches. While the former efficiently processes extensive data, it often lacks transparency and the ability to provide causal explanations. The latter method is slower and reliant on domain expertise but offers transparency [Campagner and Cabitza, 2020]. Recently, neurosymbolic approaches have attempted to merge the two approaches in an attempt to benefit from the strengths of both methods [d'Avila Garcez and Lamb, 2023]. Due to the fundamental differences between deep learning and knowledge-based approaches, it is challenging to effectively compare their output and performance. Moreover, explanations and output for both types of models differ fundamentally. This raises the question of how these explanations can be compared and evaluated. The goal of this study is to establish a framework or a set of guidelines to facilitate the comparison and combination of different AI approaches.

This study will facilitate choosing between AI methods depending on individual needs and facilitate neurosymbolic AI by gaining a better understanding of neural and symbolic components, which could allow for a more complementary combination of the two. Such integration has the potential to bridge the gap between professional practice and machine learning techniques.

3 Conclusion

This paper describes the planned steps to take over the next three years towards the goal of using argumentation and knowledge-based AI in the process of intelligence analysis at the Dutch police. This will be done using three types of studies. In the first study, insights into which types of argumentation-based explanations are most suitable for general explanations and explanations to law enforcement and judiciary personnel will be gained. The second study will provide insight into the compatibility of the reasoning used in intelligence analysis and formal argumentation. The final study will provide guidelines to facilitate the comparisons and combination of machine learning and knowledge-based AI approaches.

Finally, the results of these studies will be used to answer how knowledge-based AI can be used to provide explanations for intelligence analysis at the Dutch police.

Acknowledgements

I would like to thank Floris Bex, Matthieu Brinkuis, and Stephan Grimmelikhuijsen for their guidance. The described research is part of the 'AI4Intelligence: From Multimodal Data to Trustworthy Evidence in Court' project with file number KICH1.VE01.20.011 of the research programme Data and Intelligence which is partly financed by the Dutch Research Council (NWO).

References

[Bench-Capon, 2020] Bench-Capon, T. J. M. (2020). Before and after Dung: Argumentation in AI and Law. *Argument & Computation*, 11(1-2):221–238.

[Bex, 2011] Bex, F. J. (2011). *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*, volume 92 of *Law and Philosophy Library*. Springer Netherlands, Dordrecht.

[Borg and Bex, 2024] Borg, A. and Bex, F. (2024). Minimality, necessity and sufficiency for argumentation and explanation. *International Journal of Approximate Reasoning*, 168:109143.

[Bosc et al., 2016] Bosc, T., Cabrio, E., and Villata, S. (2016). Tweeties squabbling: Positive and negative results in applying argument mining on social media. In Baroni, P., Gordon, T. F., Scheffler, T., and Stede, M., editors, *Computational Models of Argument*, volume 287 of *Frontiers in Artificial Intelligence and Applications*, pages 21–32. IOS Press.

[Cabrio and Villata, 2013] Cabrio, E. and Villata, S. (2013). A natural language bipolar argumentation approach to support users in online debate interactions†. *Argument & Computation*, 4(3):209–230.

[Campagner and Cabitza, 2020] Campagner, A. and Cabitza, F. (2020). Back to the Feature: A Neural-Symbolic Perspective on Explainable AI. In Holzinger, A., Kieseberg, P., Tjoa, A. M., and Weippl, E., editors, *Machine Learning and Knowledge Extraction*, Lecture Notes in Computer Science, pages 39–55, Cham. Springer International Publishing.

[Crombag et al., 1994] Crombag, H., van Koppen, P., and Wagenaar, W. (1994). *Dubieuze Zaken: De Psychologie van Strafrechtelijk Bewijs*. Contact, Amsterdam.

[d'Avila Garcez and Lamb, 2023] d'Avila Garcez, A. and Lamb, L. C. (2023). Neurosymbolic AI: The 3rd wave. *Artificial Intelligence Review*, 56(11):12387–12406.

[Dechesne et al., 2019] Dechesne, F., Dignum, V., Zardiashvili, L., and Bieger, J. (2019). AI & Ethics at the Police: Towards Responsible use of Artificial Intelligence in the Dutch Police. White paper.

[Fan and Toni, 2015] Fan, X. and Toni, F. (2015). On computing explanations in argumentation. In Bonet, B. and Koenig, S., editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 1496–1502. AAAI Press.

[Guillaume et al., 2022] Guillaume, M., Cramer, M., van der Torre, L., and Schiltz, C. (2022). Reasoning on conflicting information: An empirical study of Formal Argumentation. *PLOS ONE*, 17(8):e0273225.

[Heuer, 1999] Heuer, R. J. (1999). *Psychology of Intelligence Analysis*. Center for the Study of Intelligence.

[Kruger et al., 2022] Kruger, A., Thorburn, L., and Van Gelder, T. (2022). Using argument mapping to improve clarity and rigour in written intelligence products. *Intelligence and National Security*, 37(5):607–626.

[Longo et al., 2020] Longo, L., Goebel, R., Lecue, F., Kieseberg, P., and Holzinger, A. (2020). Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions. In Holzinger, A., Kieseberg, P., Tjoa, A. M., and

Weippl, E., editors, *Machine Learning and Knowledge Extraction*, Lecture Notes in Computer Science, pages 1–16, Cham. Springer International Publishing.

[Mercier and Sperber, 2011] Mercier, H. and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–74.

[Miller, 2019] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38.

[Pirolli and Card, 2005] Pirolli, P. and Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, volume 5, pages 2–4.

[Prakken and Sartor, 2015] Prakken, H. and Sartor, G. (2015). Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214–245.

[Rahwan et al., 2010] Rahwan, I., Madakkatel, M. I., Bonnefon, J.-F., Awan, R. N., and Abdallah, S. (2010). Behavioral Experiments for Assessing the Abstract Argumentation Semantics of Reinstatement. *Cognitive Science*, 34(8):1483–1502.

[Rosenfeld and Kraus, 2014] Rosenfeld, A. and Kraus, S. (2014). Argumentation theory in the field: An empirical study of fundamental notions. In Cabrio, E., Villata, S., and Wyner, A. Z., editors, *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*, volume 1341 of *CEUR Workshop Proceedings*. CEUR.

[Saletta et al., 2020] Saletta, M., Kruger, A., Primoratz, T., Barnett, A., van Gelder, T., and Horn, R. E. (2020). The role of narrative in collaborative reasoning and intelligence analysis: A case study. *PLOS ONE*, 15(1):e0226981.

[Scheffers et al., 2024] Scheffers, R., Bex, F., and Borg, A. (2024). Related Explanations in Formal Argumentation, an Empirical Study. In Reed, C., Thimm, M., and Rienstra, T., editors, *Computational Models of Argument*, Frontiers in Artificial Intelligence and Applications, pages 265–276. IOS Press.

[Visser et al., 2023] Visser, C., Markus, A., De Poot, C., Kop, N., and Weggeman, M. (2023). Sensemaking and evidence in criminal investigations of organised crime: A literature review. *International Journal of Police Science & Management*, page 14613557231196377.