

Can We Move Argumentative XAI into the Research & Innovation Mainstream?

Timotheus Kampik^{1,*}, Antonio Rago², Oana Cocarascu² and Kristijonas Čyras³

¹Umeå University, Sweden

²King's College London, United Kingdom

³Independent Researcher

Abstract

Computational argumentation is—within the academic community—considered a promising facilitator of AI explainability. Still, the road to large scale success, e.g. in mainstream machine learning research or industry applications, appears to be long. In this write-up, we reflect on applicability challenges facing argumentative explainable artificial intelligence and sketch a set of action items that, if addressed, can help close the gap between the status quo and the ambition of substantial real-world impact.

Keywords

XAI, Formal Argumentation, Applications of Computational Argumentation

1. Introduction

Over the past decades, Computational Argumentation (CA) has emerged as a vibrant research direction within the field of Knowledge Representation and Reasoning (KR&R). Although, initially, CA research was primarily formal in nature and applications were relatively scarce, recent years have seen a substantial increase of works that present potential or actual applications of CA, often focusing on the capabilities of CA to *explain* both symbolic and sub-symbolic inferences—so-called Argumentative eXplainable Artificial Intelligence or *argumentative XAI*—as a main selling point [1, 2, 3]. Indeed, some communities members argue that CA is crucial in order to *contest* the decisions made by AI-based systems, as it allows humans to challenge inferences drawn by machines and revise outcomes if justified [4]. Despite these ambitions, applications of CA are still nascent and particularly benefits of argumentative XAI remain to be showcased. In this write-up, we give a brief overview of the current state of applications of argumentative XAI (Section 2), highlight a set of challenges (Section 3), and set out ambitions for making argumentative XAI more applicable (Section 4). We conclude with a call to action to regularly reflect on where we stand with respect to these ambitions, for example in future editions of the International Workshop on Argumentation for eXplainable AI (ArgXAI)¹, during which the discussion that resulted in this manuscript took place (Section 5).

2. The Status Quo: First Steps towards the Real World

Traditionally, the focus of CA research has been on formal definitions of CA models and their theoretical analysis. The most notable exceptions are argument mining and debating systems, which have gained attention even on a popular science level [5, 6]. This line of applications typically employs argumentative approaches—sometimes called *informal argumentation*—that are largely detached from what the core of the formal argumentation community tends to study. In argumentative XAI research, the traditional focus on formal methods is prevalent in exploring explanations within several much-studied formal

ArgXAI-25: 3rd International Workshop on Argumentation for eXplainable AI

✉ tkampik@cs.umu.se (T. Kampik); antonio.rago@kcl.ac.uk (A. Rago); oana.cocarascu@kcl.ac.uk (O. Cocarascu); kcyras@gmail.com (K. Čyras)

ORCID 0000-0002-6458-2252 (T. Kampik); 0000-0001-5323-7739 (A. Rago); 0000-0003-1552-8636 (O. Cocarascu); 0000-0002-4353-8121 (K. Čyras)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://people.cs.umu.se/tkampik/argxai/2024.html>

argumentation approaches, such as abstract [7, 8], structured [9], and quantitative/gradual [10, 11] argumentation. Basic research code for the implementation is often shared and the community is usually able to utilize an ecosystem of tools and libraries for argumentation-based reasoning [12, 13]², at least some of which offer explainability capabilities [10]. Other research lines within argumentative XAI embrace applications more directly, for example by implementing and studying explainable recommendation systems [14], image classifiers [15], and a police online fraud reporting system [16].

3. Current Limitations

Despite the promising advancements of research on applications of argumentative XAI, the community is far from achieving substantial real-world impact. Indeed, several limitations exist that stand in the way of expanding the footprint of argumentative XAI outside of the core argumentation community.

No real-world success stories. Most strikingly, argumentative XAI is apparently far from producing major real-world successes, e.g., in commercial products, public sector application, or in tools and pipelines applied by other scientific disciplines. Here, one key problem is the lack of real-world adoption of CA more broadly. This means that argumentative XAI must be viewed as a driving force for real-world applications of CA, as applications cannot be achieved by merely “enabling” explainability on top of the (non-existent) software systems that rely on CA to any substantial extent.

Few user studies. Intuitively, one would expect that explanations must be designed specifically for the users that consume them. However, argumentative XAI often focuses on mathematical properties (such as minimal collections of arguments that ensure a certain outcome), disregarding whether those properties are required or even desirable from an end-user perspective. Only a small subset of these works—e.g. [17, 14]—study the effectiveness of argumentative XAI from a human-computer interaction perspective.

Limited software support. CA is supported by a substantial ecosystem of “academic” software tools, some of which are somewhat mature argumentation reasoners (see above). As argumentative XAI is a relatively recent trend within CA, software tooling is even more limited to one-off prototypes and *research code*-level scripts. While the status quo of software tooling may be sufficient for the core community, in which researchers tend to have a precise grasp of the formal foundations, the lack of re-usable tooling makes it difficult for users without deep expertise (or with deeper expertise, merely in a slightly different area) to adopt argumentative XAI approaches.

Little relevance outside of the argumentation community. Work on argumentative XAI is, arguably, primarily conducted by the core argumentation community and disseminated in venues where it is reviewed by CA experts. While this means that papers are judged by technically highly competent reviewers, these reviewers are already “sold” on the relevance of CA.

4. Where to Go from Here: Building an Interdisciplinary Innovation Ecosystem

The limitations outlined in the previous section give rise to a range of long-term action items the community may follow up on to realize the ambition of large-scale applications of argumentation explainability. In particular, we suggest more of the following.

Interdisciplinary research. Considering that most of the argumentation community primarily has formal logic and computer science research expertise, it becomes clear that advancing applications of argumentative XAI requires more interdisciplinary outreach, in particular to research fields

²Cf. <https://people.cs.umu.se/~tkampik/argtools/>

such as psychology, human-computer interaction, and the social sciences more broadly³. Let us highlight that the community is already going into this direction and, for example, increasingly conducting research together with psychologists (e.g. [14, 19, 20]).

Collaborations with domain experts. In order to make a strong case for the usefulness of argumentative XAI, it is crucial to connect to practitioners or researchers with deep expertise in specific application domains (going beyond interdisciplinary basic research)⁴. Examples of successful connections between experts in the real world and methods based on argumentative XAI include an approach for decision support which explains its argumentative inferences to experts from the Dutch National Police [21] for validation, as well as an argumentation-based approach to reinforcement learning in which domain experts' knowledge can potentially be incorporated into argumentation frameworks for reward shaping [22].

Realistic datasets. Strengthening the case for the broad applicability of argumentative XAI requires increasing the diversity and maturity of datasets that are utilized. Ideally, the community will move from simple example datasets that happen to be available to datasets that exemplify open problems that need to be tackled and whose underlying domain is well understood, either by members of the community or experts that are willing to collaborate with the community. Identifying relevant datasets is not straightforward, and neither is the selection and curation of data in case ready-to-use datasets do not exist.

Mature software. Enabling non-experts to apply argumentative XAI capabilities requires moving from simple and potentially not re-usable scripts yielding reproducible results to maturer tooling in popular as well as scalable languages, such as Python and JavaScript (the former), as well as C++ and Rust (the latter). Ideally, such tooling is accompanied by comprehensive documentation and tutorials, e.g. in interactive or video form.

5. Conclusion

While argumentative XAI receives substantial research attention, the case for applicability has been made mostly *within* the argumentation community. To provide broader evidence and foster a broader ecosystem for applications of argumentative XAI, we recommend to: i) mature the ecosystem of re-usable software tools for argumentative XAI; ii) advance interdisciplinary work that makes the case for argumentative XAI from an empirical, human-centric perspective; iii) position work on argumentative XAI in applied research venues, or in venues that can be expected to require a compelling case to be made, such as machine learning conferences. Over the next years, we hope to reflect on progress in these directions.

Acknowledgments

This write-up is the result of an open discussion among members of the argumentation community that occurred during the 2nd International Workshop on Argumentation for eXplainable AI (ArgXAI), co-located with the 10th International Conference on Computational Models of Argument. We thank all people who participated in the discussion, as well as Francesca Toni for additional comments. The write-up is of editorial nature and has not been formally peer-reviewed.

³Cf. Miller's seminal paper for a general call to action that argues for the crucial necessity of reaching out to the social sciences in order to advance explainable AI, and also advocates for the use of argumentative approaches to AI explainability [18].

⁴This recommendation draws from the experience of two of the authors working at the interface between basic research and industry applications (to telecommunication system automation and enterprise information systems, respectively) for large tech companies.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] K. Čyras, A. Rago, E. Albini, P. Baroni, F. Toni, Argumentative XAI: A Survey, in: Z.-H. Zhou (Ed.), 30th International Joint Conference on Artificial Intelligence, IJCAI, Montreal, 2021, pp. 4392–4399. doi:10.24963/ijcai.2021/600. arXiv:2105.11266.
- [2] A. Vassiliades, N. Bassiliades, T. Patkos, Argumentation and explainable artificial intelligence: a survey, Knowl. Eng. Rev. 36 (2021) e5. URL: <https://doi.org/10.1017/S0269888921000011>. doi:10.1017/S0269888921000011.
- [3] Y. Guo, T. Yu, L. Bai, J. Tang, Y. Ruan, Y. Zhou, Argumentative explanation for deep learning: A survey, in: 2023 IEEE International Conference on Unmanned Systems (ICUS), 2023, pp. 1738–1743. doi:10.1109/ICUS58632.2023.10318322.
- [4] F. Leofante, H. Ayoobi, A. Dejl, G. Freedman, D. Gorur, J. Jiang, G. Paulino-Passos, A. Rago, A. Rapberger, F. Russo, X. Yin, D. Zhang, F. Toni, Contestable ai needs computational argumentation, 2024. URL: <https://arxiv.org/abs/2405.10729>. arXiv:2405.10729.
- [5] J. Lawrence, J. Visser, C. Reed, Translational argument technology: Engineering a step change in the argument web, Journal of Web Semantics 77 (2023) 100786. URL: <https://www.sciencedirect.com/science/article/pii/S157082682300015X>. doi:<https://doi.org/10.1016/j.websem.2023.100786>.
- [6] N. Slonim, Y. Bilu, C. Alzate, R. Bar-Haim, B. Bogin, F. Bonin, L. Choshen, E. Cohen-Karlik, L. Dankin, L. Edelstein, et al., An autonomous debating system, Nature 591 (2021) 379–384.
- [7] R. Baumann, M. Ulbricht, Choices and their consequences - explaining acceptable sets in abstract argumentation frameworks, in: M. Bienvenu, G. Lakemeyer, E. Erdem (Eds.), Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning, KR 2021, Online event, November 3–12, 2021, 2021, pp. 110–119. URL: <https://doi.org/10.24963/kr.2021/11>. doi:10.24963/KR.2021/11.
- [8] M. Ulbricht, J. P. Wallner, Strong explanations in abstract argumentation, in: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2–9, 2021, AAAI Press, 2021, pp. 6496–6504. URL: <https://doi.org/10.1609/aaai.v35i7.16805>. doi:10.1609/AAAI.V35I7.16805.
- [9] O. Arieli, A. Borg, M. Hesse, C. Straßer, Explainable logic-based argumentation, in: F. Toni, S. Polberg, R. Booth, M. Caminada, H. Kido (Eds.), Computational Models of Argument - Proceedings of COMMA 2022, Cardiff, Wales, UK, 14–16 September 2022, volume 353 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2022, pp. 32–43. URL: <https://doi.org/10.3233/FAIA220139>. doi:10.3233/FAIA220139.
- [10] T. Kampik, N. Potyka, X. Yin, K. Cyras, F. Toni, Contribution functions for quantitative bipolar argumentation graphs: A principle-based analysis, Int. J. Approx. Reason. 173 (2024) 109255. URL: <https://doi.org/10.1016/j.ijar.2024.109255>. doi:10.1016/J.IJAR.2024.109255.
- [11] T. Kampik, K. Čyras, J. Ruiz Alarcón, Change in quantitative bipolar argumentation: Sufficient, necessary, and counterfactual explanations, International Journal of Approximate Reasoning (2023) 109066. URL: <https://www.sciencedirect.com/science/article/pii/S0888613X23001974>. doi:<https://doi.org/10.1016/j.ijar.2023.109066>.
- [12] F. Cerutti, S. A. Gaggl, M. Thimm, J. P. Wallner, Foundations of implementations for formal argumentation, FLAP 4 (2017). URL: <http://www.collegepublications.co.uk/downloads/ifcolog00017.pdf>.
- [13] M. Persiani, E. Guerrero, A. Brännström, K. Kilic, T. Kampik, Fantastic argumentation tools and where to find them, in: SAFA 2024: The Fifth International Workshop on Systems and Algorithms

- for Formal Argumentation, Hagen, Germany, September 17, 2024, Technical University of Aachen, 2024, pp. 56–68. URL: <https://ceur-ws.org/Vol-3757/paper4.pdf>.
- [14] A. Rago, O. Cocarascu, C. Bechlivanidis, D. Lagnado, F. Toni, Argumentative explanations for interactive recommendations, *Artificial Intelligence* 296 (2021) 103506. doi:<https://doi.org/10.1016/j.artint.2021.103506>.
 - [15] H. Ayoobi, N. Potyka, F. Toni, Argumentative interpretable image classification, in: 2nd International Workshop on Argumentation for eXplainable AI co-located with 10th International Conference on Computational Models of Argument (COMMA 2024), Hagen, Germany, September 16, 2024, volume 3768 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3–15.
 - [16] A. Borg, F. Bex, Explaining arguments at the dutch national police, in: V. Rodríguez-Doncel, M. Palmirani, M. Araszkiwicz, P. Casanovas, U. Pagallo, G. Sartor (Eds.), *AI Approaches to the Complexity of Legal Systems XI-XII - AICOL International Workshops 2018 and 2020: AICOL-XI@JURIX 2018, AICOL-XII@JURIX 2020, XAILA@JURIX 2020, Revised Selected Papers*, volume 13048 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 183–197. URL: https://doi.org/10.1007/978-3-030-89811-3_13. doi:10.1007/978-3-030-89811-3_13.
 - [17] F. Bex, A. Borg, Related explanations in formal argumentation, an empirical study, in: C. Reed, M. Thimm, T. Rienstra (Eds.), *Computational Models of Argument - Proceedings of COMMA 2024*, Hagen, Germany, September 18-20, 2024, volume 388 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2024, pp. 265–276. URL: <https://doi.org/10.3233/FAIA240327>. doi:10.3233/FAIA240327.
 - [18] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence* 267 (2019) 1–38. URL: <https://www.sciencedirect.com/science/article/pii/S0004370218305988>. doi:<https://doi.org/10.1016/j.artint.2018.07.007>.
 - [19] S. Vesic, B. Yun, P. Teovanovic, Graphical representation enhances human compliance with principles for graded argumentation semantics, in: 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022, 2022, pp. 1319–1327. URL: <https://www.ifaamas.org/Proceedings/aamas2022/pdfs/p1319.pdf>. doi:10.5555/3535850.3535997.
 - [20] C. Seyfried, C. Reed, Y. Kamide, Defining argumentative discourse units as clauses: Psycholinguistic evidence, in: C. Reed, M. Thimm, T. Rienstra (Eds.), *Computational Models of Argument - Proceedings of COMMA 2024*, Hagen, Germany, September 18-20, 2024, volume 388 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2024, pp. 277–288. URL: <https://doi.org/10.3233/FAIA240328>. doi:10.3233/FAIA240328.
 - [21] A. Borg, F. Bex, Explaining arguments at the dutch national police, in: *AI Approaches to the Complexity of Legal Systems XI-XII - AICOL International Workshops 2018 and 2020: AICOL-XI@JURIX 2018, AICOL-XII@JURIX 2020, XAILA@JURIX 2020, Revised Selected Papers*, volume 13048, 2020, pp. 183–197. URL: https://doi.org/10.1007/978-3-030-89811-3_13. doi:10.1007/978-3-030-89811-3_13.
 - [22] Y. Gao, F. Toni, Argumentation accelerated reinforcement learning for cooperative multi-agent systems, in: *ECAI 2014 - 21st European Conference on Artificial Intelligence*, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014), 2014, pp. 333–338. URL: <https://doi.org/10.3233/978-1-61499-419-0-333>. doi:10.3233/978-1-61499-419-0-333.