Check for updates

Received: 1 July 2025 | Revised: 1 July 2025 | Accepted: 11 September 2025

DOI: 10.1002/aaai.70033

MEETING REPORT

Explainable AI, energy and critical infrastructure systems

Francesco Leofante[1] | **André Artelt**[2] | **Demetrios Eliades**[3] | **Anna Korre**[4] | **Francesca Toni**[1] | **Tim Miller**[5]

[1]Department of Computing at Imperial College London, Imperial College London, London, UK

[2]Faculty of Technology at Bielefeld University, Bielefeld University, Bielefeld, Germany

[3]KIOS Research and Innovation Center of Excellence of the University of Cyprus, University of Cyprus, Nicosia, Cyprus

[4]Department of Earth Science and Engineering at Imperial College London, Imperial College of London, London, UK

[5]School of Electrical Engineering and Computer Science at The University of Queensland, University of Queensland, Brisbane, Australia

Correspondence
Francesco Leofante, Department of Computing at Imperial College London, London, UK.
Email: f.leofante@imperial.ac.uk

Funding information
Imperial College London; Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen; European Research Council, Grant/Award Numbers: 951424, 101020934; Royal Academy of Engineering; J.P. Morgan

Abstract

The AAAI 2025 Bridge on "Explainable AI, Energy and Critical Infrastructure Systems" was held at the Pennsylvania Convention Centre, Philadelphia, Pennsylvania, USA, on February 25, 2025. The bridge gathered researchers and practitioners, bringing together innovation research across explainable AI, energy and critical infrastructure systems so they can enhance each other. The Bridge featured five keynote presentations by experts, one tutorial, poster presentations by authors who contributed their research findings, and three breakout sessions to discuss new challenges arising at the intersection of these exciting disciplines.

Environmental sustainability challenges are fuelling quests for automated solutions to manage Energy and Critical Infrastructure (E&CI) systems (e.g., power grids, telecommunication networks, transportation systems, water networks) in an economically efficient and environmentally resilient way. Artificial Intelligence (AI) techniques have gained prominence in this area, given their ability to process vast amounts of heterogeneous data under uncertainty. However, most AI techniques used to date lack the transparency that is required to understand the rationale behind their decisions, hindering their applicability in E&CI settings[1]. Furthermore, given the high-stakes nature of E&CI systems, transparency may be an important legal requirement. For instance, the EU AI Act (Regulation 2024/1689) explicitly classifies AI applications in E&CI systems as high risk and (legally) demands transparency that "allows appropriate traceability and explainability"[2]. Explainable AI (XAI) techniques have

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2025 The Author(s). *AI Magazine* published by John Wiley & Sons Ltd on behalf of Association for the Advancement of Artificial Intelligence.

the potential to satisfy these legal requirements and foster the development of supportive AI for energy management systems and critical infrastructure that is efficient and resilient, but also more transparent and socially equitable.

To reflect on how XAI can contribute to tackling challenges surrounding the application of AI to E&CI systems, we organised the Bridge on XAI, Energy and Critical Infrastructure Systems at AAAI 2025. The Bridge brought together researchers and practitioners from different fields to better understand the challenges and requirements of this new emerging area of integrating AI within traditional methods to operate E&CI systems.

Programme details. The event featured five invited talks that explored different facets of XAI's application to energy and critical infrastructure systems, their integration challenges and promises[3]. Pascal Van Hentenryck from Georgia Tech, opened the morning session with an invited talk describing ways in which AI and optimisation techniques can help improve the efficiency of operations in power grids. Additional examples of how AI can boost grid operations were provided by Luis Piloto, Senior Research Scientist at Google DeepMind, in his invited talk.

Ensuring explainability and transparency of AI-assisted decision-making in E&CI emerged as a crucial need during the bridge. Counterfactual explanations featured as a prominent tool in this area, given their ability to produce intuitive characterisations of alternative scenarios where desired outcomes would have been achieved. Kyri Baker, from the University of Colorado Boulder, demonstrated this point in her invited talk, where counterfactuals were discussed as a tool to diagnose and mitigate infeasibility in optimisation problems arising in power grid operations. The need for explainability, interpretability and safety was further discussed in the context of AI and Machine Learning (ML) applications for electric power systems. Venkat Tirupati, the Vice President of DevOps and Grid Transformation at Electric Reliability Council of Texas, and Pryia Donti, from MIT, highlighted the essential role that XAI can play to enhance transparency, foster trust and ensure efficiency in AI/ML-assisted grid operations, including applications such as cybersecurity threat detection, load forecasting and market anomaly detection. In all these scenarios, XAI techniques could provide insights into the factors that influence AI-assisted decision making, ultimately improving the accountability, reliability and compliance of those decisions.

In addition to invited talks, the Bridge featured one tutorial on differentiable programming, data-driven modelling and control, as well as posters and lightning talks exploring how applications of AI and XAI could support E&CI researchers and practitioners in their respective domains.

Finally, three breakout sessions addressed topics related to future challenges for AI in E&CI. The first session focused on explainability challenges that may arise when reinforcement learning (RL) controllers are trained to solve tasks in power grids. Participants noted that very few methods exist to explain RL controllers effectively. This may be a limitation in power grid applications, in which AI decisions need to be communicated to system operators in an intelligible and trustworthy way. The second session focused on the potential of neuro-symbolic AI systems within E&CI. Participants noted that applications in this sector require AI solutions that can ensure correctness, precision, replicability, and explainability of their decisions. As purely subsymbolic approaches may fail to achieve all these goals, this session explored how neuro-symbolic approaches may be better suited to meet the needs of E&CI. Finally, the third session explored the importance of human-centred decision-making processes for high-stakes applications in the energy sector. Participants stressed that E&CI applications require decision-making frameworks that are not only technically sound and resilient, but also aligned with human values and societal goals. Collectively, these three sessions highlighted critical research avenues that could significantly advance the field of XAI for E&CI.

Summary and outlook. Energy and critical infrastructure systems are seeing a wide range of existing and possible AI applications, from large network operations to individual consumer services. Despite this potential, there are a number of barriers that are slowing down the wider adoption and deployment of AI within E&CI systems. Specifically, discussions at the bridge revealed that the lack of explainability surrounding AI-driven decisions may significantly impact stakeholder's confidence and slow down AI uptake within E&CI. This is a notably conservative sector, where the adoption of new techniques, even those promising substantial efficiency improvements, proceeds at a slower pace given the consequential nature of decisions within these applications. Advances in XAI are important to support deployment of AI in E&CI, and the bridge provided an opportunity to explore how XAI could be used to make AI-assisted decision-making in E&CI more transparent.

The community of researchers working at the intersection of XAI, energy and critical infrastructure systems have begun to shape a vision of this important field of research, highlighting challenges and key areas for advancement in XAI. These included a growing recognition of the need for methods that can effectively explain sequential decision-making, accounting for the inherent uncertainty and temporal and physical constraints that characterise complex E&CI systems. Addressing these needs will pave the way for more trustworthy and reliable AI systems, which have the potential to unlock a new wave of innovation in E&CI.

23719621, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aaai.70033 by Imperial College London, Wiley Online Library on [16/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ACKNOWLEDGMENTS

Francesco Leofante was supported by an Imperial College Research Fellowship. André Artelt was supported by the Ministry of Culture and Science NRW (Germany) as part of the Lamarr Fellow Network. Demetrios Eliades was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Synergy grant agreement No. 951424). Francesca Toni was supported by the Royal Academy of Engineering and J.P. Morgan under the Research Chairs and Senior Research Fellowships scheme and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Advanced grant agreement No. 101020934). This publication reflects the views of the authors only.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to report.

ORCID

Francesco Leofante https://orcid.org/0000-0001-8245-9429

ENDNOTES

[1] Energy security and Artificial Intelligence. UK Parliamentary Office for Science and Technology, 2024. DOI: https://doi.org/10.58248/PN735

[2] EU AI Act https://artificialintelligenceact.eu/recital/27/

[3] Bridge resources can be found at the following link: https://www.doc.ic.ac.uk/~fleofant/aaai25-xai-eci/.

How to cite this article: Leofante, F., A. Artelt, D. Eliades, A. Korre, F. Toni, and T. Miller. 2025. "Explainable AI, energy and critical infrastructure systems." *AI Magazine* 46: e70033. https://doi.org/10.1002/aaai.70033

AUTHOR BIOGRAPHIES

Francesco Leofante is an assistant professor within the Centre for Explainable Artificial Intelligence at Imperial College London. His research focuses on trustworthy AI, with a special emphasis on counterfactual explanations and their use in AI. Since 2022, he leads the project "ConTrust: Robust Contrastive Explanations for Deep Neural Networks", a four year effort devoted to the formal study of robustness issues arising in XAI. He is regularly involved as PC member in major AI conferences such as AAAI, IJCAI, KR and ECAI.

André Artelt is a post-doctoral machine learning researcher at Bielefeld University, Germany, and a visiting researcher at the University of Cyprus. His research is centered around XAI in general and, specifically, on contrasting and counterfactual explanations. In the past years, he contributed extensively to the research field of counterfactual explanations, making many theoretical and algorithmic contributions and introducing novel applications of counterfactuals such as water distribution networks. His Python toolbox for creating and simulating complex water distribution network scenarios received the best paper award at the IJCAI-24 workshop on AI for Critical Infrastructure. He is regularly involved as PC member in major AI conferences such as IJCAI, ICML, ICLR, NeurIPS, and ECAI. He also gave a tutorial on counterfactual explanations at IJCAI 2024.

Demetrios Eliades is a research assistant professor at the KIOS Research and Innovation Center of Excellence at the University of Cyprus. His research primarily focuses on Smart Water Networks, including modeling, monitoring, and enhancing security to reduce water losses, safeguard water quality, and improve cyber-physical security. He has spearheaded developing and releasing various research tools for the scientific community, notably the EPANETMATLAB/Python Toolkits. Dr Eliades also serves as the technical coordinator in numerous research and industrial projects, where he applies AI and machine learning to enhance critical infrastructure system operation.

Anna Korre is professor of environmental engineering and Co-Director Energy Futures Lab at Imperial, the College's energy institute. She also heads the Minerals Energy and Environmental Engineering Research Group, developing engineering solutions to the supply of clean energy, the sustainable production of natural resources, and mitigating environmental impacts and risks. Her research focus is in the areas of modelling risk and uncertainty and the environmental and life cycle assessment of engineering systems. She has led and participated in numerous industry, UKRI, BEIS, The Crown Estate and EU funded projects developing engineering tools to assess the impacts of the

23719621, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aaai.70033 by Imperial College London, Wiley Online Library on [16/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

minerals and energy industries in terms of operational performance, environmental footprint and cost.

Francesca Toni is a professor in computational logic and Royal Academy of Engineering/JP Morgan Research Chair on Argumentation-based Interactive Explainable AI (XAI) at the Department of Computing, Imperial College London, UK, as well as the founder and leader of the CLArg (Computational Logic and Argumentation) research group and of the Faculty of Engineering XAI Research Centre. She holds an ERC Advanced grant on Argumentation-based Deep Interactive Explanations. Her current research focus is on explainable AI, machine arguing and computational argumentation, including argumentation in AI broadly conceived and argument mining from text.

Tim Miller is a professor of artificial intelligence in the School of Electrical Engineering and Computer Science at The University of Queensland, Australia. His research draws on machine learning, reinforcement learning, explainable AI, interaction design, and cognitive science, to help people to make better decisions. He has organised the Explainable AI workshop at IJCAI since 2019, and is a regular (senior) program committee members at the top conferences in his field, such as AAAI, IJCAI, NeurIPs, and AAMAS.

23719621, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aaai.70033 by Imperial College London, Wiley Online Library on [16/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License