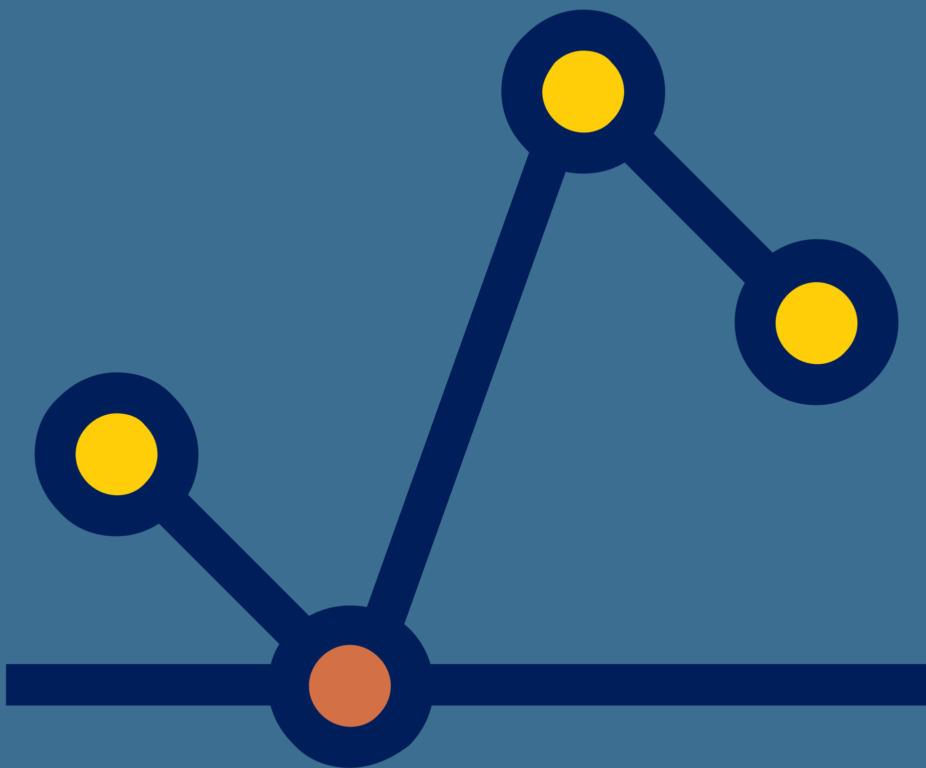
Edited by

Paola Cerchiello · Arianna Agosto
Silvia Osmetti · Alessandro Spelta

Proceedings of the Statistics and Data Science Conference



Pavia University Press

Copertina: Cristina Bernasconi, Milano

Via Salasco, 5 - 20136 Milano
Tel. 02/5836.5751 - Fax 02/5836.5753
egea.edizioni@unibocconi.it - www.egeaeditore.it

Pavia University Press
info@paviauniversitypress.it – www.paviauniversitypress.it

Prima edizione: maggio 2023
ISBN volume 978-88-6952-170-6

Preface

The development of large-scale data analysis and statistical learning methods for data science is gaining more and more interest, not only among statisticians, but also among computer scientists, mathematicians, computational physicists, economists, and, in general, all experts in different fields of knowledge who are interested in extracting insight from data.
Cross-fertilization between the different scientific communities is becoming crucial for progressing and developing new methods and tools in data science.
In this respect, the Statistics & Data Science group of the Italian Statistical Society has organized an international conference held in Pavia on the 27 and 28 of April 2023, attended by over 70 researchers from different scientific fields.
A collection of the presented papers is available in the present Proceedings showing a huge variety of approaches, methods, and data-driven problems, always tackled according to a rigorous and robust scientific paradigm.

The Statistics & Data Science group

Contents

Critical Visual Explanations. On the Use of Example-Based Strategies for Explaining Artificial Intelligence to Laypersons.

Spiegazioni visive critiche. Uso di strategie basate su esempi per spiegare l'intelligenza artificiale ai non addetti ai lavori.

Beatrice Gobbo

Long Abstract. Explainable Artificial Intelligence (XAI) has started to develop both as a field and as a collection of techniques for enabling humans to understand and interpret Artificial Intelligence (AI) mechanisms and decisions that permeate the lives of many human beings today. [16] While much effort was initially put into making AI systems and algorithms intelligible to their developers and creators, over time, interest expanded towards newcomers, domain experts and laypersons [18, 4], opening new challenges and research opportunities. For instance, promoting algorithmic fairness, accountability, trust, ethics, and awareness at all levels has become a significant challenge in terms of governance. [1] As a consequence, expanding audiences and scopes have made room for other disciplines. Indeed, surveys and State-of-the-Art Reports (STARs) on AI Explainability witness the proliferation of scientific contributions increasingly leaning towards multidisciplinary cooperation where efforts in organising and classifying AI explanations have been made considering a large variety of parameters including, for instance, *strategies*, *media* and *audience* [8, 11]. Specifically, *strategies* define how the line of explanation reasoning goes. *Media* and *audiences* define the visual, audio or text apparatus staged for supporting the explanation and its target public. [8] Among the numerous combination that could be found in the literature [11], this paper will discuss and critique from an information design perspective the use of example-based [16] visual explanations [19] for AI addressed to laypersons. While some oversimplification and misunderstanding risks have emerged when addressing with expert users [14], example-based explanations supported by visual means are considered efficient when targeted to laypersons [13, 7, 10, 15].

A study carried on by Cai and colleagues in 2019 aimed at evaluating the effectiveness of comparative and normative visual explanations for a sketch recognition system on a sample of laypersons [2] demonstrates that *normative explanations* are more effective than *comparative explanations*. In other words, it demonstrates that

Beatrice Gobbo
CIM — Centre for Interdisciplinary Methodologies, Warwick University, e-mail: beatrice.gobbo@warwick.ac.uk

the sample of lay users involved in the study, when trying to understand the reason why a sketch recognition system did not recognise their drawings (see the screenshot depicting two examples of Visual Explanation provided by the system in Figure 1), prefer to establish a norm for what drawings look like in the target class (1 , left) instead of understanding the system relying on the comparison between the user's drawing and similar drawings from alternative classes (1 , right).

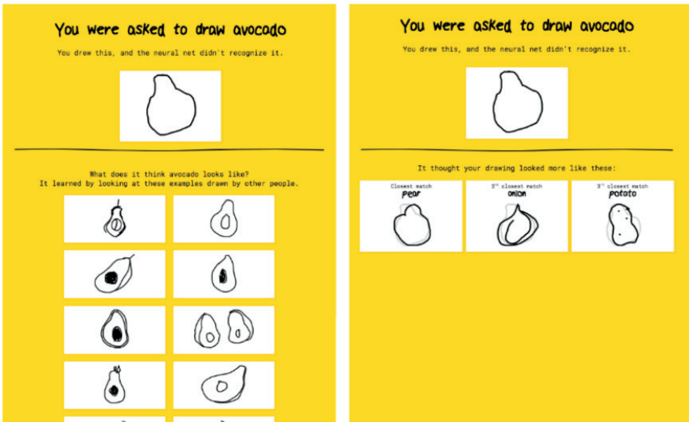


Fig. 1 The image juxtaposes normative (left) and comparative (right) example-based visual explanations. Source: [2] using QuickDraw

Hence, although example-based visual explanations are effective when dealing with laypersons, possible limitations emerging from this case study concern the risk of relying on widespread sample datasets [3] and providing biased explanations of complex, unstable and situated systems such as Artificial Intelligence machines. In this regard, it is worth mentioning that Artificial Intelligence algorithms have been defined as socio-technical *assemblages* [9], and *megamachines* [5] since they are socially, historically and materially entangled with data. Thus, the contribution revolves around the discussion on how to engage laypersons in understanding such complex machines using example-based visual explanations and how the concept of *critical information visualisation* [6] could support their design process. The principles of critical information visualisation are meant to help designers and researchers to formulate questions during the design, use, and study of information visualisations. Similarly, this paper aims to provide researchers with — preliminary — conceptual tools to critically design and stage example-based visual explanations for laypersons. [1].

- *Enrich to counteract the norm* — In line with the idea of *plurality* proposed by Dörk et al. [6], exposing multiple perspectives could enhance the audience's

[1] Note that we do not criticise the explanation of the system *per sé* — if the system is explainable — but the way the examples are displayed while explaining it.

ability to identify with the complex system. For instance, the 'Anatomy of an AI System' [17] provides a ramified visual explanation involving different actors, materials and technologies.

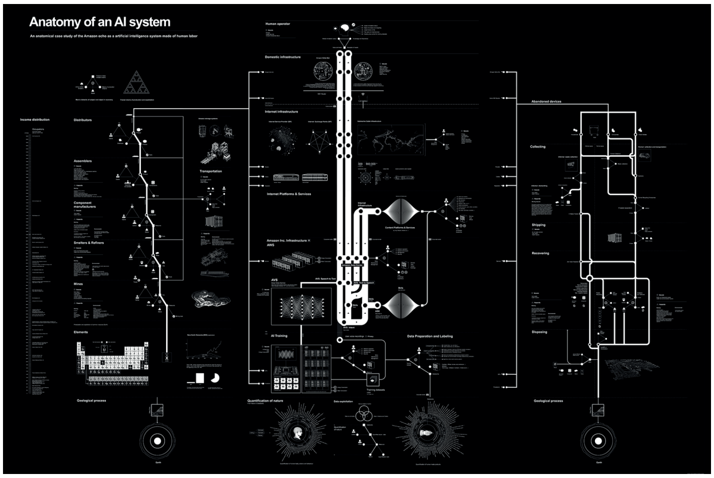


Fig. 2 The Anatomy of an AI System is a project by Share Lab, which explain how Amazon Echo by Alexa works using a multi-perspective range of examples. Source: [17]

- *Dissect to disclose the process* — Drawing on the *disclosure* principle proposed by Dörk et al., being critical in presenting example-based visual explanations implies being able to inform the audience about design decisions, such as the reason why particular examples have been presented instead of others. For instance, Mauri and colleagues present a collection of visual posters explaining a collection of statistical models (including some AI algorithms) to laypersons using examples. There, visual explanation fragments are usually accompanied by texts specifying the reason behind the choice of each example.
- *Simulate to empower the audience* — Simulation is a well-known procedure in computing and statistics. Indeed, many examples of visual explanations simulate the functioning of Artificial Intelligence systems enabling the users to interact with them. Moreover, the simulation of examples was revealed to be convincing when dealing with experts and newcomers [12]. Here, we are promoting simulation to support example-based strategies to empower the audience. Indeed, users could have the opportunity to play with their data or produce data with the algorithm in a real-time scenario (i.e., the "Simplicial Depth Measure" and "Page Rank" examples proposed by Mauri et al. [15]).

In the first part, the author presented example-based visual explanations as effective tools for explaining AI to laypersons. In the second part, the use of examples has been challenged. Finally, by drawing on the principle of *plurality*, *contingency*, *disclosure*, and *empowerment* proposed by Dörk et al., the paper outlined the need to promote a critical approach to the design of example-based visual explanation and denoted an increasing need for collaboration, where computer scientists, statisticians, social scientists and information designers cooperate in co-design settings.

References

1. Ananny, M., and Crawford, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society 20*, 3 (2018), 973–989. _eprint: https://doi.org/10.1177/1461444816676645.
2. Cai, C. J., Jongejan, J., and Holbrook, J. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray California, Mar. 2019), ACM, pp. 258–262.
3. Ciston, S. A critical field guide for working with machine learning datasets.
4. Correll, M. Ethical dimensions of visualization research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, p. 1–13.
5. CRAWFORD, K. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.
6. Dörk, M., Feng, P., Collins, C., and Carpendale, S. Critical InfoVis: exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris France, Apr. 2013), ACM, pp. 2189–2198.
7. Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., and Weisz, J. D. Expanding Explainability: Towards Social Transparency in AI systems. Publisher: arXiv Version Number: 1.
8. El-Assady, M., Jentner, W., Kehlbeck, R., Schlegel, U., Sevastjanova, R., Sperrle, F., Spinner, T., and Keim, D. Towards XAI: Structuring the Processes of Explanations. 13.
9. Gillespie, T. 167The Relevance of Algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press, 02 2014.
10. Gobbo, B. Embalming and dissecting artificial intelligence. visual explanations for the general public.
11. Gobbo, B., Elli, T., Hinrichs, U., and El-Assady, M. xai-primer.com — A Visual Ideation Space of Interactive Explainers. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (New Orleans LA USA, Apr. 2022), ACM, pp. 1–4.
12. IEEVIS. Workshop on visualization for ai explainability.
13. Jin, W., Fan, J., Gromala, D., Pasquier, P., and Hamarneh, G. Euca: A practical prototyping framework towards end-user-centered explainable artificial intelligence. *ArXiv abs/2102.02437* (2021).
14. Kim, B., Khanna, R., and Koyejo, O. O. Examples are not enough, learn to criticize! Criticism for Interpretability. In *Advances in Neural Information Processing Systems* (2016), D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29, Curran Associates, Inc.
15. Mauri, M., Vantini, S., Gobbo, B., Elli, T., Aversa, E., Benedetti, A., Briones Rojas, M. D. L. A., and Colombo, G. Making posters to understand statistics: towards a didactical approach in communication design. In *DRS2022: Bilbao* (2022), pp. 1–18.
16. Molnar, C. *Interpretable machine learning: a guide for making black box models explainable*, second edition ed. Christoph Molnar, Munich, Germany, 2022.
17. ShareLab. Anatomy of an ai system.
18. Suresh, H., Gomez, S. R., Nam, K. K., and Satyanarayan, A. Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan, May 2021), ACM, pp. 1–16.
19. Tufte, E. R. *Visual explanations: images and quantities, evidence and narrative*, twelfth printing, april 2019 ed. Graphics Press, Cheshire, 2019. OCLC: 1126611319.