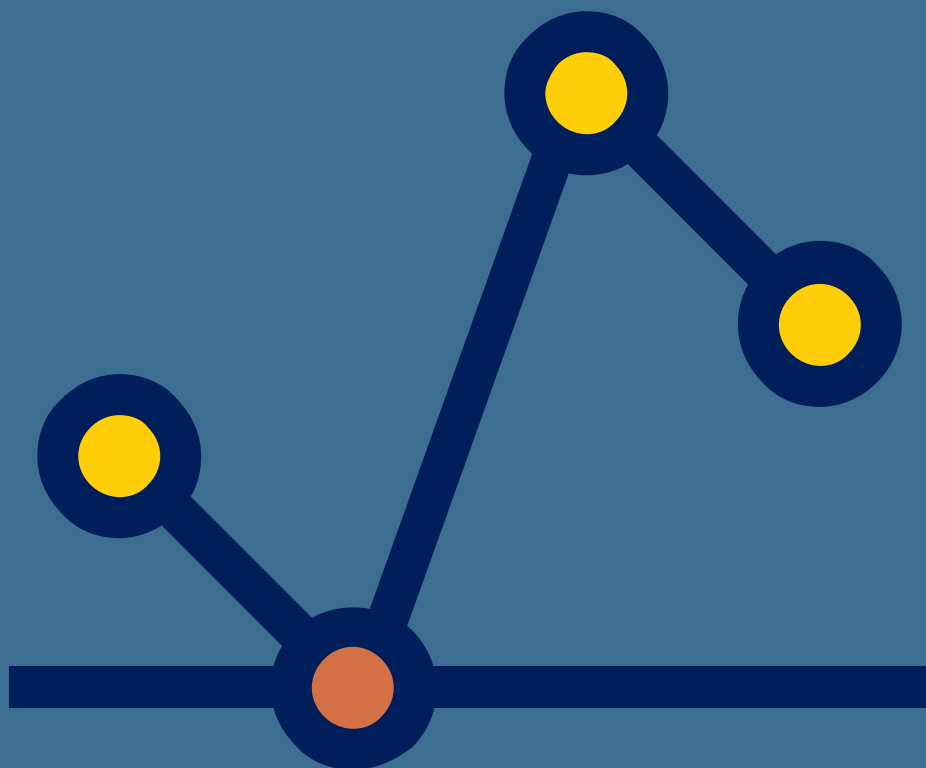


---

Edited by  
Paola Cerchiello · Arianna Agosto  
Silvia Osmetti · Alessandro Spelta

# Proceedings of the Statistics and Data Science Conference



*Copertina:* Cristina Bernasconi, Milano

Copyright © 2023 EGEA S.p.A.  
Via Salasco, 5 - 20136 Milano  
Tel. 02/5836.5751 - Fax 02/5836.5753  
egea.edizioni@unibocconi.it - www.egeaeditore.it

Quest'opera è rilasciata nei termini della Creative Commons Attribution 4.0 International Licence (CC BY-NC-SA 4.0), eccetto dove diversamente indicato, che impone l'attribuzione della paternità dell'opera e ne esclude l'utilizzo a scopi commerciali. Sono consentite le opere derivate purché si applichi una licenza identica all'originale. Il testo completo è disponibile alla pagina web <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.it>.

Date le caratteristiche di Internet, l'Editore non è responsabile per eventuali variazioni di indirizzi e contenuti dei siti Internet menzionati.

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

<b>Fractional random weight bootstrap in presence of asymmetric link functions</b> .....	1
La Rocca Michele, Niglio Marcella, Restaino Marialuisa	
<b>Innovation patterns within a regional economy through consensus community detection on labour market network</b> .....	6
Morea Fabio, De Stefano Domenico	
<b>Sparse Inference in functional conditional Gaussian Graphical Models under Partial Separability</b> .....	12
Fici Rita, Sottile Gianluca , Augugliaro Luigi	
<b>A Conformal Approach to Model Explainability</b> .....	18
Mata Naranjo Juan, Brutti Pierpaolo	
<b>A S.A.F.E. approach for Sustainable, Accurate, Fair and Explainable Machine Learning Models</b> .....	24
Raffinetti Emanuela , Giudici Paolo	
<b>Do we really care about data ethics?</b> .....	30
Ferrara Alfio	
<b>Ethical concepts of data ethics between public and private interests</b> .....	36
Durante Massimo	
<b>Being a statistician in the big data era: A controversial role?</b> .....	42
Manzi Giancarlo	
<b>Forecasting relative humidity using LoRaWAN indicators and autoregressive moving average approaches</b> .....	47
Rojas Guerra Renata, Vizziello Anna, Gamba Paolo	

<b>Interpretability of Machine Learning algorithms: how these techniques can correctly guess the physical laws?</b> .....	53
De Corato Marzio, Ferrara Alfio, Salini Silvia	
<b>The role of BERT in Neural Network sentiment scoring for Time Series Forecast</b> .....	55
Basili Roberto, Croce Danilo, Iezzi Domenica, Monte Roberto	
<b>Diagnostics for topic modelling. The dubious joys of making quantitative decisions in a qualitative environment</b> .....	61
Sciandra Andrea, Trevisani Matilde, Tuzzi Arjuna	
<b>Mapping the thematic structure of Data Science literature with an embedding strategy</b> .....	67
Irpino Antonio, Misuraca Michelangelo, Giordano Giuseppe	
<b>Critical Visual Explanations. On the Use of Example-Based Strategies for Explaining Artificial Intelligence to Laypersons</b> .....	73
Gobbo Beatrice	
<b>Visualising unstructured social media data: a chart-based approach</b> .....	77
Aversa Elena	
<b>From teaching Statistics to designers to teaching Statistics through design</b> .....	85
Mauri Michele, Vantini Simone	
<b>Forecasting Spatio-Temporal Data with Bayesian Neural Networks</b> .....	90
Ravenda Federico, Cesarini Mirko, Peluso Stefano, Mira Antonietta	
<b>Oracle-LSTM: a neural network approach to mixed frequency time series prediction</b> .....	96
Bitetto Alessandro, Cerchiello Paola	
<b>Streamlined Variational Inference for Modeling Italian Educational Data</b>	102
Gioia Di Credico, Claudia Di Caterina, Francesco Santelli	
<b>The use of magnetic resonance images for the detection and classification of brain cancers with D-CNN</b> .....	108
Mascolo Davide, Plini Leonardo, Pecchini Alessandro, Antonicelli Margaret	
<b>Modeling and clustering of traffic flows time series in a flood prone area</b> .	113
Zuccolotto Paola, De Luca Giovanni, Metulini Rodolfo, Carpita Murizio	
<b>Global mobility trends from smartphone app data. The MobMeter dataset</b> .....	119
Finazzi Francesco	

Contents

<b>Spatio-temporal statistical analyses for risk evaluation using big data from mobile phone network</b> . . . . .	124
Perazzini Selene, Metulini Rodolfo, Carpita Maurizio	
<b>A Robust Approach to Profile Monitoring</b> . . . . .	130
Capezza Christian, Centofanti Fabio, Lepore Antonio, Palumbo Biagio	
<b>The FDA contribution to Health Data Science</b> . . . . .	133
Ieva Francesca	
<b>A new topological weighted functional regression model to analyse wireless sensor data</b> . . . . .	139
Romano Elvira, Irpino Antonio, Andrea Diana	
<b>Clustering for rotation-valued functional data</b> . . . . .	145
Stamm Aymeric, Bellanger Lise	
<b>Giudici Paolo InstanceSHAP: An instance-based estimation approach for Shapley values</b> . . . . .	151
Babei Golnoosh, Giudici Paolo	
<b>A new paradigm for Artificial Intelligence based on Group Equivariant Non-Expansive Operators (GENEOs) applied to protein pocket detection</b> . . . . .	152
Bocchi Giovanni, Micheletti Alessandra, Frosini Patrizio, Pedretti Alessandro, Gratteri Carmen, Lughini Filippo, Beccari Andrea Rosario, Talarico Carmine	
<b>Clustering Italian medical texts: a case study on referrals</b> . . . . .	158
Torri Vittorio, Ercolanoni Michele, Bortolan Francesco, Leoni Olivia, Ieva Francesca	
<b>Classification of Recommender systems using Deep Learning based generative models</b> . . . . .	164
Filali-Zegzouti Sanae, Banouar Oumayma, Benslimane Mohamed	
<b>Sparse Inference in Gaussian Graphical Models via Adaptive Non-Convex Penalty Function</b> . . . . .	170
Cuntrera Daniele, Muggeo Vito M.R., Augugliaro Luigi	
<b>Bayesian causal inference from discrete networks</b> . . . . .	177
Castelletti Federico, Consonni Guido	
<b>Sign-Flip tests for Spatial Regression with PDE regularization</b> . . . . .	182
Cavazzutti Michele, Arnone Eleonora, Ferraccioli Federico, Finos Livio, Sangalli Laura M.	
<b>A novel sequential testing procedure for selecting the number of changepoints in segmented regression models</b> . . . . .	187
Priulla Andrea, D'Angelo Nicoletta	

<b>On the numerical stability of the efficient frontier</b> . . . . .	193
Fassino Claudia, Uberti Pierpaolo	
<b>Spatial regression with differential regularization over linear networks</b> . .	196
Clemente Aldo, Arnone Eleonora, Mateu Jorge, Sangalli Laura M.	
<b>An Estimation Tool for Spatio-Temporal Events over Curved Surfaces</b> . . .	201
Panzeri Simone, Begu Blerta, Arnone Eleonora, Sangalli Laura M.	
<b>Gromov-Wasserstein barycenters for optimal portfolio allocation</b> . . . . .	207
Spelta Alessandro, Pecora Nicolò, Maggi Mario	
<b>Online Job Advertisements: toward the quality assessment of classification algorithms for the occupation and the activity sector</b> . . . . .	214
Catanese Elena, Inglese Francesca, Lucarelli Annalisa, Righi Alessandra, Ruocco Giuseppina	
<b>Linear Programming for Wasserstein Barycenters</b> . . . . .	220
Auricchio Gennaro, Bassetti Federico, Gualandi Stefano, Veneroni Marco	
<b>A multi-channel convolution approach for forecast reconciliation</b> . . . . .	224
Marcocchia Andrea, Arima Serena, Brutti Pierpaolo	
<b>Hedging global currency risk with factorial machine learning models</b> . . . .	230
Giudici Paolo, Pagnottoni Paolo, Spelta Alessandro	
<b>Predicting musical genres from Spotify data by statistical machine learning</b> . . . . .	236
Biazzo Federica, Farné Matteo	
<b>The use of Bradley-Terry comparisons in statistical and machine learning models to predict football results</b> . . . . .	242
Macri Demartino Roberto, Torelli Nicola, Egidio Leonardo	
<b>A new approach for quantum phase estimation based algorithms for machine learning</b> . . . . .	248
Ouedrhiri Oumayma, Banouar Oumayma, El Hadaj Salah, Raghay Said	
<b>A comparison of ensemble algorithms for item-weighted Label Ranking</b> .	254
Albano Alessandro, Sciandra Mariangela, Plaia Antonella	
<b>Unsupervised Learning of Option Price in a Controlled Environment: a Neural Network Approach</b> . . . . .	260
Gatta Federico, Schiano Di Cola Vincenzo, Piccialli Francesco, Cuomo Salvatore	
<b>SEMgraph: An R Package for Causal Network Inference of High-Throughput Data with Structural Equation Models</b> . . . . .	266
Grassi Mario, Tarantino Barbara	



Contents

<b>Dynamic models based on stochastic differential equations for biomarkers and treatment adherence in heart failure patients</b> . . . . .	271
Gregorio Caterina, Rares Franco Nicola, Ieva Francesca	
<b>Detecting anomalies in time series categorical data: a conformal prediction approach</b> . . . . .	277
Landrò Matteo, Stamm Aymeric, Vantini Simone	
<b>The structural behavior of Santa Maria del Fiore Dome: an analysis with machine learning techniques</b> . . . . .	282
Masini Stefano, Bacci Silvia, Cipollini Fabrizio, Bertaccini Bruno	
<b>Statistics and Data Science for Arts and Culture: an Application to the City of Brescia</b> . . . . .	288
Ricciardi Riccardo, Carpita Maurizio, Perazzini Selene, Zuccolotto Paola, Manisera Marica	
<b>Detecting Stance in Online Discussions about Vaccines</b> . . . . .	294
Francesco Pierri, Pizzo Fabio, Brambilla Marco	
<b>Towards the specification of a self-exciting point process for modelling crimes in Valencia</b> . . . . .	300
Chiodi Marcello, D'Angelo Nicoletta, Adelfio Giada, Mateu Jorge	
<b>A Clusterwise regression method for distributional data</b> . . . . .	306
Balzanella Antonio, Verde Rosanna, de Carvalho Francisco de A.T.	
<b>Increasing accuracy in classification models for the identification of plant species based on UAV images</b> . . . . .	311
Simonetto Anna, Tariku Girma, Gilioli Gianni	
<b>Travel time to university as determinant on students' performances</b> . . . . .	317
Burzacchi Arianna, Rossi Lidia, Agasisti Tommaso, Paganoni Anna Maria, Vantini Simone	
<b>The FAITH project: integrated tools and methodologies for digital humanities</b> . . . . .	323
Ferrara Alfio, Picascia Sergio, Rocchetti Elisabetta, Varese Gaia	
<b>Assessing the quality of Automatic Passenger Counter data for the analysis of mobility flows of local public transport systems</b> . . . . .	328
Urbano Valeria Maria, Burzacchi Arianna, Cherubini Francesco, Arena Marika, Azzone Giovanni, Secchi Piercesare, Vantini Simone	



# 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].

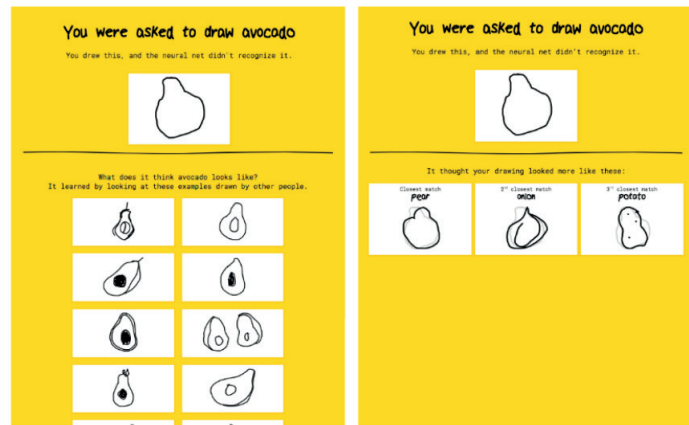
A study carried on by Cai and colleagues in 2019 aimed at evaluating the effectiveness of comparative and normative visual explanations for a sketch recognition system on a sample of laypersons [2] demonstrates that *normative explanations* are more effective than *comparative explanations*. In other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

---

Beatrice Gobbo

CIM — Centre for Interdisciplinary Methodologies, Warwick University, e-mail: [beatrice.gobbo@warwick.ac.uk](mailto: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).



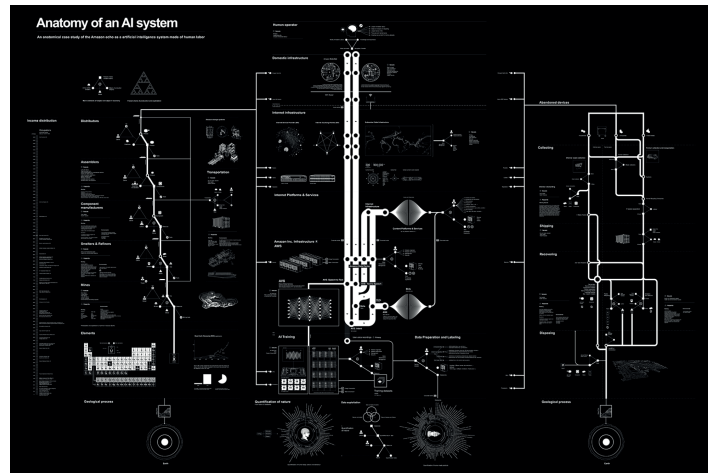
**Fig. 1** The image juxtaposes normative (left) and comparative (right) example-based visual explanations. Source: [2] using QuickDraw

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.<sup>1</sup>

- *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

<sup>1</sup> 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](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., SEVASTIANOVA, 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.