AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

Roadmap on AI technologies & applications for the Media Industry

Section: "Causality and machine learning"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Author	**Filareti Tsalakanidou** (Centre for Research and Technology Hellas – Information Technologies Institute)

This report is part of the deliverable D2.3 - "*AI technologies and applications in media: State of Play, Foresight, and Research Directions*" of the AI4Media project.

You can site this report as follows:

F. Tsalakanidou et al., Deliverable 2.3 - AI technologies and applications in media: State of play, foresight, and research directions, AI4Media Project (Grant Agreement No 951911), 4 March 2022

This report was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media (A European Excellence Centre for Media, Society and Democracy).

Causality and machine learning

Current status

During the last two decades, machine learning algorithms have had enormous success in prediction tasks that require high-dimensional inputs, including computer vision and natural language processing. This success can be attributed to “large-scale pattern recognition on suitably collected independent and identically distributed data”[1]. However, in the real world there’s very little control on data distributions, which may cause for example object detection algorithms to fail in the presence of illumination and pose variations or camera noise, issues that humans can overcome with no or very limited effort due to their inherent ability to generalise and also transfer their knowledge from one setting or domain to another.

The ability to generalise or adapt is only one of the current shortcomings of machine learning. Another limitation is explainability. Currently, ML models are mostly black boxes that produce predictions or recommendations without usually being able to provide a clear explanation as to why they predict A instead of B. This may not be considered much of a problem when an online bookstore recommends a science fiction book instead of a history book but it becomes vital when medical, economic, social or environmental predictions or recommendations are made that have real tangible impact on people’s lives. Attempts for interpretation mainly focus on how the ML model works, not how the world works thus failing to give an answer to the why question. Why one predicted outcome is more possible than another? Why one recommendation may be better for a specific person?

Moreover, what current AI systems are missing is the ability to understand ***cause-and-effect (causal) relationships*** between actions and to consider ***counterfactuals***, i.e. engage in what-if questions about possible past and future chains of events that would lead to different outcomes. This fundamental element of human intelligence, and a key to human evolution, is still missing from AI systems, limiting their ability for independent and autonomous decision-making.

Causality (or cause and effect) is the “relation between two events, one of which is the consequence (or effect) of the other (cause)”[2]. Cognition of causality is fundamental to human development. Humans start to understand such cause-effect relations from infancy, by observing the world around them and learning to predict/expect the consequences of different actions and events. Research has shown that as young as 6 months old, infants are able to “categorically perceive motion events along causal dimensions in addition to spatial and temporal dimensions”[3].

---

[1] B. Schölkopf et al., "Toward Causal Representation Learning," in Proceedings of the IEEE, vol. 109, no. 5, pp. 612-634, May 2021, doi: 10.1109/JPROC.2021.3058954.

[2] Bender, A. (2020) What Is Causal Cognition? *Frontiers in Psychology* 11(3). doi: 10.3389/fpsyg.2020.00003

[3] Muentener, P., and Bonawitz, E. B. (2017). “The development of causal reasoning” in The Oxford handbook of causal reasoning. ed. M. R. Waldmann (New York: Oxford University Press), 677–698.

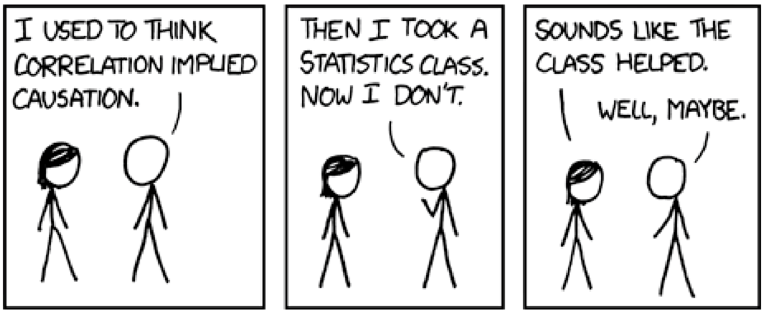

Figure 1: Correlation and causation (comic by Randall Munroe[4]).

Unlike humans, however, machine learning algorithms struggle with causality, failing to understand and determine even basic causal inferences. Take a video of Rafael Nadal and Roger Federer playing tennis as an example[5]. A human can understand that when Nadal's racket clashes with the tennis ball it will cause the ball to change its direction and go back to Federer's side. They are also able to consider counterfactuals, e.g. what would happen if Federer's shot was too low or too high for Nadal to hit the ball and how would that affect the game. Trained with millions of examples, ML algorithms can segment thousands of such video frames in seconds, detect and label different objects like humans, racquets, balls and nets in real-time, provide accurate video summaries or audio-to-text transcriptions but have difficulties in understanding basic causal relationships that a toddler can intuitively comprehend. As explained above, this difficulty is inherited in ML approaches and is a result of the main assumptions and learning tactics adopted by such methods.

Traditional ML approaches are excellent in identifying patterns and associations (correlations in the collected data) and predicting outcomes given a sufficient amount of data. However, to really take AI research to the next level and address the limitations of ML, we need to combine such traditional techniques with causal inference methods so as to be able to move beyond simple correlations to identification of causes thus, being able to answer why something happens and also transfer the acquired knowledge about causes and effects to other relevant domains (Figure 1).

Research challenges

Causality is discussed in J. Pearl's "*The Book of Why: The New Science of Cause and Effect*"[6], which argues about the need to move beyond existing data-centric ML approaches and endow AI with causation capabilities. Pearl proposes the ***Ladder of Causation,*** which constitutes a three-step approach to achieve real AI (Figure 2). The first step is "***seeing***"; here AI can learn from data and ***find associations*** that allows it to make accurate predictions. For example, the editorial board of a newspaper would like to predict which of the candidate guest opinion articles will produce the most reader engagement based on past data of user engagement with

[4] Illustration from XKCD available at https://xkcd.com/552

[5] Example inspired by B. Dickson, Why machine learning struggles with causality (2021): https://bdtechtalks.com/2021/03/15/machine-learning-causality/

[6] Judea Pearl and Dana Mackenzie. 2018. The Book of Why: The New Science of Cause and Effect (1st. ed.). Basic Books, Inc., USA.

their website and social media. This is where the majority of ML algorithms currently operate, offering statistical predictions that are only accurate if the conditions remain the same. In our example, a recommender trained with past article engagement data would have failed to offer an accurate prediction about the interestingness of a medical article on viruses when the pandemic had just begun.

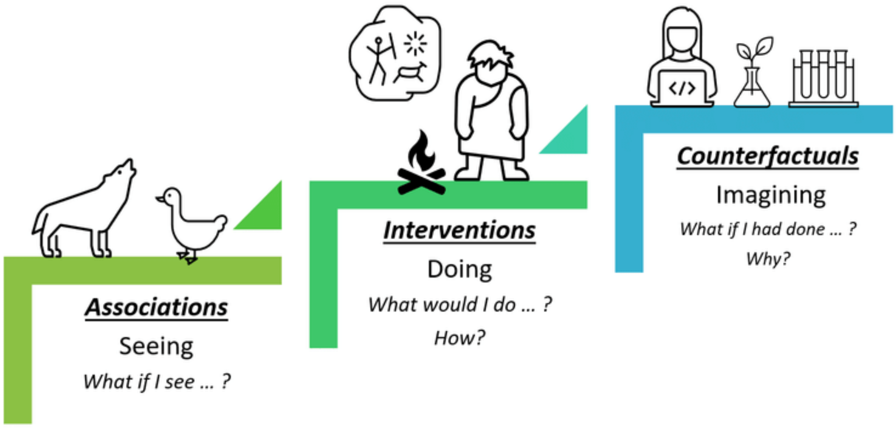

Figure 2: The Ladder of Causation[7]

Second step is "***doing***", which involves ***interventional reasoning***, i.e. trying to predict the outcomes of specific actions (interventions) that change the current conditions. For example, the editorial board considers publishing more articles of younger guest columnists inspired by music and entertainment industry trends – how will this decision affect readership among older audiences; will it attract younger audiences? In terms of machine learning, we are dealing with a distribution shift that changes the underlying conditions and thus the statistical relations between variables. Obviously, this type of prediction requires the development of a causal model that understands the causal relations that affect reader engagement with an article, considering many different variables.

The third step is "***imagining***" and involves ***counterfactual reasoning***, i.e. examining different scenarios of what could have happened if the conditions were different. For example, the editorial board would like to know what the effect on readership would have been, had they not published so many articles critiquing the government before the last election. These kinds of predictions require causal models that depend also on unobserved data. They are critical in AI since they allow to reflect on past decisions and test hypotheses.

Causal AI is still in a nascent state but the field is expected to grow significantly in the next years, aiming to overcome existing ML limitations and deliver more human-like machine intelligence. Unlike machine learning models, causal models describe the causal mechanism of a system based on observed data, a set of variables and assumptions. They are able to incorporate data distribution changes when interventions are applied to the system thus being better in

[7] Image taken form Carey, Alycia & Wu, Xintao. (2022). The Fairness Field Guide: Perspectives from Social and Formal Sciences (source file: https://www.researchgate.net/figure/Pearls-Ladder-of-Causation-The-first-rung-associations-only-allows-predictions-based_fig2_357875366)

generalising than ML models[8]. The main approaches proposed include Causal Bayesian Networks (CBN) and Structural Equation Modelling (SEM)[9]. These causal models comprise of a statistical model and a causal graph that reveals the causal relations between the different variables. To identify the causal relations, two types of search algorithms are proposed[10]: the first exploits conditional independence relations in the data to find a Markov equivalence class of directed causal structures; the second finds a unique causal structure under certain assumptions, considering a noise term that is independent from causes. Newest methods try to exploit deep learning techniques to recognise simple cause-and-effect relationships[11]. Recently, Netflix proposed computational causal inference (CompCI) as an interdisciplinary field across causal inference, algorithms design, and numerical computing that "*addresses engineering needs and human needs for scalability, and directly benefits the deepening relationship between experimentation and personalisation in products and algorithms*"[12]. Recent surveys on causal discovery[13,14] provide a good overview on both background theory and proposed methods.

Identifying causality in real-world settings by exploiting the recent breakthroughs in deep learning and data science is the main challenge of the field. This will hopefully allow efficient understanding of the underlying relations of different variables and thus of how a real-world system, physical or other, operates as a whole. The integration of causal reasoning and inference has the potential to improve the generalisation capabilities of AI and facilitate easy domain adaptation, exploiting the identified causal relations between variables. For example, in the game industry agents currently have to learn to play a game from scratch, even if the game is of the same genre and includes similar strategies. Causal AI can help distill previous knowledge in new games, similarly to a young gamer that uses her previous playing experience but also other life experiences to play a new game. An additional challenge that causal AI inspires to address is explainability of current ML methods, aiming to demystify and provide clear explanations on how decisions are made and actions are driven in AI systems that currently seem like a black box to the average user and in many cases to the experts that have developed them.

Fairness and bias mitigation is another challenge of AI systems where causality can help. According to J. Loftus[15], fairness and causal inference are dual problems since in causal inference, you try to understand the effect of a certain variable while in fairness you try to make a certain variable (like gender or race) not have any effect. To this end, counterfactual fairness

---

[8] J. Ramachandran, Causal Learning – The Next Frontier in the Advancement of AI (2021): https://www.course5i.com/blogs/causal-learning-in-ai/

[9] Peter Spirtes (2010). Introduction to Causal Inference. *Journal of Machine Learning Research* 11(54), 1643–1662.

[10] Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524

[11] Y.Bengio, T. Deleu, N. Rahaman, R. Ke, S. Lachapelle, O. Bilaniuk, A. Goyal, C. Pal (2020). A meta-transfer objective for learning to disentangle causal mechanisms in Int. Conf. on Learning Representations 2020 (ICLR2020), https://openreview.net/pdf?id=ryxWIgBFPS

[12] Jeffrey C. Wong (2020). "Computational Causal Inference", Netflix: https://arxiv.org/abs/2007.10979

[13] Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524

[14] Matthew J. Vowels, Necati Cihan Camgoz and Richard Bowden (2021), D'ya like DAGs? A Survey on Structure Learning and Causal Discovery, https://arxiv.org/abs/2103.02582v2

[15] The Alan Turing Institute, Fairer algorithm-led decisions (2018): https://www.turing.ac.uk/research/impact-stories/fairer-algorithm-led-decisions

in models can be ensured by using causal models.[16] As M. Kusner points out[15], causal models will allow "*to reimagine any individual as a different race or gender, or any different attribute, and make a prediction on that imagined person*." Ensuring that algorithmic outcomes are the same in the real world and a 'counterfactual world' will lead to fairer AI.

According to J. Bengio, one of the pioneers of deep learning, after the great success of ML-based AI what we need now are "*machines that understand the world, that build good world models, that understand cause and effect, and can act in the world to acquire knowledge*"[17,18].

Societal and media industry drivers

Vignette 1: Making successful programming choices in big TV networks

Emilia is the Head of the Entertainment section of a TV network and she has to make a lot of different decisions on a daily basis: select the pilots they will order to series, order new pilots based on submitted scenarios, make decisions about casting of new shows, adjust programming to increase ratings, select which shows will be offered in the network's streaming platform, decide which shows need further promotion and what kind of promotion, envisage ways to beat the competition, etc. Emilia and her team are assisted by an AI TV programming assistant with causal capabilities. The assistant has been trained with heterogeneous historic user engagement data (TV ratings, ratings on sites like Rotten Tomatoes and IMDB, online reviews by TV critics and audience, social media comments, screenings with live audience, surveys, etc.) for the network but also other networks, streaming services, production companies and content creators/ distributors, etc. and integrates causal models from different domains (like psychology, social sciences, advertisement, etc.). Given that the ratings of the prime-time zone on Thursday evenings have been dropping recently, Emilia needs to make quick changes to the schedule. To this end, she asks the AI assistant to determine the possible causes for this drop. The assistant identifies two main causes: the legal drama at the 9:00-10:00 pm slot seems to have caused a lot of online critique for promoting an image of the legal and incarceration system that contradicts the living experiences of many people in minority communities while also failing to embrace diversity in terms of casting. In addition, in a competitor network a new sci-fi series seems to draw the younger audience to that channel.

Emilia is faced with different questions. To beat the competition, should she re-schedule the legal drama to another night and put in its place a high-ratings adventure series with significant following among young audiences that currently plays on Mondays? Will that cause the ratings of the adventure series to temporarily fall? How will it affect advertisers? Or should she urgently put on the Thursday time-slot one of the new shows that were scheduled for later this year? Which one of these shows has the potential to beat the competitors' sci-fi show? Why is the audience drawn to that show? With regard to the critique for the legal drama, would different storylines be enough? Or should they also introduce new cast members? The AI assistant is able

[16] M. Kusner, J. Loftus, C. Russell, R. Silva, Counterfactual fairness (2018) at https://arxiv.org/abs/1703.06856
[17] B. Dickson, System 2 deep learning: The next step toward artificial general intelligence (2019): https://bdtechtalks.com/2019/12/23/yoshua-bengio-neurips-2019-deep-learning/
[18] YouTube video "Yoshua Bengio: From System 1 Deep Learning to System 2 Deep Learning (NeurIPS 2019)": https://www.youtube.com/watch?v=T3sxeTgT4qc&ab_channel=LexClips

to examine these scenarios and estimate the outcome on ratings based on a series of intervention and counterfactual scenarios. Based on the analysis, Emilia decides to promote one of the new shows in the Thursday timeslot, taking her chances with a romantic comedy with sci-fi elements and a diverse cast that seems to appeal to a young female audience. In addition, she decides to give the green light for changes in the cast of the legal drama based on the suggestions offered by the AI assistant about the profile of new cast members and a relevant analysis of how different elements of diversity affect ratings on different audiences.

Vignette 2: Improving newspaper readership and newsroom operation using causal AI for successful editorial choices

Ariadne is the new Editor in Chief of a big newspaper. She was hired because both readership and online engagement of users had dropped considerably while trust in the medium as estimated by a relevant survey had fallen significantly. Ariadne must analyse the reasons why this happened and take measures to reverse the situation. She is assisted by an AI newsroom assistant that provides a full analysis on the main reasons readership has fallen. The report includes interesting findings: the audience has stopped reading the articles of specific political commentators because they consider them partisan; articles about the pandemic attract fewer readers because readers are tired of the newspaper's highly-negative coverage; younger audiences consider the newspaper 'old' because of lack of diversity and different voices; online audiences are not satisfied by the format of the online experience, etc. Ariadne decides that a revamp is in order but first wants to understand how journalistic and editorial decisions were made up to now. The newsroom assistant offers a causal analysis of the factors that influence the journalistic choices of different editors and journalists in the newsroom, offering evidence of influence by political interests and of biases with regard to gender and age. Ariadne considers the evidence provided and makes some first decisions to re-assign journalists to different topics and enforce stricter rules with regard to non-partisan coverage of political news. In addition, she considers hiring a new group of diverse journalists to cover social issues but also entertainment and media, aiming to engage younger readers and adopt a less political profile for the newspaper. Ariadne also hires a graphical designer to redesign the newspaper website aiming to maximise appeal on younger but also older audiences. Using causal inference, the AI newsroom assistant as well as an AI user experience assistant (an AI system that uses causal inference to estimate the impact of platform design changes to user experience and test counterfactual scenarios) estimate how these interventions may affect readership and engagement among different demographics as well as advertisement. After a few months, the decisions made by Ariadne are re-evaluated and a set of counterfactual questions is examined by the AI assistants to estimate what could have been different if different decisions were made.

Future trends for the media sector

Causality can be a game changer in the media sector in several ways. We highlight some of these opportunities with an eye to AI4Media use cases:

- ***Improve personalisation of services*** and ***increase user engagement*** by capturing the causal relations between services, offered content, user interfaces, user profile/ behaviour, global/local trends, advertisement strategies, etc.
- ***Enhance recommender systems*** by a) identifying causal relationships between user attributes, user choices, and media content to estimate the user reaction when exposed to different content, b) mitigating biases through counterfactual testing, c) explaining why specific content was recommended.
- Help the fight against disinformation by ***understanding the mechanisms for disinformation spread and influence on social media*** (including disinformation topic, susceptibility of different groups of people, role of networks of friends, role of politician/political ideology/polarisation, impact of social media and traditional media, format of disinformation content, effect of fact-checking etc.) exploiting existing knowledge from psychology and social sciences[19,20]. In addition, it can improve the trustworthiness of existing deepfake or disinformation detection tools by providing explanations on why some content is considered false or fabricated.
- ***Understand how journalistic and editorial decisions are made*** and which are the factors that influence them the most, by considering personal behaviours in the newsroom, media organisation structure and codes of conduct, political/social/economic national and international environment, influence of advertisers or economic and political interests, public opinion trends, cultural differences, systemic biases etc.[21]
- ***Improve general or targeted advertisement*** and customise campaigns for different products or issues by detecting the main drivers that will convince different groups of customers and estimating the impact of different campaign strategies based on these findings.
- Produce ***automatic summaries of movies, sports or user videos*** that will explain not only who or what we see but how the action or inaction of different characters affects the plot or what relationships the different characters form between them.
- ***Solve the generalisation problem*** of current object detection algorithms thus allowing detection of objects in different settings and under different conditions and improving significantly search efficiency in large audiovisual databases.
- ***Improve and facilitate game design*** by a) allowing gaming experience from one game to be transferred to another, b) integrating causal models for real-world social interactions to the gaming environment to improve human-machine interaction and allow for natural story-telling, c) identifying why different groups or types of players interact in different ways or adopt different strategies, aiming to maximise satisfaction and engagement, d) estimate how adoption of different game rules, strategies or

[19] Lu Cheng, Ruocheng Guo, Kai Shu, Huan Liu, Causal Understanding of Fake News Dissemination on Social Media (2021) at https://arxiv.org/abs/2010.10580
[20] S. T. Smith, E. K. Kao, D. C. Shah, O. Simek and D. B. Rubin, "Influence Estimation on Social Media Networks Using Causal Inference," 2018 IEEE Statistical Signal Processing Workshop (SSP), 2018, pp. 328-332, doi: 10.1109/SSP.2018.8450823
[21] Lukas P. Otto and Isabella Glogger (2020) Expanding the Methodological Toolbox: Factorial Surveys in Journalism Research. *Journalism Studies 21*, 947-965.

capabilities will affect player engagement, e) estimate how different game elements may affect the mental health of users exploiting existing knowledge from psychology.[22]

- ***Remove bias and improve explainability of content moderation systems*** by understanding whether moderation policies are biased against specific groups or beliefs and by providing explanations about why some content is automatically removed. Causal reasoning and counterfactuals can also help refine moderation strategies by estimating the impact of moderation policy changes and assessing past decisions.

Goals for next 10 or 20 years

In the next couple of decades, AI applications will be endowed with causal and inference capabilities, while repositories of causal models will be available, modelling everyday phenomena but also expert concepts thus allowing the development of new AI systems with efficient understanding of causes and effects in new domains. AI systems will be able to discover new causal models and to verify and improve them based on ML and collected evidence from active experimentation in a new media domain.[23,24]

[22] Katarina Gyllenbäck, Putting into play – A model of causal cognition on game design (2018): https://katarinagyllenback.com/2018/12/17/putting-into-play/

[23] K. Hartnett, To Build Truly Intelligent Machines, Teach Them Cause and Effect (2018): https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/

[24] Y. Gil and B. Selman. A 20-Year Community Roadmap for Artificial Intelligence Research in the US. Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI). Released August 6, 2019. arXiv:1908.02624 https://cra.org/ccc/resources/workshop-reports/

ARTIFICIAL INTELLIGENCE FOR
THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu