Bias, journalistic endeavours, and the risks of artificial intelligence

Leiser, M.R.; Pihlajarinne, T.; Alén-Savikko, A.

Bias, journalistic endeavours, and the risks of artificial intelligence

By Dr M. R. Leiser[1]

Abstract:

Artificial intelligence is increasingly used throughout all processes of the news cycle. AI also has untapped corrective potential. By learning to point readers to diverse, quality, and/or legitimate news after exposure to 'fake news', 'false narratives', and disinformation, AI plays a powerful role in cleaning up the information ecosystem. Yet AI systems often 'learn' from training data that contains historical inaccuracies and biases, with results proven to embed discriminatory attitudes and behaviours. Because this training data often does not contain personal information, regulation of AI in the news production cycle is largely overlooked by legal commentators. Accordingly, this chapter lays out the risks and challenges that AI poses in both journalistic content creation and moderation, especially through machine-learning in the post-truth world. It also assesses the media's rights and responsibilities for using AI in journalistic endeavours in light of the EU's AI draft regulation legislative process.

Keywords: artificial intelligence, machine-learning, automated journalism, fake news, explainability, best-efforts accuracy

Introduction

We are living in a world of fast-developing technologies enabled by machine-learning systems that analyse structured data to infer the probability of an outcome. To address the legal, ethical, and societal challenges associated with allegations and concerns that machines will soon be able to duplicate and replicate the human mind, artificial intelligence (AI) is presently subjected to significant attention from eager legal researchers and policy makers.[2] Of course, the promoters of AI and its associated uses have a dirty little secret – there is no such thing as 'artificial intelligence'. Rather, AI describes a series of technologies used by humans to do the heavy lifting required in the era of big datasets and computational analytics. Its potential ranges from identifying correlations in datasets indiscernible to the human mind to increasing the efficiency of production across a range of industrial applications.

Increasingly, AI is used in the production of news and other journalistic endeavours.[3] Not only are machine-learning systems perceived to be replacing humans in the 'creative' process, but AI also enables journalists by personalizing, recommending, fact-checking, labelling, and translating vast arrays of user-generated and viral content. The deployment of AI systems for the purpose of personalizing content provides a way for platforms to make recommendations that the platform believes might be beneficial to the user. However, commingled within 'personalization' are recommender systems that are presented by platforms as a means of helping users identify content that they might find appealing, but in reality boost advertising revenue, and can lead to filter bubbles that reinforce narrow or inaccurate viewpoints. As a result of vast amounts of collected data and complex algorithmic judgements, certain recommendations can effectively reinforce discriminatory beliefs (direct/indirect as well as discrimination by association, and discrimination by perception) and/or encourage harassment and victimization.

AI also has a role to play in investigative journalism to 'extract references to real-world entities, like corporations and people, and start looking for relationships between them, essentially building up context around each entity' in big datasets such as the Panama Papers (13.5 million documents) or the ICIJ's data set of 2.5 million documents relating to the offshore holdings and accounts of over 100,000 entities across

[1] Dr M R Leiser is Assistant Professor of Law and Digital Technologies at Leiden University in The Netherlands.
[2] Rosario Girasa. "Artificial Intelligence as a Disruptive Technology", Springer Science and Business Media LLC, 2020; See also Riya Sil, Abhishek Roy, Bharat Bhushan, A.K. Mazumdar. "Artificial Intelligence and Machine Learning based Legal Application: The State-of-the-Art and Future Research Trends", 2019 International Conference on Computing, Communication, and Intelligent Systems
[3] For a survey of AI use across European Newsrooms, see Fanta, A (2017). Putting Europe's robots on the map: Automated journalism in news agencies. Reuters Institute Fellowship Paper, 9.

four large databases.[4] Its promise lies in its capabilities to do the heavy-lifting and analytics of huge amounts of data, freeing journalists to undertake more probing and investigative reporting. AI systems have both an important creative potential and an editorial function in 21st century journalism. AI can find patterns or flag outlier events for further investigation. The promise of AI also plays an important role in identifying misleading content but can direct people to alternative and better sources of information.

With AI becoming instrumental to our information ecosystem, regulators have undertaken to address an identified normative concern associated with machine-learning systems: both implicit and explicit biases in the training data can lead to discriminatory effects. Accordingly, concepts like 'fairness in machine learning', 'accountability', 'algorithmic transparency' as well as 'explainability' alongside various approaches to 'ethical AI' have preoccupied academia, civil society, and policymakers.[5] Significant critique has also been undertaken of AI's role in journalistic creation, moderation, and fact-checking disinformation.[6] This chapter attempts to bridge the gap between the regulatory environment for artificial intelligence in the European Union, on the one side, and its use in journalism on the other. The first section explains how machine-learning systems work. The second section identifies vulnerabilities in the deployment of AI in the newsroom. The third section examines the EU's response to the 'rise and risks of AI', before concluding with a discussion of the responsibilities for newsrooms that deploy automated journalism, fact-checking, content creation, and other forms of AI-generated news.

Understanding AI and machine learning systems

AI systems are defined as "software (and possibly also hardware) systems…, that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal."[7] Programmers have developed AI systems using natural language processing to write simple articles such as sports or stock market reports. GPT-3, OpenAI's language generator, can generate endless texts in response to input parameters provided by humans;[8] for example, Tencent's "Dreamwriter" can write an article in 0.5 seconds and up three hundred thousand articles a year.[9] Smartphone applications such as Prisma provide their

[4] Can Artificial Intelligence Like IBM's Watson Do Investigative Journalism?, Fast Company, 12 November 2013, available at https://www.fastcompany.com/3021545/can-artificial-intelligence-like-ibms-watson-do-investigative-journalism, accessed 25 February 2021.

[5] Burrell J, 'How the Machine "thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society http://bds.sagepub.com/lookup/doi/10.1177/2053951715622512; Butterworth M, 'The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework' Computer Law & Security Review; https://www.sciencedirect.com/science/article/pii/S026736491830044X>; Datta A, MC Tschantz and A Datta, 'Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination' (2015) 2015 Proceedings on Privacy Enhancing Technologies 92; Diakopoulos N, 'Algorithmic Accountability: On the Investigation of Black Boxes' (New York: Tow Center for Digital Journalism, Columbia University, 2014) http://towcenter.org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/; Diver L and B Schafer, 'Opening the Black Box: Petri Nets and Privacy by Design' (2017) 31 International Review of Law, Computers & Technology 68 https://doi-org.ezproxy.is.ed.ac.uk/10.1080/13600869.2017.1275123; Doshi-Velez F et al., 'Accountability of AI Under the Law: The Role of Explanation' [2017] arXiv:1711.01134 http://arxiv.org/abs/1711.01134; Gillespie T, 'The Relevance of Algorithms' (2014) 167 Media technologies: Essays on communication, materiality, and society; Kendall G and G Wickham, Using Foucault's Methods (London ; Thousand Oaks, Calif: Sage Publications, 1999); Kroll JA et al., 'Accountable Algorithms' (Rochester, NY: Social Science Research Network, 2016) SSRN Scholarly Paper ID 2765268; http://papers.ssrn.com/abstract=2765268>; Pasquale F, The Black Box Society: The Secret Algorithms That Control Money and Information (Cambridge: Harvard University Press, 2015); Selbst AD and J Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233; Wachter S, B Mittelstadt and L Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76

[6] See, eg, Broussard, Meredith, Nicholas Diakopoulos, Andrea L. Guzman, Rediet Abebe, Michel Dupagne, and Ching-Hua Chuan. "Artificial intelligence and journalism." Journalism & Mass Communication Quarterly 96, no. 3 (2019): 673-695; Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press; McStay, A. (2018). Emotional AI: The rise of empathic media. Sage; Hansen, M, Roca-Sales, M, Keegan, J M, & King, G (2017). Artificial intelligence: Practice and implications for journalism; On a wide range of aspects of journalistic endeavours, see Marconi, F (2020). Newsmakers: Artificial Intelligence and the Future of Journalism. Columbia University Press.

[7] High-Level Expert Group on Artificial Intelligence, High-Level Expert Group on Artificial Intelligence, available at https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines, at 6, accessed 16 February 2021.

[8] For an example of this, see 'A robot wrote this entire article. Are you scared yet, human?', The Guardian, 8 September 2020, available at https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3, accessed 16 February 2021.

[9] For an introductory video of "Dreamwriter", available at https://v.qq.com/x/page/z0713387ge88.html, accessed 16 February 2021.

users with an AI-based digital lens to change or ameliorate their photos, while ZAO offers their users an opportunity for creating short deepfake videos with their own faces.[10] Google's AI poetry uses neural network techniques to train its AI system;[11] it then uses an autoencoder to write full sentences.[12] Furthermore, generative Adversarial Networks (GANs) facilitate projects such as fake picture generators or Grover, a fake news generator.[13]

Generative AI systems can be divided into two categories: fully autonomous generative AI and co-creative generative AI. The latter is rarer as it involves real-time human-AI interaction during the process. The user and the AI system generate outputs in response to inputs provided by the other party. In traditional decision or prediction systems, outputs are based on a 'handcrafted' model: data is pumped into a handcrafted model of pre-determined algorithms and pre-set parameters. As Leiser and Dechesne state, "handcrafted systems are those that answer questions directed at classifying items (i.e., predicting discrete values), or predicting continuous values (such as risks, price development, etc.). Humans are left to interpret the outcomes".[14] However, machine-learning models operate on a set of pure correlations without "explicit pointers" for humans to interpret. Deep learning is a type of machine learning that uses deep neural networks to train a computer to perform human-like tasks, like recognizing speech or images. Machine-learning systems turn this 'training data' into a model that can predict or classify new data on the basis of patterns distilled from the training data. Private traits and attributes are predictable from the digital records of human behaviour. However, training data does not get stored in the model. This is mostly done by machine-learning algorithms, where the algorithms reconstruct relationships and dependencies between the characteristics in the training data and the target output. The resulting model then contains a 'logic' of the dependency of the output on the input for the given task, which it has derived from the training data. Machine-learning systems will develop capabilities without any way of reverse engineering the data from which the system learned, nor is it possible to fully understand the logic inside the 'black box'.[15]

Most regulatory attention in this space concerns lack of transparency about the logic used in AI systems either trained on personal data or for the purposes of decision-making that has an impact on individuals. For example, in its final report on 'Disinformation and fake news', the UK's Department of Culture, Media and Sport specifically called for the extension of protections of privacy law "to include models used to make inferences about an individual".[16] Data Protection Regulators have also issued guidelines about the obligation to provide meaningful information about the logic involved in automated decisions.[17] With

[10] For Prisma, see Vlad Savov, Prisma will make you fall in love with photo filters all over again, The Verge, 19 July 2016, available at https://www.theverge.com/2016/7/19/12222112/prisma-art-photo-app. For ZAO, see Zak Doffman, Chinese Deepfake App ZAO Goes Viral, Privacy of Millions 'At Risk', Forbes, available at https://www.forbes.com/sites/zakdoffman/2019/09/02/chinese-best-ever-deepfake-app-zao-sparks-hugefaceapp-such as-privacy-storm/, accessed 16 February 2021.

[11] Deep learning is a subset of machine-learning. It uses deep neural networks, deep belief networks, recurrent neural networks and/or convolutional neural networks for machine-learning processes: it uses these architectures to model its predictive computational statistics; See, G Ras, M Gerven, & W Haselager, Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges, ArXiv 1803.07517 (2018).

[12] M Burgess, Google's AI has written some amazingly mournful poetry, WIRED, 16 May 2016, available at https://www.wired.co.uk/article/google-artificial-intelligence-poetry, accessed 16 February 2021.

[13] T Karras, et al, Analyzing and improving the image quality of stylegan, arXiv preprint arXiv:1912.04958 (2019), demonstrations available at https://thispersondoesnotexist.com. For Grover, see also rowanz, Code for Defending Against Neural Fake News, available at https://github.com/rowanz/grover, demonstrations available at https://thisarticledoesnotexist.com.

[14] Leiser, M R, & Dechesne, F (2020). Governing machine-learning models: challenging the personal data presumption. International Data Privacy Law, 10(3), 187-200.

[15] "Black box(es)" is a semi-colloquial term used to describe opaque machine-learning models, which are traditionally, although need not be, deep-learning based; See Pasquale, F (2015). *The black box society*. Harvard University Press.

[16] Department of Culture, Media and Sport, 'Disinformation and "Fake News": Final Report', Eighth Report of Session 2017–19, 18 February 2019 at para 48, available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179105.htm#_idTextAn%20chor005, accessed 17 February 2021; See also Wachter, S, & Mittelstadt, B (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. Colum. Bus. L. Rev., 494 and Edwards, L, & Veale, M (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. Duke L. & Tech. Rev., 16, 18; Kaminski, Margot E "The right to explanation, explained." Berkeley Tech. LJ 34 (2019): 18

[17] Information Commissioner's Office, and the Alan Turing Institute, "Explaining decisions made with AI", available at https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf, accessed 17 February 2021; see also Article 29 Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, available at https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=49826, accessed 17 February 2021.

concerns that machines are dehumanizing decision-making, both profiling and general automated decision-making *about humans* can only take place when robust legal protections are in place, the principles of data protection are adhered to, and data-subject rights can be upheld.[18] These issues manifest themselves in the general belief that AI challenges the set of legal guarantees put in place in Europe to combat discrimination and ensure equal treatment.[19]

These forms of machine-learning models also play an important role in modern data-driven journalism. AI systems, trained for the purpose of news creation, can search for independent input, and with zero or limited human intervention. These systems can also operate without processing any personal data[20] – the caveat that activates the European Union's data protection regime.[21] An AI system that analyses crime data for hotspots, for example, would not fall under the remit of the GDPR, unless the data subject is identifiable.[22] Understandably, much of the work in this area has focused on historical biases that are embedded in the very training data that machine-learning systems are built.[23] For example, 'predictive policing' is sold to financially challenged law enforcement agencies (LEAs) as a 'neutral' method to counteract unconscious biases, yet increasingly deploy data mining techniques to predict, prevent, and investigate crime.[24] However, research indicates that predictive policing can adversely impact minority and vulnerable communities. For example, using historical data to assist in deployment can lead to more arrests for nuisance crimes in neighbourhoods primarily populated by people of colour. Algorithms employed to help determine criminal sentences in the USA inadvertently discriminated against African Americans.[25] Not only does historical data risk discriminatory effects, but data integrity, too. Dörr and Hollnbuchner posit that missing items can lead to bias in content generation.[26] These effects are an artefact of the specific technology and will take place regardless of any measures implemented to mitigate the machine's bias.[27]

AI in the newsroom

The growing datafication and algorithmicizing of society, the emergence of the platform economy, the mediatization of everyday life, and growing adoption of machine-learning and AI-powered tools and services are transforming both newsrooms and media services. It is not unreasonable to foresee that the practice of journalism would join numerous other disciplines characterized by the ubiquity of interconnected intelligent systems with autonomous capacities. *The New York Times* (NYT) 'Editor' project

[18] General Data Protection Regulation, Arts 13-21.
[19] Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law, available at https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1, accessed 11 June 2021.
[20] GDPR, Art 4(1).
[21] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance); Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA; Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).
[22] This is not without considerable controversy. Some have argued that all information could theoretically relate to an individual – see Purtova, N (2018). The law of everything. Broad concept of personal data and future of EU data protection law. Law, Innovation and Technology, 10(1), 40-81; However, the very broad concept of personal data could make the entire data protection regime unmanageable – see Koops, B J (2014). The trouble with European data protection law. International data privacy law, 4(4), 250-261.
[23] Mehrabi, N, Morstatter, F, Saxena, N, Lerman, K, & Galstyan, A (2019). A survey on bias and fairness in machine learning. arXiv preprint arXiv:1908.09635.
[24] The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm was used to predict the risk ratings of offenders on a scale from 1-10, with the latter being the highest risk. If the algorithm predicted a lower score, this helped judges decide whether offenders could go on parole or probation: see F. Zuiderveen Borgesius, Discrimination, Artificial Intelligence, and Algorithmic Decision-making, 2018 Strasbourg: Council of Europe, Directorate General of Democracy, at 15 and J Angwin et al, Machine bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. 2016 ProPublica.
[25] J Angwin et al, Machine Bias, ProPublica, 23 May 2015, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, accessed 23 February 2021.
[26] Konstantin Nicholas Dörr & Katharina Hollnbuchner (2017) Ethical Challenges of Algorithmic Journalism, Digital Journalism, 5:4, 404-419, page 9.
[27] Selbst, Andrew D, Disparate Impact in Big Data Policing (February 25, 2017). 52 Georgia Law Review 109 (2017)

applies tags to traditionally written news articles.[28] *The Washington Post* (WP) covers financial news and local sports events via various forms of 'automated journalism',[29] a broad term used to describe the use of AI, i.e., software or algorithms, to automatically generate news stories with no contribution by human beings, apart from that of the programmers who developed the algorithm. It covers algorithmic, automated, and robot journalism, and bots that write news. An AI algorithm independently collects and analyses data and then writes a news article.

Because it reduces costs and could broaden its audience and increase its market share, the WP incorporated AI to cover and write simple local stories.[30] *Associated Press*, *Forbes*, *The Los Angeles Times*, and *ProPublica* all use various forms of automated journalism.[31] Automated journalism is based on natural language generation (NLG) technology, which generally permits creation of text-based journalism from a dataset consisting of digitally structured data: "Early examples of the use of NLG technology to automate journalism are mostly confined to relatively short texts in limited domains but are nonetheless impressive in terms of both quality and quantity. The text produced is generally indistinguishable from a text written by human writers and the number of text documents generated substantially exceeds what is possible from manual editorial processes"[32]

Of course, crime is a favourite subject of the news media, with crime stories estimated to make up between 12.5 and 40 per cent of local news.[33] Chermak's analysis of six print and three broadcast media organizations revealed that "print media present nine crime stories a day, on average, and electronic media four crime stories per day".[34] More recently, Curiel et al's social media analysis revealed that an astounding 15 out of every 1000 tweets were about crime or *fear* of crime.[35] Despite social media suffering from a strong bias towards violent or sexual crimes, little correlation exists between social media messages and crime. Social media is not useful for detecting trends in crime but demonstrate insight into the amounts of *fear* about crime.

Given the cost effectiveness and processing capacity of modern computing, machine-learning is being rapidly deployed across newsrooms.[36] Yet AI systems used in newsrooms are often trained on crime reports. In AI discourse, and especially machine-learning, "models" are arrived at. The first step is inputting (relevant) data into the machine, the data inputted largely depending on what the machine would be used for or be doing ultimately. Secondly, the machine identifies the relevant patterns, dots, differences, and especially similarities in the data inputted to it. The third stage is model creation. This is based on steps 1 and 2 – basically, the machine develops a model that can be used for a task when data similar to step 1 is inputted, and so on. Machine-learning's predictive analytics is used to analyse historical and 'real-time' data to make predictive decisions in, not only news reporting, but fact-checking the authenticity of an unverifiable news story. As discussed in the previous section, the accuracy of these predictive decisions increases with the amount of data processed, including the training data upon which the AI system is modelled.

---

[28] Washington Post PR Blog, The Washington Post experiments with automated storytelling to help power 2016 Rio Olympics coverage, available at https://www.washingtonpost.com/pr/wp/2016/08/05/the-washington-post-experiments-with-automated-storytelling-to-help-power-2016-rio-olympics-coverage/, accessed 23 March 2021.

[29] Nicole Martin, Forbes, (Feb 8, 2019), "Did A Robot Write This? How AI Is Impacting Journalism", available at https://www.forbes.com/sites/nicolemartin1/2019/02/08/did-a-robot-write-this-how-ai-is-impacting-journalism/?sh=563c1e779575, accessed 23 March 2021; see Keohane, Joe. 2017. "What news-writing bots mean for the future of journalism" Wired. February. https://www.wired.com/2017/02/robots-wrote-this-story/.

[30] Lucia Moses, DigiDay, (17 Sept 2017), "The Washington Post's robot reporter has published 850 articles in the past year", available at https://digiday.com/media/washington-posts-robot-reporter-published-500-articles-last-year/, accessed 23 March 2021.

[31] A Graefe, Guide to Automated Journalism, in Columbia University Academic Commons, 2016.

[32] DCaswell, K Dörr, Automated Journalism 2.0: Event-driven narratives, in Journalism Practice, 2017, p. 2

[33] Grosholz, J, & Kubrin, C (2007). Crime in the news: How crimes, offenders and victims are portrayed in the media. Journal of Criminal Justice and Popular Culture, 14, 59-83.

[34] Chermak, S (1994) "Crime in the News Media: A Refined Understanding of How Crimes Become News", 95-129 in Media, Process, and the Social Construction of Crime: Studies in Newsmaking Criminology, edited by G Barak. New York: Garland Publishing at 711.

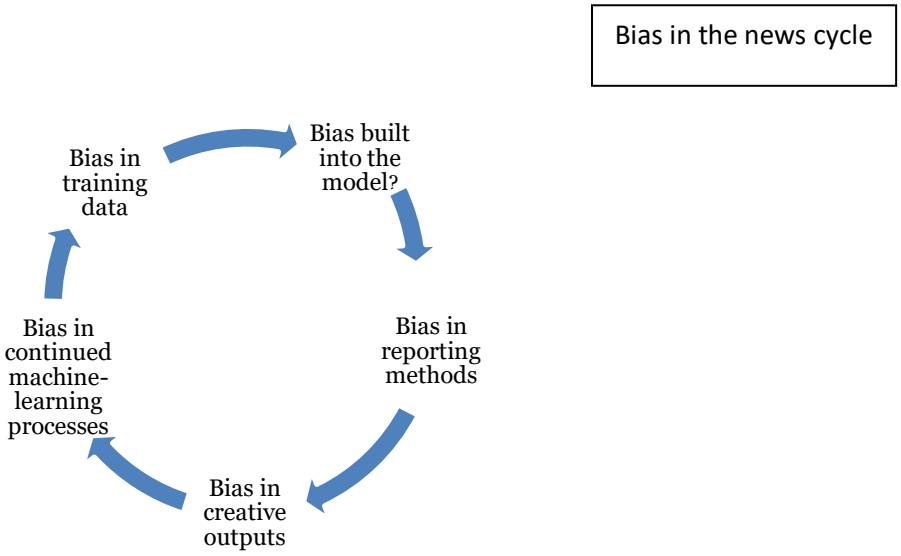
[35] Prieto Curiel, R, Cresci, S, Muntean, CI et al Crime and its fear in social media. Palgrave Commun 6, 57 (2020). https://doi.org/10.1057/s41599-020-0430-7

[36] Automated Journalism – AI Applications at New York Times, Reuters, and Other Media Giants, available at https://emerj.com/ai-sector-overviews/automated-journalism-applications/, accessed on 23 February 2021.

AI is already used to facilitate automated reporting of murders and other forms of violent crime. For example, an AI system retrieves homicide data directly from a coroner's office, which in turn generates leads for reporters to expand with details about the victim's life and family.[37] AI could be used to match details about the deceased's life with details from social media and public registries. Incredibly, this has been touted as an example of automated journalism operating *without bias*,[38] ignoring the numerous instances that data could contain errors and the bias in decision-making by the journalist that chose what to report from the available data, nor that any errors in reporting could appear in the training data of other AI systems designed to search for patterns – for example, crime trends.

The outcomes of any machine-learning system will be trained on a data set for the purpose of creating a new output correlating with the set of inputs on its own with zero or limited human intervention. Automated journalism operates by either independently writing and publishing news articles without input from a journalist or by 'cooperating' with a journalist who can be deputized to supervise the process or provide input to improve the article. All methods are dependent upon access to, and availability of the structured data needed to generate news articles. Thus, any simple error in the coroner's reporting could theoretically infect the entire cycle of *news reporting*, undermining the integrity of the system. Worryingly, the use of predictive policing by LEAs could facilitate further unconscious bias in news reporting, which in turn would affect real-world policing, which in turn would affect the outcomes of predictive policing. The general advantages of this method are the speed with which data can be collected and articles can be written, fewer errors in output, and cost savings. Yet the quality of automated journalism depends on the training data. However, not only is perfect training data never possible, human error, prejudice, and misjudgement can enter into the journalistic lifecycle at multiple points. Consequently, biases are introduced at any point in the news delivery process, from the preliminary stages of data extraction, collection, and pre-processing to the critical phases of news formulation, model building, and reporting.

Figure One: Therein lies the challenge in preventing bias in the newsroom



[37] N Lemelshtrich Latar, The Robot Journalist in the Age of Social Physics: The End of Human Journalism? in G Einav (ed), The New World of Transitioned Media, New York, 2015, 74.
[38] Monti, M (2019). Automated journalism and freedom of information: Ethical and juridical problems related to AI in the press field. Opinio Juris in Comparatione, 1, 2018.

AI in fact-checking

There are growing efforts by journalists, policymakers, and technology companies towards finding effective, scalable responses to online disinformation and false information. Whether by design or coincidence, false online content appears to exploit a specific conjunction of technological and psychological factors. In a content analysis of 150 fake and real news items, fake news titles were found to be substantially more negative in tone than real news titles.[39] The furore over 'fake news' has exacerbated long-standing concerns about political dishonesty, harmful conspiracy theories, malicious rumours, and deceptive campaigns; for example, online conspiracies about Covid-19 spores emanating from 5-G masts have caused real-world arson attacks,[40] while deceptive campaigns about election integrity incited an insurrection at the US Capitol.[41] Online disinformation is generally understood as intentional dissemination of false and misleading information via the Internet so as to mislead its recipients for politically and financially motivated reasons. Terms like "fake news", "disinformation" and "misinformation" are frequently conflated in the surrounding discourse, but the emerging preference is for using the term "disinformation" as a descriptor. To avoid infringements of collateral rights, the European Union has made a conscious decision to refrain from implementing legislative remedies for online disinformation until such time as self-regulation has been conclusively proven ineffective. Non-state responses within the European Union stem primarily from measures like credibility labels, transparency in political advertisements, restrictions on artificial amplification of engagement statistics, and media literacy initiatives implemented by online platform providers. Other non-state responses include fact-checking and other trust-building initiatives from traditional media houses, as well as continued research and awareness campaigns by civil society organizations.

In 2006, Facebook and Instagram started working together with third-party fact-checking organizations and individuals (3PFC) in many countries to ensure that content uploaded by users on the platform is truthful and to avoid dissemination of disinformation. Fact-checkers have to be certified by the International Fact-Checking Network (IFCN). They can mark content as "true", "partly true", "false", "partly false", "false title", "not applicable for evaluation", "satire", "hoax", "opinion" and "not evaluated". In December 2019 it was announced that a pilot programme would recruit part-time contracted "community reviewers" to expedite its fact-checking process. In 2019 Facebook started allowing fact-checkers to check ads and flag them as false. US fact-checkers were also gradually allowed to remove paid ads they thought were false. In 2020, Facebook started belatedly acting against Holocaust deniers, anti-vaxxers, and QAnon, the dangerous movement responsible for many right-wing conspiracies.

Unsurprisingly, AI is increasingly used for detecting fake news, fact-checking, image verification, and video authentication.[42] A natural language processing engine can go through the subject of a story along with the headline, main body text, and the geo-location. Further, AI will find out if other sites are reporting the same facts. In this way, facts are weighed against reputed media sources. Using predictive analytics backed by machine-learning, a website's reputation can be predicted through considering multiple features like domain name and Alexa web rank. When it comes to news items, the headline is key to capturing the attention of the audience. The technology has grown in significance as it tries to understand pages' context without relying on third-party signals.

Artificial Intelligence has been instrumental in discovering and flagging fake news headlines by using keyword analytics. A key tool in NLP is a neural network architecture that encodes words to a latent space, decodes to a translation, typo, or classification. The 'neural' part of the architecture learns which words to

[39] J Paschen. Investigating the emotional appeal of fake news using artificial intelligence and human contributions. Journal of Product & Brand Management, 29:223–233, 2019

[40] Sky News, 'Coronavirus: 90 attacks on phone masts reported during UK's lockdown', available at https://news.sky.com/story/coronavirus-90-attacks-on-phone-masts-reported-during-uks-lockdown-11994401, accessed 16 February 2021.

[41] The QAnon conspiracy theory and a stew of misinformation fueled the insurrection at the Capitol, available at https://www.insider.com/capitol-riots-qanon-protest-conspiracy-theory-washington-dc-protests-2021-1, accessed 17 February 2021.

[42] Zhou, X, & Zafarani, R (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.

focus on and for how long.[43] In theory, after these tools are combined, they can train models efficiently and in parallel. However, using these types of AI systems to online fact-check massive amounts of user-generated content either *ex-ante* or *ex-post,* for example, public posts would require an unprecedented amount of computational power; for example, the BERT natural language processing system uses over 100 million parameters.[44] As well as raising questions about whether platforms are monitoring content, using AI for online fact checking *at scale* would require an exponential increase in computing power relative to linear generation of new content. However, AI might have a role to play at the generative level. Rather than dealing with the burden of content moderation downstream, Facebook uses AI to mitigate creation of fake accounts[45] and to detect word patterns that could indicate fake news.[46] It does not attempt or purport to draw its own conclusions about the accuracy of a story. As it can extract and manipulate information from text, NLP could be instrumental in fact-checking online content.

The same approach cannot be said for checking the *accuracy, trustworthiness, and validity* of user-generated content. Nor can NLP address the amplification effect of algorithms designed to make people share and engage with as much content as possible by showing them things they were most likely to be outraged or titillated by. They were not created to filter out what was false or inflammatory. The upstream problem—and the one that is ultimately far more difficult to resolve—is whether it is possible for AI to establish whether an online claim found in user-generated content is true or false,[47] nor is it yet capable of determining what Bernal labels as 'false narratives'.[48] There is also the challenge of identifying misinformation where content may have been deliberately and intentionally fabricated, may or may not be true but is not verifiable, but is produced with the intention of making a profit, and/or pushing a certain ideological or political agenda but is believed by a user to be accurate and shared accordingly. Even using AI to identify and remove inaccurate content is debatable: the user would never appreciate the corrective effect, or the social shaming associated with the marketplace of ideas.

The legitimacy of collaborative AI to counter disinformation will largely depend on the perceived impartiality and credibility of participating news organizations.

We have not yet realized the true potential of artificial intelligence in combating fake news. The future needs more sophisticated tools that can harness the power of artificial intelligence, big data, and machine learning to stop fake news making ripples in the user world. Technically speaking, the main problems associated with automated journalism, in terms of narrative and critical considerations, surround their low quality. Yet the effects of AI and AI systems are not only going to refashion human relationships but redistribute labour and creativity. Accordingly, examination is required of these transformative effects on journalism and news production.

Regulating AI in the European Union

While newsrooms forge ahead with automated journalism and fact-checkers increasingly rely on AI to search for patterns of analysis that indicate deceptive content, the EU has been struggling to come up with a single framework for the regulation of artificial intelligence.[49] The deployment of AI inside institutions traditionally responsible for democratic accountability without the appropriate safeguards and a lack of coherent ethical and legal safeguards raises alarms. Recognizing this problem before the cancer of

[43] Available at https://medium.com/@edloginova/attention-in-nlp-734c6fa9d983, accessed 23 February 2021.
[44] Facebook said it "disabled" 1.2 billion fake accounts in the last three months of 2018 and 2.19 billion in the first quarter of 2019; see Phsy Org, 'Fake Facebook accounts: the never-ending battle against bots', available at https://phys.org/news/2019-05-fake-facebook-accounts-never-ending-bots.html, accessed 23 February 2021.
[45] Rob Lever, Phys Org, (24 May 2019), "Fake Facebook accounts: the never-ending battle against bots", available at https://phys.org/news/2019-05-fake-facebook-accounts-never-ending-bots.html#:~:text=Facebook%20says%20its%20artificial%20intelligence,before%20they%20can%20post%20misinformation, accessed 23 March 2021.
[46] 'Facebook is using AI to remove fake news', available at https://www.clickatell.com/articles/digital-marketing/facebook-using-ai-remove-fake-news/, accessed 23 February 2021.
[47] https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/, accessed 16 February 2021.
[48] Bernal, P (2018). Facebook: Why Facebook Makes the Fake News Problem Inevitable. N Ir Legal Q, 69, 513.
[49] Black, J, & Murray, AD (2019). Regulating AI and machine learning: setting the regulatory agenda. European journal of law and technology, 10(3).

disinformation metastasizes is a fundamental principle of the responsible AI movement.[50] The AI HLEG expert group42 published its Ethics Guidelines for Trustworthy AI in April 2019.[51] According to the Guidelines, trustworthy AI should meet three criteria throughout the system's entire life cycle:

(1) lawful - respecting all applicable laws and regulations

(2) ethical - respecting ethical principles and values

(3) robust - both from a technical perspective while considering its social environment.[52]

The concept of 'transparency' has been posited as a prerequisite to machine-learning and AI systems to allow one to grasp "some sense of understanding the mechanism by which the model works".[53] However, both supporters and critics argue that the very nature of the technology means that complete transparency is an unachievable goal.[54] In the purest sense, every instance of AI trained on personal data would be incompatible with the requirements of Article 5(1) (a) GDPR: personal data cannot be processed in a transparent (and concise) manner in a way that is understandable to the data subject.[55] Thus, transparency gives way to the concept of interpretation, or the process of translation of "an abstract concept (e.g., a predicted class) into a domain that the human can make sense of".[56] There can be different levels of understanding depending on a person's age, mental condition, and education so that "meaningful information about the logic involved"[57] under GDPR rules may vary. Frustratingly, the GDPR does not explicitly provide examples of what level of understanding needs to be explained. Some opine that application of the criteria like 'average person' tests[58] may solve the problem.[59] However, considering protection of personal data as a fundamental right under the EU Charter,[60] such an approach does not exactly mitigate the discriminatory effects that its proponents claim the GDPR is tasked with eliminating.[61] Even the Unfair Commercial Practices Directive provides more protection: if a clearly identifiable group who are particularly vulnerable can be identified, the impact of commercial practice should be assessed from the perspective of the average member of *that group* [Emphasis Added].[62]

Other efforts in Europe include the European Data Protection Supervisor's (EDPS) recent public consultation on the necessity of a "digital ethics" framework to address technological developments such as AI and robotics;[63] also, the European Union Agency for Fundamental Rights report on "Getting the Future Right: Artificial Intelligence and Fundamental Rights"[64] and the Council of Europe's CAHAI Secretariat report "Towards Regulation of AI Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law".[65] The Software and Information Industry Association (SIIA) developed its own "Ethical Principles for Artificial Intelligence and Data Analytics" in late 2017 to help in well-developed ML/AI.[66]

---

[50] Available at https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/, accessed 12/03/2021

[51] AI HLEG Ethics Guidelines For Trustworthy AI, available at https://ec.europa.eu/digital-singlemarket/en/news/ethics-guidelines-trustworthy-ai, accessed 27 March 2021.

[52] Id at 5.

[53] Ribana Roscher; Bastian Bohn; Marco F Duarte; Jochen Garcke et al "Explainable Machine Learning for Scientific Insights and Discoveries" doi: 10.1109/access.2020.2976199 [accessed 27.03.2021, but compare to Yavar Bathaee "The Artificial Intelligence Black Box and the Failure of Intent and Causation" Harvard Journal of Law & Technology Vol.31, No. 2 Spring 2018, 906-919, 929

[54] Recital 58 GDPR

[55] Ribana Roscher et al (n 52), quoting Montavon, G, Samek, W, & Müller, K R (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73, 1-15; See also Leiser and Dechesne (n 12).

[56] Methods for interpreting and understanding deep neural networks", Digit Signal Process., vol 73, 1-15, Feb 2018.

[57] Arts 13(1)(f) 14(1)(g) and 15(1)(h) GDPR

[58] Similar to that of the "average consumer" who is reasonably well informed, and reasonably observant and circumspect, see CJEU in *Severi*, C-446/07, ECLI:EU:C:2009:530, para 61 and the case-law cited

[59] See Guide to GDPR (n 12), 120

[60] Rec 1 GDPR; see also Art 21(1) EU Charter: Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.

[61] Rec 71 GDPR

[62] Unfair Commercial Practices Directive 2005/29/EC, Art 5(3)

[63] https://edps.europa.eu/data-protection/our-work/ethics_en, accessed 23 March 2021.

[64] https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf, accessed 23 March 2021.

[65] https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a, accessed 23 March 2021.

[66] https://www.pr.com/press-release/735528

The rapid development of AI technologies has stimulated a variety of responses from the EU to the phenomenon.[67] Its Declaration expressed the need for a workable definition of AI and AI systems, for determining ethical guidelines for its use, alongside liability considerations associated with deployment of AI.[68] Its efforts were intended to build a human-centric 'ecosystem of excellence' and 'ecosystem of trust' for industry deployment of AI. The EU eagerly reiterated the importance of investment in terms of both money and data into research and application of AI. Furthermore, the Communication from the Commission called for the EU to "strengthen fundamental research and make scientific breakthroughs…facilitate the uptake of AI and the access to data", "supporting testing and experimentation" and "encourage the wider availability of privately-held data".[69] Additionally, the JRC's report "Artificial Intelligence: A European perspective"[70] provided different accounts of the developing technology alongside possible impacts, examining artificial entities when they involve unique cognitive or behavioural implications. The report stresses that known psychological attributes and systematic biases appear to be further amplified by digital media.[71]

At the precipice of a significant technological development, with concerns about unforeseen harms, and without an emerging winner from the heterogeneity of competing ethical approaches (as well as concerns about favouring ethics in lieu of regulatory intervention backed up by sanctions), the European Commission introduced a legislative process to address the risks that AI poses to safety and fundamental rights. The proposal for the Regulation of Artificial Intelligence[72] harmonizes rules in a risk-based and 'future-proof' manner to provide predictable and sufficiently clear conditions under which enterprises can develop AI applications and plan their business models, while ensuring that the EU and its Member States maintain control over regulatory standards, so not forced to adopt and live with standards set by others.

The regulation uses a risk-based approach to regulating AI. Applications with minimal or no risk are permitted without restrictions;[73] high risk AI is permitted, subject to specific transparency obligations;[74] higher risk applications are permitted, subject to compliance with AI requirements[75] and *ex-ante* conformity assessment;[76] while forms of 'unacceptable AI' (subliminal manipulation; exploiting children or mentally disabled persons; general purpose social scoring; remote biometric identification for law enforcement in publicly accessible spaces (with exceptions)) are deemed unacceptable and prohibited.[77]

---

[67] European Commission. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. (2020). Available here.; European Commission. White Paper. On Artificial Intelligence - A European approach to excellence and trust. (2020). Available here.; European Commission. Liability for Artificial Intelligence and other emerging digital technologies. (2019). Available here. European Parliament. European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics. (2019). Available here.; AI HLEG. Ethics Guidelines for Trustworthy AI. (2019); Available here.AI HLEG. A definition of AI: Main capabilities and scientific disciplines. (2019). Available here.; AI HLEG. Policy and Investment recommendations for Trustworthy AI. (2019). Available here; Council of Europe. Guidelines on Artificial Intelligence and Data Protection. (2019). Available here; Council of Europe. Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data. (2019). Available here; Council of Europe. Report on Artificial Intelligence Artificial Intelligence and Data Protection: Challenges and Possible Remedies. (2018) Available here; EDPB. Guidelines 3/2019 on processing of personal data through video devices. (2020) Available here. EDPS. Opinion 3/2018. EDPS Opinion on online manipulation and personal data. (2018) Available here. ICO. Guidance on the AI auditing framework. Draft guidance for consultation. 2020. Available here; EU Science Hub, "*Artificial Intelligence: A European Perspective*", https://ec.europa.eu/jrc/en/publication/artificial-intelligence-european-perspective, accessed 16 February 2021.

[68] https://ec.europa.eu/jrc/communities/en/node/1286/document/eu-declaration-cooperation-artificial-intelligence

[69] European Commission Brussels, 'Communication from The Commission', 25.4.2018; COM (2018) 237 final; 'Artificial Intelligence for Europe', {SWD (2018) 137 final}.

[70] Lewandowsky, S, Smillie, L, Garcia, D, Hertwig, R, Weatherall, J, Egidy, S, … & Leiser, M (2020). Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making, available at https://publications.jrc.ec.europa.eu/repository/handle/JRC122023, accessed 11 June 2021.

[71] ibid, 45.

[72] Proposal for a Regulation laying down harmonized rules on artificial intelligence, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence, accessed 05 June 2021; Communication on Fostering a European approach to Artificial Intelligence, https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence, 05 June 2021.

[73] Art 69: no mandatory obligations, but possible voluntary codes of conduct for AI with specific transparency requirements.

[74] Art 52: notify humans that they are interacting with an AI system unless this is evident; notify humans that emotional recognition or biometric categorization systems are applied to them;

[75] Title III, Ch 2.

[76] Title III, Annexes II and III; CE marking and Process (Title III, Ch 4, Art 49).

[77] Title II, Art 5.

Recital 68 of the Draft Regulation suggests that 'certain AI systems intended to interact with natural persons or to *generate content*' [italics added] may pose risks of impersonation and deception; therefore, AI-generated content must comply with transparency obligations regardless of whether classified as high risk. Manipulated images, audio, or video content should also contain a disclosure that the content has been artificially created or manipulated by labelling not only the output but the source material.

Journalistic responsibility when using machine-learning systems

Automation bias leads decision makers to assume that quantitative methods are superior to qualitative methods, and to reduce the task at hand to applying the quantitative data available. This undermines and devalues the necessary complex contextualization that human reasoning applies. With trust in legacy media a rather fluid dynamic, and public attitudes to the credibility of social media as a replacement to traditional news outlets, and both legal and political fallout from the use of automated decision-making, there is a general attitude among Europeans that the use of AI should be transparent and discernible. A recent Eurobarometer study focusing on AI found that 80% of the representative EU population sample think that they should be informed when a digital service or mobile application uses AI.[78] A recent representative survey probed the German public's attitudes towards use of online AI and use of machine learning to exploit personal data for personalization of services.[79] Attitudes towards personalization were found to be domain-dependent: Most people find personalization of political advertising and news sources unacceptable. The degree of moral outrage elicited by reports of immoral acts online has been found to be considerably greater than for encounters in person or in conventional media.[80]

The diffusion between a designer's intention and the actual behaviour of an AI system creates a "responsibility gap" that is difficult to bridge with traditional notions of responsibility[81] and is subject to ongoing debate (e.g., the EU's recent statement on artificial intelligence by the Group on Ethics in Science and New Technologies).[82] Traditionally, responsibility for any journalistic error would attach to the newsroom through a variety of ethical obligations, regulatory frameworks, and most importantly, tort (defamation/libel) law. However, autonomous learning machines are fed data sources, learn without supervision, and produce outputs that cannot be predicted. In a normative sense, responsibility means being able to explain actions that you were able to control. Someone will be responsible to the extent that they know the circumstances and facts around decisions that they undertake. Thus, responsibility can be ascribed to a principle of 'control'. AI systems and machine-learning models turn that principle on its head. At present, there are AI systems in newsrooms that are able to decide on a course of action and to act without human intervention. The rules on which they act are not fixed, but change during the operation of the AI system, by the system itself. The machine learns and produces a series of actions, where traditional ways of attributing responsibility are not compatible with the control principle. No-one has enough control over the machine's actions to be able to assume responsibility for them. These constitute the "responsibility gap".

Explainability

The second example of explainability refers to the obligation to make the internal logic of AI systems discernible to human beings.[83] This does not require disclosing the inner working of the logic, nor does it equate to algorithmic transparency.[84] The ethos behind explainability lies in distinguishing what input produced what undesired effect in order to justly allocate responsibility for that effect. As they may be asked to explain, any media or news organization using AI should be prepared to explain its workings and

[78] EU Barometer Public Opinion, available at https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/%20instruments/STANDARD/surveyKy/2255, accessed 16 February 2021.
[79] A Kozyreva, S Herzog, P Lorenz-Spreen, R Hertwig, and S Lewandowsky. Artificial intelligence in online environments: Representative survey of public attitudes in Germany. 2020.
[80] Crockett, M J (2017). Moral outrage in the digital age. Nature human behaviour, 1(11), 769-771.
[81] ibid, 45.
[82] ibid, 45, referring to the EU's recent statement on artificial intelligence by the Group on Ethics in Science and New Technologies, available at https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-%09be1d-%2001aa75ed71a1, accessed 23 March 2021
[83] N Gill, P Hall, & N Schmidt Proposed Guidelines for the Responsible Use of Explainable Machine Learning (2020)
[84] For a detailed account of why this is not feasible, see Leiser and Dechesne, (n 12).

rationale and should understand the reasons behind its output. When processing personal data inside an AI system or when personal data appears in the training set, the legal requirements for explainability come from a variety of hard and soft law measures.

Under Article 5(1)(a) GDPR[85], a media organization that uses an AI system will have to ensure that any personal data be processed in a 'lawful', 'fair' and 'transparent' way. The latter requires that information disclosure be discernible by the data subject whose data is subjected to processing.[86] Analysis of personal information by AI systems could amount to (a) 'processing of personal data', and (b) 'profiling.[87] Although these provisions keep data subjects abreast as to the "generalities" of data processing, Articles 13 and 14 provide the legal basis for data subjects to be provided with 'meaningful information about the logic provided' where relevant.[88] The subject of an AI-generated news report could exercise their rights against the media organization if acting as a data controller.[89] The Information Commissioner's Office (ICO), the UK's regulator for data protection, has stressed the need for explainability in the use of AI systems.[90] In this regard, the ICO stresses that explainability is needed, not only for regulatory compliance and system accuracy, but also to ensure that data subjects are informed.[91] Thus, explainability should be interpreted and applied widely. The regulator notes that the approach of institutions to 'explaining [machine-learning]-assisted decisions should be informed by the importance of putting the principles of transparency and accountability into practice, and of paying close attention to context and impact'.[92] It is advised that these require appreciating the –

- Purported aim of the modelling.
- Type of modelling derived and as implemented.
- Variables and/or data to be used in when processing, inclusive of their integrity, validity, and availability.
- Data set on which the model is trained.
- Purported and literal impact of the model.
- Target audience of the explanation.
- Required explanation for said audience's intelligibility.
- Any other reasonable consideration.

The ICO envisages explainability to be tailored, not only to the relevant intelligence of the audience, but to the context in which the AI system is used; therefore, any machine-learning that uses personal data, including content creation, moderation, and fact-checking, should be explainable, not just to data subjects, but to anyone with a stake in accessing good quality and/or corrective journalism. Because of algorithmic opacity and AI systems' autonomous and ever-evolving learning curve that transforms data into an incomprehensible form, along with obscurity in AI decision-making processes, anything less than the evolving legal standard for transparency and procedures for ensuring explainability will likely involve push-back by regulators.

Disclosure & transparency

[85] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) 2016.
[86] The 'Transparency Principle'; See also GDPR, Rec 64.
[87] GDPR, Art 4(2) & (4).
[88] GDPR, Rec 39, 58 & 60.
[89] The extent of the 'right to an explanation' is hotly contested. For various takes on the extent of the right, see Edwards, L, & Veale, M (n 16), 16, 18.; S Wachter, B Mittelstadt & C Russell, Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR, 31 Harvard Journal of Law & Technology 2 (2018) 841–887; F Doshi-Velez, M Kortz, R Budish, C Bavitz, S Gershman, D O'Brien, et al, Accountability of AI Under the Law: The Role of Explanation, Working Draft (2019) 1-21; A Selbst & J Powles Meaningful information and the right to explanation, 7 International Data Privacy Law 4 (2017) 233-243; S Wachter, B Mittelstadt & L Floridi, (n 5) 76–99.
[90] The Information Commissioner's Office (ICO), Guidance on the AI Auditing Framework: Draft Guidance for Consultation (2020); The ICO, Explaining Decisions Made with AI Draft Guidance for Consultation Part 1, 2 & 3 (2019).
[91] ibid.
[92] The Information Commissioner's Office, Explaining Decisions Made with AI Draft Guidance for Consultation Part 2 (2019), 4.

As AI systems in newsrooms are data hungry and require massive amounts of information, *data transparency* is crucial for building trustworthy AI. Explanations may help ordinary citizens understand how data was processed within an AI system, but this does not amount to the same as disclosure. As Kissinger asserts, "there is a fundamental problem for democratic decision-making if we rely on a system that is supposedly superior to mere humans but cannot explain its decisions".[93]

By shining light on what is actually explained, explanations serve as a means to verify the accuracy of the explanation. However, explanations amount to a single branch of transparency.[94] Bloch-Wehba argues that "true algorithmic transparency goes far beyond an explanation of a challenged action to the individual that is affected".[95] Transparency is a complex construct that evades simple definitions. It can refer to explainability, interpretability, openness, accessibility, and visibility.[96] Overall transparency encompasses disclosures about the AI system, the logic involved, information about the data, and how that data is used. The European Parliament's Governance Framework for Algorithmic Accountability and Transparency report states: "transparency may relate to the data, algorithms, goals, outcomes, compliance, influence, and/or usage of automated decision-making systems (i.e., algorithmic systems) and will often require different levels of detail for the general public, regulatory staff, third-party forensic analysts, and researchers."[97]

A data processing impact assessment (DPIA) is necessary when *data processing* operations are likely to result in a high risk to the rights and freedoms of natural persons. A DPIA is mandatory in cases of the following non-exhaustive activities, characteristic of a majority of AI applications:

- systematic and extensive evaluation of the personal aspects of an individual, including profiling
- processing of sensitive data on a large scale
- systematic monitoring of public areas on a large scale.[98]

If a DPIA discovers that such risk exists and cannot be mitigated, the AI operator is obliged to consult the data protection regulator.[99] The most severe regulatory action against an operator who cannot demonstrate ability to comply is "temporary or definitive limitation including a ban on processing",[100] but it has certain time limits.[101] The EU commission has proposed introducing a requirement to undertake an overall AI impact assessment to ensure that any regulatory intervention is proportionate, and distinguishing those being "high risk" from the remainder. Two cumulative criteria clarify when and how AI should be specified as bearing high risk: (1) a sector where significant risks can be expected to occur and (2) application in such a manner, when significant risks are likely to occur.[102] In addition to this general category, some types of activity are considered as always bearing high risk, e.g. applications for recruitment and other situations that can have an impact on the rights of workers, use for purposes of remote biometric identification.[103]

[93] H Kissinger, How the Enlightenment Ends: Philosophically, Intellectually—in Every Way—Human Society Is Unprepared for the Rise of Artificial Intelligence, 2018 The Atlantic, available at https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/, accessed 26 March 2021.
[94] Brkan, M (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. International journal of law and information technology, 27(2), 91-121.
[95] Bloch-Wehba, H (2019). Access to Algorithms. Fordham L. Rev., 88, 1265.
[96] Felzmann, H, Fosch-Villaronga, E, Lutz, C, & Tamò-Larrieux, A (2020). Towards transparency by design for artificial intelligence. Science and Engineering Ethics, 1-29.
[97] Koene, A, R Richardson, Y Hatada, H Webb, M Petel, D Reisman, C Machado, J L Violette, and C Clifton. A governance framework for algorithmic accountability and transparency, 2018. EPRS/2018/STOA/SER/18/002, 2018.
[98] Art 35(3) GDPR.
[99] Rec 84 and Arts 35(1) GDPR.
[100] Art 58(1)(f) GDPR.
[101] Art 36(2) GDPR.
[102] Brussels, 19.2.2020 COM(2020) 65 final White Paper On Artificial Intelligence - A European approach to excellence and trust, 17
[103] ibid, 18.

Conclusion

The world is abuzz with the prospects and promises of AI and its ability to process large datasets accurately in order to derive predictive outcomes. Four factors have ensured AI's success as we see today in almost every sector. These include: "exponential increased computer processor capabilities; emergence of global digital networks; advances in distributed computing (hardware and software); and especially the emergence of Big Data".[104] The indescribable amounts of available personal data for fuelling AI, increased processing power, access to cheaper and greater storage capacity has ensured advances in creating ML-model.[105] AI holds great promise and utility for news media and fact-checkers. However, as is widely acknowledged, many machine-learning applications function as 'black-boxes' that have been built using vast amounts of 'historical data'.[106] Predictive profiling offers a unique approach to threat mitigation that begins from the point of view of the aggressor/adversary and is based on an actual adversary's methods of operation, their modus operandi. This method is applicable to securing virtually any environment and to meeting any set of security requirements. The post-crime orientation of criminal justice is increasingly overshadowed by the pre-crime logic of security. Frameworks for preventing crime are not as concerned with gathering evidence, prosecution, conviction, and subsequent punishment as in targeting and managing through disruption, restriction, and incapacitation those individuals and groups considered to be a risk. Using unexplainable, unaccountable, irresponsible AI will end the system of checks and balances. Worryingly, few insights can be derived about the internal logic of AI systems.[107] The absence of understanding the logic behind machine-learning is grave, not only from a journalistic integrity perspective (given the necessity of warranting algorithmic transparency) but also a broader societal perspective.[108]

The consequences of using biased training data in journalistic endeavours that rely on machine-learning is a prime example. Subjects of a news article could face the social stigmatization of being labelled a 'suspect' or even a 'criminal'.[109] Under the present system of checks and balances, the burden of proof is on the person discriminated against to show that this was the result of a) bad data and/or b) an algorithm, and/or an automated decision. This would require a person subjected to a decision to have access to the model used by the newsroom, the training data, and the raw data from, for example, the coroner's office. Therefore, automated journalism and AI systems used in newsrooms are an unchecked power with indeterminable consequences for society. This can lead to further discrimination, catastrophic economic and social losses, as well as loss of reputation and, in some cases, infringement of civil liberties. It is important that everyone affiliated with media production and consumption, including readers, have some kind of understanding of what artificial intelligence actually is and how it operates. This fundamental understanding will not only shape how we use it but enable us to use it in a way that actually serves society, rather than just the technology.

---

[104] Subramanian, Ramesh (2017) "Emergent AI, Social Robots and the Law: Security, Privacy and Policy Issues," Journal of International Technology and Information Management: Vol 26: Issue 3, Article 4; at 84

[105] "…it is data, in many cases personal data, that fuels these systems, enabling them to learn and become intelligent" (see The Norwegian DPA Report supra., at 5)

[106] Wired, Machine Learning and Cognitive Systems: The Next Evolution of Enterprise Intelligence (Part I) (2020), available at https://www.wired.com/insights/2014/07/machine-learning-cognitive-systems-next-evolution-enterprise-intelligence-part/, accessed 16 February 2021; Wired, Location Intelligence Gives Businesses a Leg Up Thanks to Real-Time AI (2020), available at https://www.wired.com/wiredinsider/2019/06/location-intelligence-gives-businesses-leg-thanks-real-time-ai/, accessed 16 February 2021.

[107] Jason Brownlee, What is Deep Learning? (2019), available at https://machinelearningmastery.com/what-is-deep-learning/, accessed 16 February 2021; see also N Gill, P Hall, & N Schmidt, *Proposed Guidelines for the Responsible Use of Explainable Machine Learning* (2020).

[108] G Ras, M Gerven, & W Haselager Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges. ArXiv 1803.07517 (2018)

[109] G Sinha, To suspect or not to suspect: Analysing the pressure on banks to be 'Policemen', 15 Journal of Banking Regulation Vol. 1 (2014) at 75-86; SAS Institute, What is next-generation AML? The fight against financial crime fortified with robotics, semantic analysis and artificial intelligence (2020) at 8, available at https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers-ebooks/sas-whitepapers/en/next-generation-aml-110644.pdf, accessed 16 February 2021.