



Artificial Intelligence: opportunities and concerns in the era of Big Data.

Ethical and practical issues with decision-making and Generative AI in the era of ChatGPT.

Sztuczna inteligencja: możliwości i obawy w erze Big Data.

Etyczne i praktyczne kwestie związane z podejmowaniem decyzji
i generatywną sztuczną inteligencją w erze ChatGPT.

Giulia Varvarà

ORCID: 0009-0007-8664-1085

e-mail: giulia.varvarao2@gmail.com

Centrum Nauki Experiment

Artificial Intelligence has made impressive progress in the past few decades, and, especially in the last couple of years, so many different systems able to simulate human response in a very realistic and coherent way have been made available to the general public. This has opened the road to new possibilities for the use of AI in a variety of contexts, starting from the use of generative AI to the shortening of otherwise long working tasks (such as programming). However, it also created unforeseen issues that have yet to be addressed, since there is a lack of legal and ethical guidelines for the use of these new AI tools. This article analyses some of the most controversial applications of these new AI systems, highlighting both problems and ethical concerns, as well as the possible ways in which they can be dealt with in the future.

Key words: Artificial intelligence, Ethics, Generative artificial intelligence, Data-driven decisions.

Introduction

Machines that can act and interact like humans have always been in the human imagination, as much of the sci-fi literature can prove, but in the past few years, they seem to become increasingly close to reality.

While there are no sentient robots or anything similar yet, the field of Artificial Intelligence has made great strides in the last decade and now AI is becoming increasingly present in our daily life.

But what is AI exactly?

The Oxford Dictionary defines Artificial intelligence as the “study and development of computer systems that can copy intelligent human behaviour”. This includes visual perception, speech recognition, decision-making, and translation between languages. Even though it may seem like a futuristic concept, it isn’t.

The first article on how to create intelligent machines was written by Alan Turing in 1950 (Turing, 1950), who also created the imitation Game (also called the Turing test) to test machines’ intelligence. The term Artificial Intelligence per se was created only 6 years later in an 8-week-long summer project at Dartmouth College, where researchers from different scientific areas put their minds together to create a new line of research aimed at building machines that would simulate human intelligence (Haenlein, Kaplan, 2019).

Scientists have been working on the development of AI since then and, while some important results have been achieved in the early years, the most impressive jumps ahead are way more recent, and have become possible due to the advent of Artificial Neural Networks and Big Data.

Now AI is employed in various applications, from customer service to education, entertainment, identification systems, speech recognition, language translation and the list goes on. Even very sensitive fields have seen increasing employment of AI, like healthcare, employment decisions and even art, raising different types of concerns.

While it is clear that AI is a very powerful tool capable of improving and speeding up many tasks and is now part of our daily lives, it does not come without different types of concerns that ought to be taken into account when using it.

This article aims to analyse some of the most controversial uses of AI, pointing out both the advantages and issues brought by the use of these new AI-powered products, as well as the actions that must be taken to ensure their ethical use.

Bias, Privacy and Explainability: AI in decision-making processes

One of the most interesting and controversial applications of AI nowadays is in the context of decision-making processes, which are used in many different fields, from employment and business to healthcare.

The main feature of AI is its ability to learn, but to do so it requires data to train on, which is why the era of big data allowed AI applications to progress so much.

This implies that one of the most important limitations is that every model that requires training is going to be just as good as the data that is trained on, which raises many different issues.

One of the most relevant is the possibility of making biased decisions that reflect the hidden biases in the training datasets themselves.

An example of this was the AI recruiting tool that Amazon was trying to develop around 2014, which turned out to not select employees in a gender-neutral way when it came to software developer jobs and other technical posts (Dastin, 2018). This was because the data on which the model was trained showed male dominance in the tech industry, which brought the model to penalise resumes not coming from men.

The aforementioned case is just one example of how an AI tool trained on a biased dataset can have a very negative influence and enlarge the gap between social groups instead of improving fairness. While Amazon stated that this recruiting tool was never used and the project was abandoned, there are other areas in which AI is still going to be used for decision-making. One of them is Healthcare.

Healthcare is one of the most delicate and complicated fields in which AI could be employed. In healthcare, the risk of making wrong decisions, like giving incorrect diagnoses and treatments, which will deeply influence the lives of patients is a notable and constantly present problem. When AI is introduced in the decision-making process, the risk is exacerbated and becomes more controversial since, for example, a wrong decision could also result in biases in the dataset on which the models are trained.

There are so many different cases and causes for the patients' illnesses that usually the training data is not able to represent all the variations and complexities in the health problems of the patients because it is simply not large enough (Wang, Preininger, 2019).

Moreover, most of the medical data and the treatment protocol available on which the models can be trained are more likely to exist for illnesses that usually affect white men, since most of the studies and clinical trials are made on this group. This means

that algorithms trained on such data will most likely make poorer predictions for other groups (black people, women, youngsters etc.). This is not only a health issue because it means that patients belonging to other groups will receive worse care, but, since healthcare is a way to promote social justice, it will increase the gap between the different social groups. There is a real risk of overfitting the model to a dataset that does not represent the actual population (Morley et al., 2020). Overfitting means that the algorithm would be tuned too finely to the data available in the training set and in turn will lead the model to perform more poorly on new data, meaning that the model will not be generalizable, and its output cannot be used for data that are different from the training set. If an overfitted model is used for data different from the training ones, the results are going to be very poor and, when talking about such sensitive areas of application, also dangerous.

Moving the focus from the individual-based level to policy decisions, when AI is used to predict illness trends to deploy resources in the best possible way, there is a risk that the models are susceptible to distortion, which could lead to wrong predictions. An example of this is the case of the Google Flu trend in 2013 (Vayena, 2015) when the algorithm overestimated the spread of influenza because it misclassified illnesses with similar symptoms as confirmed cases of influenza.

While overestimating may not seem a great problem, we have to keep in mind that the system is constantly learning from the data that it receives on the performance of the measures employed. This means that a misguided intervention not only means a non-optimal use of resources but also can affect how future interventions will be carried out and how many resources will be made available for similar scenarios (Morley et al., 2020).

This use of AI for predicting events is not only employed in healthcare but also when it comes to the justice system for predicting crime rates in different areas, deploying resources, organising policing and deciding on specific criminal cases.

As it is possible to imagine, also this application is extremely controversial given the extremely biased nature of the datasets made by criminal records, which are fed to models that will in turn, give output that is most likely to discriminate against minorities that have been believed to be more prone to crime and which as consequences had run into more legal problems.

These policing algorithms are among the ones that Cathy O’Neil describes in her book “Weapons of Math Destruction” (O’Neil, 2018), in which she illustrates how Big Data is sometimes used to increase social injustice and threaten democracy.

What these dangerous applications of Big Data that she describes have in common is this hidden bias that comes from the datasets from which the models have been trained, but also their opaqueness.

Opaqueness is one of the issues concerning many forms of AI and is also described as the “Black-box problem”.

The term Black box comes from the flight recorder machine, but it is now used as a metaphor to describe systems that work in a hidden way that is not easily possible to understand.

This is the case with most of the AI algorithms, like unsupervised machine learning, Artificial Neural Networks and most deep learning approaches, since it is incredibly hard even for experts to decipher how these algorithms learn and produce outputs.

While these methods usually prove to be very effective, they do not explain how conclusions are reached, raising transparency, accountability and responsibility issues, especially when they are applied to social problems.

The delicate nature of the areas in which these approaches are used (analysis of health conditions, job applications, prison terms etc.) stresses the need for transparency in the decision-making processes to avoid decisions that are detrimental and unjust. This is why some algorithm decision-making systems are now including, together with the black-box AI model, a second AI model that can explain the black-box outputs.

While this seems an easy solution, explainable AI models have been accused of low fidelity and of not being able to properly capture how a black-box system forms its decisions (Peters, 2022).

Explainable AI (XAI) is the research field that was born out of this need to have understandable AI approaches and, as the name says, explainable.

However, XAI does not come without its challenges, which depend both on the people and on the nature of the data. Most people will lack the expertise to understand the explanations. Moreover, even when some variables are left out to avoid bias in data such as racial profiling and discrimination in general, the algorithms can still make biased choices because the excluded characteristics can be reflected in other variables. For example, race or income level can be inferred by addresses or gender can be inferred by the schools attended.

Another challenge lies in the fact that, given their learning nature, algorithms are not static but change over time, which means that the explanations also need to be dynamic since they can become obsolete. Moreover, since the data are collected from

different sources, the decisions made by the AI are also influenced by multiple algorithms, which are involved in the original data generation and extrapolation process, and contexts, which interact and can interfere with one another and make the explanation more complicated.

More difficulties come from the nature of the problems addressed as well because they tend to be ambiguous, and the outputs tend to be context-specific, so elaborating a general explanation of the AI algorithms and their decision systems that goes beyond the specific application and can explain the general working of the system becomes trickier. Moreover, the fact that the causality between input and output is explained does not necessarily mean that the algorithm is using that causality to reach its decisions, since its internal workings could follow a completely different learning path.

Lastly, even if an explanation for the algorithms could be found, there would still be the need to check that it stays the same over time as not only the algorithm's mechanisms but also the explanation of its working itself is subject to changes over time (de Bruijn et al., 2022).

Therefore, XAI aims to create more transparent and less opaque AI approaches, but it might not always provide acceptable explanations or increase the level of trust, on the contrary,

it might even produce the opposite results. The relationship between transparency and trust is not straightforward, and knowing exactly how a system is using the data fed to it might make users more sceptical and worried about their data usage.

AI needs to be trusted to be used, especially since it handles a lot of data, which is often extremely personal. This is particularly the case of healthcare applications since patients and their data are extremely linked and, moreover, individuals are encouraged to share more and more information about their health. Who has access to these data and how they are used and shared through many different institutions and for which purpose is a matter of extreme importance, which makes Privacy and data handling one of the main concerns about AI.

The first privacy-related concern is how data is acquired, since in this new technological age, every human creates more data than he is aware of and this information can be used in ways that are not known or clear to the person from which the data are taken.

The way these data are used raises additional concerns, since they can map health problems just as well as someone's behavioural habits and preferences (Li, Zhang, 2017). Moreover, since AI usually employs cloud computing, cybersecurity and pos-

sible data breaches are other concerns that need to be kept in mind, not only when thinking about AI in decision-making, but in general when dealing with applications that require and produce a lot of data.

Generative AI, opportunities and controversies

AI has been used in a variety of fields, raising several different issues, but the application that attracted more attention and created more debate recently is for sure Generative AI.

Generative AI can be defined as a type of Artificial intelligence that, when given a prompt, can create text, images etc.

The most well-known example is ChatGPT, the AI chatbot developed by OpenAI that became incredibly popular at the end of 2022.

This chatbot makes use of a type of models called Generative Pre-trained Transformers (GPT), which are a type of Large Language Models (LLM) that pre-trains on a large amount of text data and then fine-tunes on specific tasks, like question-answering or text classification (Ray, 2023).

During the pre-training, the model uses as sources a copious number of web pages, articles and books in an unsupervised way, which means it does not require human-imputed labels or annotations. In this phase, the model learns to recognise and generalise language patterns, which it will then use to predict the next word needed in a sentence given the previous ones.

There have been many different versions of the GPT model since its first release in 2018 and the famous ChatGPT, released in November 2022, is built upon GPT3.5 and GPT4, making it able to create coherent and realistic responses in conversation settings.

While ChatGPT still requires very clear and specific prompts to create meaningful answers, it can understand the context in text-based conversations, produce linguistically coherent, correct and context-accurate responses, and it can be adapted to various tasks.

These characteristics, together with the lack of need to be trained for new tasks and its scalability, make it a very attractive tool to use in many different contexts, from education to content generation or even programming and code writing.

If these types of chatbots can be so helpful in diminishing and speeding up workloads, why are they also controversial?

Some of the issues are the same as explained above for the case of decision-making AI, like the risk of biases in the source material which can be spread by the model: the output generated by this type of application is going to end up in the source material as well, propagating any bias. This is of particular concern when such tools are used in fields like research, where the risk of biased outcomes must be carefully addressed.

Moreover, there is a need for quality assessment of the generated results as GPT models create very coherent responses that can be both high-quality or very low-quality or inappropriate depending on the case (Dwivedi et al., 2023).

This is true also because being trained on a very large amount of data, there is a high risk of overfitting, which means that the generalization ability of the model needs to be constantly monitored and improved. In a similar way, there is a need to check the factual accuracy of the chatbot responses since this type of language models might generate text that is not reliable, accurate and in some cases even fake.

A practical example of these not usable and misleading outputs can be found in asking ChatGPT for references (e.g., peer-reviewed papers) on a specific topic since when it is not able to provide them, instead of stating so the model fabricates good-sounding but fake references (Hillier, 2023). The fact that these chatbots can make up responses when unable to give a satisfying answer to the user highlights how great the need for GTP-generated responses must be carefully analysed before being used.

Other challenges of GPT models are also common to the decision-making AI applications, such as the explainability issues, since these models are hard to interpret and explain, or the privacy concerns, which come from the fact that these models access a very large amount of data from many different sources, raising concerns on how these data are obtained, stored and used.

Another source of preoccupation is the risk of overreliance on these models as they become more advanced, which could reduce critical thinking and individual problem-solving skills (Ray, 2023).

Generative AI also has to face safety concerns since it can create harmful content like hate speech or fake news and, moving from ChatGPT to other types of generative AI, also deepfakes.

Deepfakes are videos created with the use of Machine Learning algorithms superimposing the expression of one person onto another one's face or using existing videos of someone and reverse engineering their speech to have them say anything. The results of this AI application can be very realistic, and most people can have trouble

understanding whether they are looking at a real video or a deepfake. This affects the perception of true and false and the confidence and trust in news on social media and on the internet in general.

An example of how dangerous deepfakes can become is the incident that happened on February 3rd 2023 in which a deepfake video of US President Joe Biden making a transphobic speech went viral on Instagram (Hudnall, 2023).

Online civic culture is damaged by such a climate of lack of trust: users will be less likely to behave collaboratively and responsibly towards others when sharing news themselves as well (Vaccari, Chadwick, 2020). Citizens might also just start missing news altogether to avoid this uncertainty.

Deepfakes are not the only visual outputs created using Artificial intelligence. OpenAI has also created a system for creating images from a text prompt called DALL-E and, even though there are limitations on the number of images generated for free, its main perk for users is that there is no restriction on the use of generated images, even to commercial purposes (Cooper, 2022).

Moreover, other projects have recreated the results of DALL-E, allowing anyone to have access to free versions of this application. The most well-known one is DALL-E mini, which uses a smaller model than the original one used by Open AI and produces lower quality results, but is available to everyone for free.

Again, a model is just as good as the data that it is trained on, so even in this application, the risk of bias is very much real. When given a prompt, AI will output the image that it thinks represents the most input given the training data fed to the model. This means when asked for a nurse, it will produce images of women, when asked for a scientist will give images of men, when asked for a beautiful woman it will output back skinny white girls (Parra, Stroud, 2023). Overall the model is generating mostly white people, men in the cases of roles demonstrating a higher level of education, and women in the cases of caretaking ones. It pictures very well the biases existing in the training dataset which is also limited to images described in English, preventing other cultures from being represented (Dayma, 2021). Moreover, the content represented in pictures and art also affects what people think themselves and others can do.

A very peculiar issue raised by this AI application is the question of copyright and ownership, which needs to be addressed at both ends.

The most controversial part of systems like DALL-E, ChatGPT and similar, is that they train on extremely large datasets containing pretty much everything available on the internet, not necessarily asking the owners of such material the permission to use it.

While in chatbots the use of copyright material for training the models might be hidden and not so clear to recognize, that is not the case when it comes to images. Boris Dayma, the mind behind DALL-E mini, admitted that, especially in the early version of the model, watermarks were often present in the generated output (Dayma, 2021).

Artists employ watermarks as a way to protect their art pieces from being used without permission. The presence of watermarks in AI art means that the models have been trained on them without the consent of the authors, who will see their style reproduced by AI applications without having any say in the matter.

Many websites and platforms have been created to check if images of someone or someone's art have been used to train AI models, but knowing if these images have been used does not ensure that the owner can take back control over them.

This can be demonstrated by the case of Robert Kneschke, a German photographer who found out that some of his material had been included in the LAION dataset, which is in turn used to train AI image generators. He asked for his pictures to be removed but had instead been told that the dataset contains only the links to his images which are available on the internet, so they couldn't remove them (Growcoat, 2023).

There has been an intense debate going on in the past few years about whether the use of copyrighted material for training AI should be considered "fair use" and then be allowed.

"Fair use" is defined as the copying of copyrighted material for a limited and transformative purpose. It could be argued that this is exactly what AI does, since it creates new material from the source data. On the other hand, the way the algorithms are designed might lead to the AI creating outputs that are identical to the original works, which would infringe the original creator's right, or that are derivative works, which are also protected under copyright law. Lastly, AI-generated art could create market harm since it will compete if not replace the original works since these can be acquired and used for free while the original work requires the author's permission (Torrance, Tomlinson, 2023).

The relationship between ownership rights and AI is brand new and there is still no proper regulation of the use of copyrighted pieces in training sets, but it is necessary to find a way to balance the need for AI to train and develop and the right of the authors to protect their artworks.

The other side of the ownership issue in generative AI regards the output: Who is the owner of the generated materials?

The first obvious candidates are the authors of the training material, without which the model could have not been trained, then there are the developers of the model, the model itself, the user who gave the input that generated the material or also no one at all (Avrahami, Tamir, 2021).

The same question arises also in the area of research, where AI tools like ChatGPT can be used to help scientists generate research ideas and hypotheses but also in writing down the paper to send to journals (Ray, 2023). Should these chatbots then be listed as co-authors as well?

The definition of Author in the case of copyright law is different in different countries, but most states use the so-called “human authorship principle”, which means that to be protected by copyright, a work must be produced by a human, which in turn means that in most countries AI cannot be listed as the author at all. However, the fact that it cannot be listed as an author under copyright law does not mean that it cannot be considered as one in the field of an academic paper (Lee, 2023).

In academia, work is published not if it can be copyrighted but according to whether it can contribute to academia, this means that even chatbots could be listed as co-authors.

Right now most journals still refuse to list AI as an author, but this is not because they are not humans, like in the case of creative work, but because the available AI systems do not satisfy the authorship criteria, as stated in the editorial policies of the *Nature* journal. In particular, current AI tools do not comply with the need for accountability, which is one of the criteria for the attribution of authorship, which requires the author to be accountable for the work submitted and this cannot be effectively applied to Large Language Models (LLM), like ChatGPT, which cannot then be listed as authors.

The fact that it is just the impossibility of being able to hold LLMs accountable for their work presently and not their not being humans that prevent chatbots from being listed as co-authors leaves the door open for future more developed AI systems to be included as co-authors in academia.

Conclusions

In this article, we have examined the most controversial applications of AI given its state of the art, and we have pointed out where there is a need for intervention, especially from the regulation and ethics side.

AI needs plenty of data to be able to become a useful tool, which means that it is necessary to ensure that these data, which can be extremely sensitive, are protected

every step of the way, from the moment they are imputed in the datasets to when they become a new output.

Moreover, there is a need to carefully check what is used in the training to avoid biases, which are always dangerous, especially because AI will take these biases and then propagate them, shaping also the vision of the world of AI users.

A lot of attention should also be put to the way data is retrieved, making sure the owners of such data have agreed to them being used by AIs.

This article also highlighted the need for new regulation about the use of copyrighted pieces in training datasets and how to deal with ownership rights at the other hand as well, since it should be decided who should be the owner of the AI outputs and there is a lack of clarity, especially from the legal point of view, about that.

Some authorities are starting to recognize this need as well. The European Union, for example, voted on an “*Artificial Intelligence Act*” in June 2023 and lawmakers are now trying to finalise the new legislation which will aim to classify AI systems with different requirements and obligations tailored on a ‘risk-based approach’. According to this classification, some AI systems presenting “unacceptable” risks will be prohibited, while “high-risk” ones will be authorised if they respect some requirements and the AI systems presenting only “limited risk” will be allowed but still will be subject to very light transparency obligations (Madiega, 2023).

This is a first step towards better AI regulation but, especially given the much slower nature of legislation compared to technological development, it needs to be accompanied by an effort to educate also the general population about the risks and problematic issues of AI as well.

This article has analysed how AI can be a very powerful tool, but can also be dangerous if used carelessly and without a good understanding of the mechanism behind it.

We put the emphasis on how different applications and types of models present the same contentious points, which highlights how such issues are an intrinsic part of AI and no type of use, even the most playful ones like generating images, will not be subject to them. This indicates a pressing need not only to face these issues from a technical, legal and ethical point of view but also for a proper education about AI.

The general population should be educated about the advantages but also the risks and the problematic points of these systems, in order to acquire enough literacy of AI and at least a rough understanding of how the processes involved actually work.

Given the increasingly important role of AI in everyday life, everyone must become responsible and aware, as ignorance should not be the excuse for unethical or

problematic behaviours. Finally, we stress the need for legislative advances to be accompanied by an AI-literacy programme. We believe that schools, formative centres and workplaces should start to provide lectures about AI, enabling a wide variety of users to understand the mechanics of such tools as soon as they start to use them.

There are already some efforts in this direction, especially in schools (Casal-Otero et al., 2023), but for now, they seem to be more concentrated on developing technical and applied skills and are not much focused on the ethical side of AI. Moreover, these literacy programs always face the practical problem of the lack of teaching staff educated in the subject.

The intention of this article is not only to point out the need for new regulations, duty of the legislators, but also the need for general awareness and education, to which more people can contribute.

So we summarised the points that present more ethical and technical issues in the most used AI applications, accessible also to the general public and present in everyone's everyday life. Moreover, a basic explanation of how common AI applications work, which concerns arise and the need for both legislation and education is given in a language suitable for non-experts in order to bring awareness.

AI is a tool and, as John Knoll (one of the minds behind Photoshop) said once "Any tool can be used for good or bad. It's really the ethics of the artist using it."

But AI is a tool available not only to artists but to everyone. The power that it can have in people's vision of the world, how it can affect anyone's life now that most of its applications are available to everyone, means not only that it should be more clearly regulated, but especially that people should be better educated about its pro and cons and ethical problems.

Every user should approach AI responsibly and know enough about its internal work to use it with the right caution and not as the fast and easy solution to everything or blindly believe it.

Bibliography

- Avrahami, O., Tamir, B. (2021). Ownership and Creativity in Generative Models. *ArXiv*. /abs/2112.01516
- Casal-Otero, L. *et al.* (2023) ‘Ai literacy in K-12: A systematic literature review’, *International Journal of STEM Education*, 10(1). doi:10.1186/s40594-023-00418-7.
- Cooper, D. (2022) *Is dall-e’s art borrowed or stolen?*, *Engadget*. Available at: https://www.engadget.com/dall-e-generative-ai-tracking-data-privacy-160034656.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAL2kuMch9qy7ShbU_7eef88Z1DO-q69cTXNqkFiKWqDTJHGGu9wZ4SIqaz1NuVsTAuySEAGMartNgHEK-9oxZnfWH0uhdT8tv6LrfVcm3T7vx4Tv83e1LOWTLUCbgF_9DtLhoYCOy-IxTHxZKRLFYHZXUeER2rurshypkjj7ZAtOzlp [Access: 01 August 2023].
- Dastin, J. (2018) *Amazon scraps secret AI recruiting tool that showed bias against women*, *Reuters*. Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> [Accessed: 31 July 2023].
- Dayma, B. (2021) *Dall-e Mini explained*, *W&B*. Available at: <https://wandb.ai/dalle-mini/dalle-mini/reports/DALL-E-Mini-Explained-with-Demo--Vmlldzo4NjIxODA#the-datasets-used> [Accessed: 01 August 2023].
- De Bruijn, H., Warnier, M. and Janssen, M. (2022) ‘The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making’, *Government Information Quarterly*, 39(2), p. 101666. doi:10.1016/j.giq.2021.101666.
- Dwivedi, Y.K. *et al.* (2023) ‘Opinion paper: “so what if chatgpt wrote it?” multidisciplinary perspectives on opportunities, challenges and implications of Generative Conversational AI for Research, practice and policy’, *International Journal of Information Management*, 71, p. 102642. doi:10.1016/j.ijinfomgt.2023.102642.
- Gille, F., Jobin, A. and Ienca, M. (2020) ‘What we talk about when we talk about trust: Theory of trust for AI in Healthcare’, *Intelligence-Based Medicine*, 1–2, p. 100001. doi:10.1016/j.ibmed.2020.100001.
- Growcoat, M. (2023) *Ai image dataset demands money from photographer who requested removal of his photos*, *PetaPixel*. Available at: <https://petapixel.com/2023/04/26/ai-image-dataset-demands-money-from-photographer-who-requested-removal-of-his-photos/> [Access: 01 August 2023].

- Haenlein, M. and Kaplan, A. (2019) 'A brief history of artificial intelligence: On the past, present, and future of Artificial Intelligence', *California Management Review*, 61(4), pp. 5–14. doi:10.1177/0008125619864925.
- Hillier, M. (2023) *Why does chatgpt generate fake references?*, *TECHE*. Available at: <https://teche.mq.edu.au/2023/02/why-does-chatgpt-generate-fake-references/> [Access: 04 August 2023].
- Hudnall, H. (2023) *Fact check: Video altered to show Joe Biden making transphobic remarks*, *USA Today*. Available at: <https://eu.usatoday.com/story/news/factcheck/2023/02/09/fact-check-video-edited-show-joe-biden-making-transphobic-remarks/11211453002/> [Access: 04 August 2023].
- Lee, J.Y. (2023) 'Can an artificial intelligence chatbot be the author of a scholarly article?', *Science Editing*, 10(1), pp. 7–12. doi:10.6087/kcse.292.
- Li, X. and Zhang, T. (2017) 'An exploration on artificial intelligence application: From security, privacy and ethic perspective', *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* [Preprint]. doi:10.1109/icccbda.2017.7951949.
- Morley, J. *et al.* (2020) 'The ethics of AI in health care: A mapping review', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3830408.
- O'Neil, C. (2018) *Weapons of math destruction: How big data increases inequality and threatens democracy*. London, UK: Penguin Books.
- Papp, D., Krausz, B. and Gyuranecz, F. (2022) 'The AI is now in session – The impact of digitalisation on courts', *Cybersecurity and Law*, 7(1), pp. 272–296. doi:10.35467/cal/151833.
- Parra, D. and Stroud, S.R. (2023) *The ethics of AI Art*, *Center for Media Engagement*. Available at: <https://mediaengagement.org/research/the-ethics-of-ai-art/> [Access: 01 August 2023].
- Peters, U. (2022) 'Explainable AI lacks regulative reasons: Why ai and human decision-making are not equally opaque', *AI and Ethics*, 3(3), pp. 963–974. doi:10.1007/s43681-022-00217-w.
- Ray, P.P. (2023) 'CHATGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope', *Internet of Things and Cyber-Physical Systems*, 3, pp. 121–154. doi:10.1016/j.iotcps.2023.04.003.
- Turing, A.M. (1950) 'Computing Machinery and intelligence', *Mind*, LIX(236), pp. 433–460. doi:10.1093/mind/lix.236.433.

- Torrance, A. W., & Tomlinson, B. (2023). Training Is Everything: Artificial Intelligence, Copyright, and Fair Training. *ArXiv*. /abs/2305.03720
- Vaccari, C. and Chadwick, A. (2020) ‘Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news’, *Social Media + Society*, 6(1), p. 205630512090340. doi:10.1177/2056305120903408.
- Vayena, E. (2015) ‘Ethical challenges of Big Data in public health’, *European Journal of Public Health*, 25(suppl_3). doi:10.1093/eurpub/ckv169.024.
- Wang, F. and Preininger, A. (2019) ‘Ai in health: State of the art, Challenges, and Future Directions’, *Yearbook of Medical Informatics*, 28(01), pp. 016–026. doi:10.1055/s-0039-1677908.
- Zhang, Y. *et al.* (2021) ‘Ethics and privacy of Artificial Intelligence: Understandings from Bibliometrics’, *Knowledge-Based Systems*, 222, p. 106994. doi:10.1016/j.knosys.2021.106994.
- Editorial policies: Artificial Intelligence* (no date) *Nature news*. Available at: <https://www.nature.com/nature-portfolio/editorial-policies/ai> [Access: 04 August 2023].
- Artificial Intelligence* (no date) *Artificial-Intelligence noun - Definition, pictures, pronunciation and usage notes* | *Oxford Advanced Learner’s Dictionary at Oxford-LearnersDictionaries.com*. Available at: <https://www.oxfordlearnersdictionaries.com/definition/english/artificial-intelligence> [Access: 04 August 2023].
- European Parliamentary Research Service and Madiega, T. (2023) *Artificial Intelligence Act*.