Harvard Data Science Review • Issue 6.3, Summer 2024

Data Scientists Discuss AI Risks and Opportunities

David Banks[1] Gerard de Melo[2] Sam (Xinwei) Gong[3] Yongchan Kwon[4] Cynthia Rudin[5]

[1]Department of Statistical Science, Duke University, Durham, North Carolina, United States of America,

[2]Department of Artificial Intelligence and Intelligent Systems, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany,

[3]LinkedIn, Sunnyvale, California, United States of America,

[4]Department of Statistics, Columbia University, New York, New York, United States of America,

[5]Department of Computer Science, Duke University, Durham. North Carolina, United States of America

The MIT Press

Published on: Jul 31, 2024

DOI: https://doi.org/10.1162/99608f92.740dcedb

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

ABSTRACT

This is an edited summary of a virtual panel conversation that took place on December 19, 2023, concerning the risks and benefits of AI systems . The many topics that are covered include AI impacts on education, the economy, cybercrime and warfare, autonomous vehicles, bias and fairness, and regulation. In addition, the role for data scientists is discussed. But the field is moving quickly, and some of the issues and concerns may have changed by the time this discussion is published.

We note that the discussion was refereed, which led to some post hoc changes to the actual conversation. Most of the participants are mostly comfortable with the new points that have been attributed to them.

Keywords: autonomous vehicles, cybercrime, ethics, regulation

---

Discussion

David Banks (DB): What do you see as the major risks and major benefits of AI systems?

Yongchan Kwon (YK): I think we live in a fascinating time. Almost every industry sector is being rapidly changed by AI. It is impacting biological research, chemistry, robotics, and even politics and education. The change is broad, undeniable, and inevitable.

I would like to highlight its impact on education. AI makes personalized learning and teaching processes more efficient and productive. Let me give an example. I am teaching a Python class for first-year MS statistics students at Columbia. There are many students, and a typical question I get is “Hi Professor. Here is my code. It should work, but it doesn’t. Can you tell me what I’m doing wrong?” In previous years, such questions were time intensive. I had to read all the code, and diagnose the problem. It could easily take 30 minutes to an hour. But after ChatGPT, it became easier and faster for the students to debug code. Also, ChatGPT is available 24/7, so most problems get solved even when I am sleeping. As a result, it is highly beneficial to me too. By saving time, I can more efficiently concentrate on improving course materials and refining assignment questions, leading to an elevated quality in my teaching. That is the greatest benefit and virtuous cycle I am seeing in the university.

Sam (Xinwei) Gong (SG): I am most excited about how AI can contribute to the advancement of science. I was really impressed a few years ago when I read the paper from Google on AlphaFold (Jumper et al., 2021). Predicting protein structure was a hard problem for many years, and the AI behind AlphaFold took a quantum

leap in that space. I look forward to AI transforming materials science, helping doctors diagnose diseases, and contributing to our understanding of physics, chemistry, and biology.

Regarding risks, AI systems can clearly be used by bad actors. A more subtle risk is overreliance on AI outputs. For example, Steven Schwartz, a lawyer with Levidow, Levidow & Oberman, is accused of using ChatGPT to generate a legal brief that turned out to cite fictitious cases, falsely thinking that ChatGPT would not hallucinate.

Of course, there are also doomsday scenarios, with AI making smarter AI systems. Some fear that if the objective function of an AI is not aligned with human objective functions, then it might lead a super-intelligent AI to take over the world. I don't think it's completely crazy to worry about that kind of Skynet possibility.

But coming back to ground, one area that regulators and practitioners such as myself should focus upon is AI fairness. AI models are often trained with data generated or labeled by humans, and thus tend to reproduce or can even exacerbate any biases humans have. For example, when asked to generate images of people in different professions, a common text-to-image model would show those in more lucrative professions with lighter skin colors (Nicoletti & Bass, 2023). Fairness is especially important when AI is used to allocate economic opportunities. If an AI system is used to recommend job candidates to recruiters, we certainly don't want it to treat candidates differently based on personal attributes unrelated to their qualifications.

Gerard de Melo (GDM): AI has already done much to reduce language barriers, and it is beginning to substantially help people with visual impairment and other disabilities or special needs. I think applications in health care will be very important, especially as we move further toward personalized medicine.

In terms of risk, I agree with Sam that there is real danger that we will start to outsource a lot of key decision-making to AI. We don't necessarily do this because we think it is the right thing to do, but often simply because it is convenient, practical, or the only way to scale a system up. For example, if one is concerned about content moderation in social media, it really isn't possible for human beings to keep up with all the material that is produced, and so the only way forward is to offload much of the decision-making to AI. Someone can appeal an AI decision to remove content, but I fear we may wind up in a Kafkaesque situation where overwhelmed human moderators reviewing the appeals may overrely on AI scoring rather than thoroughly reviewing each appeal sufficiently well.

As AI becomes more powerful, I think we are likely to willingly grant AI more agency and authority. We may allow AIs to make reservations and purchases for us, select insurance plans and investment portfolios, and so forth. They may wind up controlling traffic, or being largely responsible for managing city budgets. So, we need to be very sure that they make decisions according to principles that humans can understand and with which they agree.

My greatest concern is AI involvement in warfare. It is possible that AI probabilities are being or will be used to identify buildings that are military targets, or even to steer lethal autonomous weapons.

And there is real concern about how AI misinformation, disinformation, or AI-generated deep fakes may affect political discourse and the mechanisms of democracy. Algorithms already influence what we read and what type of information we consume. This will continue and become amplified once we have AI-generated content that is personalized and further optimized to be engaging and persuasive.

There are a lot of risks, but overall, I am quite hopeful. Think back to the early days of the Internet. If someone had told us that the Internet would cause riots, stolen elections, toppled governments, and shorter attention spans, we might not have gained the benefits of fast communication and global access to information and services, which are all arguably net gains. If we take steps to mitigate the risks, the benefits of AI could be amazing.

Cynthia Rudin (CR): Things like a chatbot raises the floor for everyone in terms of writing. It is like free editing, which is a huge boon in a world in which so many of our brilliant colleagues don't speak English as a first language.

The benefits for health care depend upon the availability of data, the ability to train AI systems to align with human values, and the creation of systems of governance and redress. Right now, there is a lot of data hugging, in which data owners decline to share even anonymized data, which is super annoying. And the gains from individually tailored tutoring could be huge. As Gerard said, personalized assistants would be great—an AI could manage your travel, book your dental appointments, act as a secretary, and give you more time to do things you enjoy.

Perhaps self-driving cars will become safer than human drivers. Right now, maybe not so much, but the potential is there. This will depend upon the built environment and infrastructure that evolves to support autonomous vehicles, as well as the legal and insurance environment surrounding accountability. We shall have robots that can communicate with humans. We shall have smarter automated customer service. And one can make better illustrations for academic papers and other purposes. A grandfather could create a personalized comic book for his five-year-old granddaughter, even if he couldn't do an illustration before AI.

Also, facial recognition technology would make international travel faster and, presumably, lead to better security and safety. Hopefully, it can help detect human trafficking.

In terms of risk, I think the current major threat is disinformation and misinformation. Fake information can be used to start wars, as with Facebook and Myanmar (De Guzman, 2022). Right now, a lot of fake information is being used to persuade nations to lean away from democracy. These large AI systems amplify the power of a

handful of bad actors to inflict damage. An example is what happened to the stock market when the fake picture of the Pentagon being attacked went viral (Bond, 2023).

I listed facial recognition as a benefit, but it is also a risk. The technology can be inaccurate, leading to false arrests (Hill, 2023), but the root cause is usually overzealous policing that ignores the guidance provided by such AI systems. And it can be used in ways that limit individual freedom. A person can stand outside a mosque or a synagogue or an abortion clinic with their cell phone and take a picture of everyone who walks in. Someone who is in a witness protection program probably doesn't want a lot of people walking around with facial recognition technology. Additionally, there are concerns about the accuracy of facial recognition technology.

Even the collection of large databases of biometric data puts people at risk, which is why Europe has recently banned such collections (Heikkilä, 2023). We know such data sets can often be easily hacked. Recently, 23andMe was breached (Lyngaas, 2023), exposing the genetic data of millions of people. I cannot begin to guess what damage an AI could do with such information in the future.

Another risk is to creative people. Writers and artists may be displaced by AI systems. The sad thing is that these creatives generated the text and images upon which the large language models were trained, which may lead to them eventually losing their jobs. Local news organizations are also threatened---AI systems are taking their data and repackaging it without compensating those who worked to gather the news in the first place.

YK: I'd like to add my concern about the homogenization of AI-generated images. The homogenization I referred to indicates that AI models produce degenerate outputs lacking diversity, even in scenarios where multiple variations are possible. For instance, I prompted an AI startup's text-to-image model, suspected to be DALL-E, to create "images of a girl with roses" and all the images showed an Asian child with red roses. We want to have more diversity in the visual representations.

I'm seeing this kind of sameness in class as well, and think it damages the educational experience. For both last year and this year, I gave an assignment in which students were to use an API to collect and download JSON data, then wrangle it into a pandas DataFrame form. The assignments were rather similar, but students this year had the option to use ChatGPT. The majority of students successfully submitted their solution that accurately fulfilled the requirements of the assignment. Interestingly, last year, each student's solution was quite distinct, but this year, all the answers were essentially the same. In this case, it is good that students can learn how to collect and wrangle data in Python by leveraging AI technologies, but I worry that AIs will narrow people's thinking, especially in terms of education.

DB: I know that you have recently given testimony to Senator Schumer on regulating AI. Can such systems be

regulated, and if so, how should we do it?

CR: I repeat, I don't think we should permit black-box solutions for high-stakes decisions. There could be some exceptions, for example, when we know that the AI is 100% accurate, or where interpretability doesn't mean anything, or when there is no way to create an interpretable system. But the default should be to ban uninterpretable AIs from situations that matter.

Also, like Europe, I believe we should have no creation, storage, or use of large biometric databases, or algorithms that use such databases, without some licensure or certification process controlled by the government. Our government currently doesn't know how to do that for AI. The government inspects restaurants for food safety and tests people before issuing drivers' licenses and certifies buildings and many other things, but it doesn't have the authority to restrict people from collecting large biometric databases that could put people in danger. An AI system could affect many more people than a single unsanitary restaurant, so it seems wrong not to have protocols in place for inspecting an AI that has access to millions of people's biometric data.

We should also figure out laws about copyright. I'm not a copyright lawyer, but I have concerns about an AI system being trained on material I produce, and then recycling it without acknowledgement. Of course, this harks back to the problem of local news reporters generating content that is repackaged without compensation.

I think the government needs to use existing monopoly laws to prevent abuse by a small number of dominant AI systems. Right now, there are a number of lawsuits proceeding through the courts that address different aspects of my concerns (Grynbaum & Mac, 2023). A number of states and the federal government are suing big tech companies for monopolistic practices, and they're winning (Fung, 2023; Mac & Hill, 2022; Stempel, 2023). These big tech companies are taking our data and using it in ways that we would not want, cases in which if we had the option to opt out of the automated data collection, we would have exercised it. The data are ours, and we should control it.

We need to monitor large AI systems, especially recommender systems, for health repercussions, for spread of misinformation, and for societal impacts more broadly. President Biden has issued an executive order that does exactly this (Exec. Order No. 14110, 2023). It says we should track these AI systems, see what they do, and then regulate accordingly.

I think there was a missed opportunity by the Supreme Court, which held that for companies using recommendation systems, the recommendations are not content ("Gonzalez v. Google LLC," 2023; Leven, 2023;). Even if those recommendations contain very noxious ideas, the company is not responsible for the harm that is done. So, during the pandemic, it was entirely legal for AI recommender systems to suggest quack medicines, or to recommend YouTube videos that reinforce racial stereotypes, or to spread fake news.

GDM: Being based in Europe, I've been following the proposed EU AI Act legislation closely.[1] Overall, I think it is going in the right direction. We need to regulate specific applications, especially high-stakes ones, while leaving the door open to innovation in less critical domains. We need to regulate certain use cases, but not all applications of AI.

Surveillance is one of the major issues currently being debated. During the pandemic, many employers gained the ability to track in great detail specific activities their people were doing while working from home (Tung, 2020). That may make sense in certain limited settings, but at larger scales it leads to a dystopian society. Similarly, in online education platforms, every response is tracked, and students may feel pressure to constantly engage with the system. Another issue being discussed is whether AI-generated material should be marked as such. China already in 2022 presented their "Provisions on the Management of Deep Synthesis Internet Information Services," which require visible watermarks on AI-generated material (Chinese Central Cyberspace Affairs Commission, 2022).

Looking ahead, I think many of the current open source AIs remain reasonably benign. Malicious actors can use them for spam or perhaps impersonation, but there also numerous positive use cases, such as boosting productivity. For current models, transparency and testing requirements seem the most useful. At some point, we may reach a point where more restrictions are needed, when perhaps it should no longer be possible for everyone to download uncensored AI models lacking any safeguards. Of course, the concern is that if one country enacts regulation, other countries with laxer rules will likely remain, leading to an arms race in AI capability. National borders do little to prevent AI access.

There are hopeful signs. The G7 Hiroshima Principles (G7 2023 Hiroshima Summit, 2023) and the Bletchley Park Summit (Bletchley Park, 2023) are all moving regulation forward in a balanced way. But AI is evolving quickly, and regulators will have difficulty keeping up, especially in terms of fairness and interpretability of AI systems (O'Neil et al., in press). We shall need periodic risk assessments, red teaming, incident reporting, and certification processes. And the big unsolved problem is accountability. When an AI system makes mistakes, who or what is liable?

SG: I like the idea of tagging AI-generated content. Many artists and writers will find their livelihood suffer due to competitions from AI. Tagging would be one component of a system to ensure fair compensation for derivative work based on human creations. More importantly, tagging AI-generated content can mitigate the risk that such content gets used in nefarious activities such as fraud or spreading fake news.

Before moving into regulation, I would personally like to see more focused discussions on principles, especially around AI alignment and AI fairness. For example, there are many different definitions of AI fairness and accuracy, and there are mathematical proofs that not all of them can be simultaneously satisfied

(Chouldechova, 2017; Kleinberg et al., 2017), which is a challenge for regulation. A perfect solution can never be achieved, but I hope academics, industry scientists, and regulators can come together to find an approximate consensus for what fairness criteria should be applied on specific AI systems, especially in high-stakes applications such as hiring, health care, and content moderation where accuracy is important.

For example, in content moderating, one concept is counterfactual fairness. If a social media platform decides to take down a post that says 'Chinese people spread COVID,' then it should also take down posts that say 'Americans spread COVID' or 'Mexicans spread COVID.' But, of course, fairness in content moderation goes beyond this narrow principle. One wants the distribution of a predictor to remain the same if one changes a protected attribute while holding constant features that are not causally dependent upon that attribute.

DB: If I were trying to regulate an AI, I would start with Asimov's three laws of robotics. To paraphrase, (1) An AI cannot harm a human being, or, through inaction, allow a human to be harmed. (2) An AI must obey a human being, except insofar as it conflicts with the first law. (3) An AI shall protect itself, except insofar as it conflicts with the first or second law. Obviously, this needs to be built out, for example, by specifying different kinds of harm (e.g., psychological harm, fake news, etc.).

YK: Regarding regulation, we need to focus on the transparency and interpretability of AI models. Both are difficult to achieve. Probably a regulator would need to know the details of the training dynamics of an AI system: which data were used, how they were collected, what was the model architecture, what were optimization hyperparameters, how was it optimized, and so forth. I am working on a way to figure out which training data most control the output of an AI system, using an approximate influence function.

One economic aspect that is becoming very important is the data marketplace. Here companies (or people) buy and sell data, often for purposes of targeted e-marketing. Such data can be used to train AI systems, or for other purposes. Data are often acquired without explicit permission from the individuals in the data set, and they receive no compensation for the data they provide. To advance the data marketplace and, consequently, improve AI systems built upon it, it is essential to discuss key questions such as: 'What regulations should be applied for data marketplaces?' 'What would be a good notion of data valuation?' and 'With these valuation methods, how can we standardize and enhance the efficiency of the data marketplace?' (Ghorbani & Zou, 2019). This is an area where I think more regulation and attention are appropriate.

DB: Cynthia, I think you were the first person to mention autonomous vehicles. If widely adopted, they have the potential to put millions of people out of work: truck drivers, delivery persons, and Uber/Lyft/cab drivers. But they also have the potential to substantially lower the price of goods in stores, by reducing net transportation costs. What are your thoughts on economic fairness here?

CR: I'm not an economist, so I cannot answer that. But AI threatens to put many more people out of work than just drivers. Script writers, artists, cashiers, and many other professions are at risk. But we cannot go backward---the world doesn't work that way. We shall have to retrain people quickly.

We have already seen a lot of economic disruption. Electronic commerce is dominated by a small number of companies, and their actions have significantly changed the way people purchase goods. Their AI recommender systems do an amazing job of using the personal data we just hand over to them to enable evermore effective advertising, and ever-greater concentration of wealth in those few firms. And it is impossible for competition to arise, because only businesses that have our data can train the recommender systems. The net effect will be even greater income inequality.

SG: I'm not an economist either, but as a layperson I have been intrigued by the idea of universal basic income (UBI). With the advancement of AI, our society will become collectively far more productive, even though many people may lose their jobs. UBI can ensure that everyone can lead a dignified life. I have seen small-scale studies that suggest that UBI would improve people's well-being and create virtuous cycles for the society (e.g., Verho, 2022). However, it's worth noting that, of course, these studies were not performed in a world where AI has already taken over many people's jobs and some studies even boosted employment among the study subjects.

GDM: I agree that AI will make us all a lot more productive. At some point, many human jobs may consist of monitoring and checking the work of AI systems. This trend will continue as AIs become more involved in autonomous cars, planes, and ships, robotic manufacturing, and perhaps even care for senior citizens.

Personally, I think this is a good thing. In Europe, we have a lot of demographic aging, and similar trends apply to China, Japan, and North America. If properly managed, replacing human stoop labor by smart physical machines is a step that could make all our lives better. If we become more productive, I think we should have better safety nets to ensure that the benefits are shared more equitably.

DB: What do you see as the role for data scientists in a world hurtling toward very sophisticated AI?

SG: Just so that we are on the same page in terms of definitions, in industry there are often two distinct but related job titles in the same company: one is typically called AI engineer or machine learning engineer and the other is typically called data scientist. The AI engineer role is, namely, responsible for developing AI systems,

so the day-to-day work is heavy on coding, algorithm development, and system design. On the other hand, the data scientist role focuses more on experimental design and business insights, so it is heavier on statistics.

In that context, people hired to be data scientists may not build the deep neural networks that drive AI systems, but they understand how to conduct experiments that compare different architectures, training strategies, and so forth. In a world that is bound to be more reliant on AI systems, I think people with a data science background can bring more rigor to the understanding and the usage of these systems to boost trust and safety. One such area is uncertainty quantification: how to come up with accurate and efficient uncertainty estimates for AI predictions and how to leverage the uncertainty estimates for better long-term outcomes. Another direction involves the understanding of failure modes. For example, how can we detect that a large language model is hallucinating. Lastly, coming back to AI fairness, data scientists are well equipped to measure AI biases and investigate their root cause.

CR: Data scientists need to figure out if AI systems are working as intended. That will be tricky, because they are increasingly used for increasingly complex tasks. We have to figure out if there are unexpected outcomes. And we have to figure out their impact on society, which is a causal question, and statisticians are very good with causal inference.

I think AI will play a large role in health care. That is a high-stakes application, and we need to have people with expertise in health care to supervise the AIs that get built for that purpose.

GDM: I teach both data science and AI courses, so I have always felt that these areas are tightly intertwined. AI is driven by Big Data, so we need data scientists to work on better data collection, data preprocessing, and outcome assessment, perhaps even risk analyses.

At the same time, there is a trend toward data science workflows becoming more automated. Data scientists want to have AI help in cleaning and managing data---this will amplify our ability to derive insights from data. However, we need to ensure that the AI systems are actually doing the right thing, and not introducing unwarranted data transformation steps.

DB: There is a famous account of how the discovery of the ozone hole was delayed because statistical software smoothed over the ‘outliers’ in the satellite measurements (Kandel, 1991):

> The discovery of the ozone hole was announced in 1985 by a British team working on the ground with “conventional” instruments and examining its observations in detail. Only later, after reexamining the data transmitted by the TOMS instrument on NASA’s Nimbus 7 satellite, was it found that the hole had been forming for several years. Why had nobody noticed it? The reason was simple: the systems

> processing the TOMS data, designed in accordance with predictions derived from models, which in turn were established on the basis of what was thought to be "reasonable," had rejected the very ("excessively") low values observed above the Antarctic during the Southern spring. As far as the program was concerned, there must have been an operating defect in the instrument.

GDM: Right now, large language models are not yet very good at handling complex tabular data. Data scientists still have much to contribute. The kind of work we do will evolve, but with collaboration from domain scientists and AIs, we should be able to do more.

YK: I agree about the need to measure alignment and ensure that the AI outputs are correct and reliable. I think data scientists will continue to need a solid foundation in statistics as they will make critical decisions based on both data and the outputs of AI models. Without a deep understanding of the randomness of data and models, their ability to accurately interpret and draw meaningful insights may be limited. Related to this point, I think we should pay more attention to what is the right question to ask. AI will make coding and data management and report writing easier, freeing us to structure better research questions.

DB: Thank you all for your fascinating observations. We live in interesting times.

Addendum

DB: This panel discussion has been through two rounds of revision, which is good because none of us want to publish rubbish. But there are ethical issues in asking the panel to edit our words in ways that disagree with our opinions, and the panel struggled with this a bit in the first revision. I think we found a compromise that respects the referees' suggestions while staying true to our beliefs.

In the second round of revision, one of the referees was concerned that our discussion suggests that AI has put us in new territory. He or she lists a number of previous problems caused by expert systems, such as use by the military to select incorrect targets or financial meltdowns caused by algorithmic trading. The referee feels that failing to connect recent AI advances to previous issues arising from predictive analytics means that we shall always be 'reinventing the wheel' when seeking solutions.

My view is that, of course, AI issues are not completely new, but the last 2 years have seen an explosion of interest in the risks and benefits, and that growth changes the way we think about managing AI. In a similar way, climate change action was a long time coming, but people are now beginning to take it seriously, following Canadian forest fires, sunny day flooding in Miami, and extreme weather events in the Midwest. The new scope and scale *do* change things. For example, consider privacy. In the 1930s, one could hire a private

investigator to surveil anyone, but it was expensive. The arrival of the Internet made it so easy to violate privacy. Similarly, large language models make it easy to generate deepfakes and certain kinds of cybercrime.

I also disagree about reinventing the wheel. I wish we had enough prior experience with AI systems to confidently know how autonomous vehicles and large language models should be regulated, or AI-enhanced cybercrime could be curtailed. But I do not see how knowing that large language models are a continuation of predictive analytics offers a clear path to guidance on guardrails.

Disclosure Statement

David Banks, Gerard De Melo, Sam (Xinwei) Gong, Yongchan Kwon, Cynthia Rudin have no financial or nonfinancial disclosures to share for this article.

References

Bletchley Park. (2023). *Artificial Intelligence (AI) Safety Summit*. https://bletchleypark.org.uk/bletchley-park-makes-history-again-as-host-of-the-worlds-first-ai-safety-summit/

Bond, S. (2023, May 22). *Fake viral images of an explosion at the Pentagon were probably created by AI*. NPR . https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai

Chinese Central Cyberspace Affairs Commission. (2022). *Provisions on the management of deep synthesis internet information services*. http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

De Guzman, C. (2022, September 28). Meta's Facebook algorithms 'proactively' promoted violence against the Rohingya, new Amnesty International report asserts. *Time Magazine*. https://time.com/6217730/myanmar-meta-rohingya-facebook/

Exec. Order No. 14110, 3 C.F.R. 75191 (2023). https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

Fung, B. (2023, September 26). *US government and 17 states sue Amazon in landmark monopoly case.* CNN. https://www.cnn.com/2023/09/26/tech/ftc-sues-amazon-antitrust-monopoly-case/index.html

G7 2023 Hiroshima Summit. (2023). *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems*. https://www.mofa.go.jp/files/100573471.pdf

Ghorbani, A., & Zou, J. (2019). Data Shapley: Equitable valuation of data for machine learning. *PMLR*, 97, 2242–2251. https://proceedings.mlr.press/v97/ghorbani19c.html

Gonzalez v. Google LLC. (2023, December 9) In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Gonzalez_v._Google_LLC&oldid=1189094067

Grynbaum M., & Mac, R. (2023, December 27). The Times sues OpenAI and Microsoft over A.I. use of copyrighted work. *The New York Times*. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Heikkilä, M. (2023, December 11). Five things you need to know about the EU's new AI Act. *MIT Technology Review*. https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/

Hill, K. (2023, August 6). Eight months pregnant and arrested after false facial recognition match. *The New York Times*. https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html

Jumper, J., Evans, R., Pritzel, A. Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kandel, R. S. (1991). *Our changing climate*. McGraw Hill.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In C. H. Papadimitriou (Ed.), *8th Innovations in Theoretical Computer Science Conference (ITCS2017)* (Vol. 67, 43:1–43:23). Schloss Dagstuhl -- Leibniz-Zentrum fuer Informatik. https://doi.org/10.4230/LIPIcs.ITCS.2017.43

Leven, R. (2023, February 21). *What to watch in a U.S. Supreme Court hearing on Section 230*. Berkeley College of Computing, Data Science, and Society. https://data.berkeley.edu/news/what-watch-us-supreme-court-hearing-section-230

Lyngaas, S. (2023, December 5). *Hackers access profiles of nearly 7 million 23andMe customers*. CNN. https://www.cnn.com/2023/12/05/tech/hackers-access-7-million-23andme-profiles/index.html

Mac, R., & Hill, K. (2022, May 9). Clearview AI settles suit and agrees to limit sales of facial recognition database. *The New York Times*. https://www.nytimes.com/2022/05/09/technology/clearview-ai-suit.html

Nicoletti, L., & Bass, D. (2023, June 14). Humans are biased. Generative AI is even worse. *Bloomberg*. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

O'Neil, C., Sargeant, H., & Appel, J. (in press). Explainable fairness in regulatory algorithmic auditing. *West Virginia Law Review*, 127. https://ssrn.com/abstract=4598305

Stempel, J. (2023, September 27). *Apple is ordered to face Apple Pay antitrust lawsuit*. Reuters. https://www.reuters.com/legal/apple-is-ordered-face-apple-pay-antitrust-lawsuit-2023-09-27/

Tung, L. (2020, November 27). *Microsoft 365's Productivity Score: It's a full-blown workplace surveillance tool, says critic*. ZDNET. https://www.zdnet.com/article/microsoft-365s-productivity-score-its-a-full-blown-workplace-surveillance-tool-says-critic/

Verho, J., Hämäläinen, K., & Kanninen, O. (2022). Removing welfare traps: Employment responses in the Finnish basic income experiment. *American Economic Journal: Economic Policy*, *14*(1), 501–522. https://doi.org/10.1257/pol.20200143

Weiser, B., & Schweber, N. (2023, June 8). The ChatGPT lawyer explains himself. *The New York Times*. https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html

---

©2024 David Banks, Gerard de Melo, Sam (Xinwei) Gong, Yongchan Kwon, and Cynthia Rudin. This article is licensed under a Creative Commons Attribution (CC BY 4.0) International license, except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. During the interval between the panel discussion and publication, a modified version of the proposed legislation was passed and can be found at the European Parliament website: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence ↩