The AI Act Should Be Technology-Neutral

By Patrick Grady | February 1, 2023

The EU aims to implement the world's first artificial intelligence (AI) regulation, the Artificial Intelligence Act, intended to allow people and businesses "to enjoy the benefits of AI while feeling safe and protected."[1] Unfortunately, the AI Act's broad definition of AI penalizes technologies that do not pose novel risks.[2] To resolve this, policymakers should revise the definition of AI to only apply to specific AI approaches that create significant challenges.

Policymakers have long valued the principle of technology neutrality, which holds that laws and regulations should avoid privileging or penalizing one set of technologies over another.[3] Technology neutrality does not necessarily demand that the exact same rules apply to different technologies. For example, if policymakers believe AI systems present novel risks not found in non-AI systems, they can and should address those risks.

This report shows that the AI Act is not, despite the intention of the European Commission, technology-neutral. Instead of addressing unique concerns about uninterpretable machine learning (ML) systems—a subset of AI systems—the Act would apply to a much broader set of AI systems that do not need regulatory intervention.[4] The result is legislation that would create significant overreach and potential harm to the EU's AI ecosystem. A better definition would limit the scope of the proposed law to only those technologies that pose novel risks.

AI development is not linear—it has gone and continues to go through various periods of flourishing ("springs") and stagnations ("winters"). The last "AI winter" has passed, but the EU is falling behind its global competitors—China and the United States—in AI research, investment, and

adoption. If the EU's AI Act limits innovation, as the current version of the bill is likely to do, the next AI winter could be strictly European.[5]

DEFINING ARTIFICIAL INTELLIGENCE

The definition of AI is crucial because it determines the AI Act's scope. Unfortunately, there is no universally agreed upon definition among practitioners, researchers, and developers. [6] AI is a moving target: As John McCarthy, the computer scientist who coined "artificial intelligence" put it, "As soon as it works, no one calls it AI anymore."[7] Indeed, the definition of AI has proved to be one of the most contentious parts of the AI Act.

In its initial proposal, the European Commission uses the following description:

> "Artificial intelligence system" (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.[8]

Annex I of the AI Act lists the following techniques and approaches:

> (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
>
> (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference, and deductive engines, (symbolic) reasoning and expert systems;
>
> (c) Statistical approaches, Bayesian estimation, search and optimization methods.

By using a list of techniques to define AI, the EU risks capturing too little, by not considering future AI methods, and too much, by including AI methods that have little to no bearing on policymakers' concerns.

THE OPPORTUNITY COST

The AI Act regulates "high-risk" uses of AI and identifies eight "high-risk" categories for which AI systems must undergo conformity assessments, and fulfill transparency requirements and post-market monitoring obligations. Compliance will be expensive: Conformity assessments alone may cost up to €400,000 and will ultimately be borne by consumers and EU businesses competing in global markets.[9]

The Commission admits that it does not know how many applications fall into the "high risk" category. Its impact report estimates up to 15 percent of all AI applications but the study supporting this estimate concedes that

"the actual proportion is unknown and will depend on the definition of high-risk AI systems."[10]

The AI Act will price AI systems out of the EU's market and hamper the arrival of others. Consider in table 1 the following "high risk" use cases that already improve—or promise to improve—the lives of Europeans.

Table 1: Existing and potential "high risk" AI use cases

High-Risk Category (Annex III)	Existing Use Cases	Potential Use Cases
"Biometric identification and categorisation of natural persons"	Age estimation technology verifies customers' ages in retail, and the ages of social media users[11]	Gaze detection to ascertain levels of alertness in a driver[12]
"Management and operation of critical infrastructure"	Cities use AI to manage traffic dynamically, set speed limits, and adjust pricing on highways[13]	Autonomous robot fleet that can fulfill inspection and maintenance of water, gas, and electricity facilities, reducing costs and improving safety[14]
"Education and vocational training"	Education providers (e.g., Pearson) use AI to co-grade essays; Massive Open Online Courses (e.g., Khan Academy, Coursera) use AI in education and adaptive learning tools[15]	Virtual reality in classrooms to examine ancient artifacts, communicate with historical figures, and explore ancient cultures[16]
"Employment, workers management and access to self-employment"	Recruiters use AI to screen high volumes of CVs; gig platforms use AI to find and allocate work for freelancers[17]	AI-powered virtual recruitment platforms and job fairs can facilitate hiring, onboarding, and training[18]
"Access to and enjoyment of essential private services and public services and benefits"	Public services use AI to improve delivery and accessibility through web portals, digital applications, and robots, and by optimizing internal workflows to guide case workers[19]	An emergency first response system that utilizes advanced speech recognition[20]

High-Risk Category (Annex III)	Existing Use Cases	Potential Use Cases
"Law enforcement"	Enforcement authorities use AI to tackle fraud and sex offenders and prevent terrorist attacks[21]	Deep learning-enabled computer vision to automate monitoring and analysis[22]
"Migration, asylum and border control management"	Over 1 billion e-passports rely on AI to significantly improve the efficiency of border controls[23]	An airport security "smart tunnel" that checks people while they walk through it[24]
"Administration of justice and democratic processes"	Lawyers use AI to automate data extraction and topic modeling[25]	Deep learning can support judges and accelerate case resolutions[26]

Critics of AI often overlook how the technology has quietly but dramatically improved many aspects of consumers' lives—for example, protecting them from fraud, filtering harmful content online, providing routing services anywhere and with any means, translating the world's languages, securing payments with facial recognition, tracking health data and well-being, diagnosing illnesses and diseases, improving customer service, optimizing deliveries, recommending new music, and improving voice and video calls.

Many new use cases may never arrive if the AI Act passes as proposed because innovators will face burdensome requirements. As the EU experiences its so-called "Digital Decade," its unicorns and most promising AI start-ups are already turning elsewhere.[27] The EU may further deter investment and hibernate its AI ecosystem for a long AI winter. To avoid this fate, the EU should revisit its definition of AI to be as technology-neutral as possible.

A TECHNOLOGY-NEUTRAL APPROACH TO REGULATING AI

AI will drive significant growth in the digital economy over the next decade, and the EU is in danger of being left behind.[28] The AI Act, introduced in reaction to fears about an out-of-control technology's potential impact on society, threatens to deter research and investment. Rather than succumb to the latest technology panic, the EU should take a technology-neutral approach when regulating AI, to create a level playing field between processes that use AI and those that do not. Doing so should not be controversial: In its proposal, the Commission explains that the definition of AI "aims to be as technology-neutral and future-proof as possible."[29] But,

as drafted, it threatens the most rudimentary notion of technology neutrality.

Technology-neutral regulations avoid unfairly favoring one technology over another, especially when there is no relevant difference between them. The European Commission's directive on regulating digital technologies defines "technology neutrality" as regulation that "neither imposes nor discriminates in favor of the use of a particular type of technology" but notes that technology neutrality "does not preclude the taking of proportionate steps to promote certain specific services where this is justified, for example, digital television as a means for increasing spectrum efficiency."[30]

The problem with technology-specific regulations is they can soon become obsolete, by failing to anticipate future innovations and requiring a perpetual amendment to stay relevant. For example, policymakers trying to regulate transistors or integrated circuits in the 1960s could never have imagined how these technologies would give rise to personal computers, mobile devices, and the Internet. Technology-specific regulations often hold back the adoption of superior technology.

Some people will always misuse new tools. But instead of unduly restricting the latest technology and thus holding back both positive and negative uses, policymakers should focus on holding those who misuse it properly accountable for their actions.[31]

The legislative conversation has narrowed the Act's scope, but not well enough. Policymakers should amend the AI Act to be as technology-neutral as possible. In this case, it should only apply to uninterpretable ML, since these only systems introduce novel risks and thus can justify a technology-specific approach.

To see why, the next section of this report evaluates the set of AI technologies in the scope of the AI Act—statistical, search, and optimization methods; AI; ML; and uninterpretable ML.

Statistical, Search, and Optimization Methods

For thousands of years, humans relied on the abacus for calculation. Mechanical calculators arrived in the 17th century (and were met with a familiar moral panic [see box 1]) and then were replaced by digital calculators in the second half of the 20th century. Calculators have since revolutionized mathematics, physics, chemistry, engineering, and astronomy. But imagine if governments had created a law limiting its use in "high-risk" cases, including construction and accounting. Such prohibitions would have undercut its development and use.

Box 1: Calculator panic

Richard Thornton, editor of the *Primitive Expounder*, a 19th Century American journal, railed against mechanical calculators with uncanny prophesy:

"[S]uch machines, by which the scholar may, by turning a crank, grind out the solution of a problem without the fatigue of mental application, would by its introduction into schools, do incalculable injury. But who knows that such machines when brought to greater perfection, may not think of a plan to remedy all their own defects and then grind out ideas beyond the ken of mortal mind!"[32]

And yet, compared with the lowly abacus, the calculator has introduced the capacity for great harm. It has substantially exacerbated errors and helped financial criminals, money launderers, and predatory lenders on a scale and speed otherwise impossible. Consider the criticism of a Dutch fraud risk scoring tool (*fraudescorekaart*).[33] It crudely assigns people as most likely to have committed fraud if they are, for instance, single, in a low-paying job, or poorly educated, or live in a less-favored neighborhood. The tool is not a complicated AI system, but rather a simple spreadsheet—a big calculator. After producing controversial outcomes, several municipalities decided to stop using the system.[34]

Nevertheless, it would have been a mistake to regulate calculators—and so it would be to include statistical methods in the AI Act. The decision to use the fraudescorekaart tool was a policy decision. Humans controlled all the inputs (employment, living situation, family, age, education, fraud history, etc.) and the weightings (some of which are entirely subjective rather than based on statistical substantiation) that together assigned fraud risks.[35] The correct response is to penalize those who curated and decided to use the technology, not to regulate Excel spreadsheets.

Unfortunately, this was not the Commission's approach. Fearing the potential for fraudescorekaart-esque scandals, the initial draft of the AI Act included in its scope the use of a lot of basic software (e.g., Annex III captures, basic linear regression models, and statistical programs used in spreadsheets) across all eight so-called "high-risk" sectors. Yet, this type of basic software does not impose novel risks, so creating unique rules for it violates the basic tenets of technology neutrality.

Artificial Intelligence

Some people think of advanced robots and self-driving cars when they hear the term "AI," but the term itself simply refers to computer systems that can simulate human-like actions, such as reasoning, perceiving, and acting. Therefore, many AI systems are quite basic.

Consider the Australian Robodebt case, wherein the Australian government deployed an AI system to automate its debt-recovery process.[36] Robodebt searched for discrepancies, decided whether former or current welfare recipients owed debt, and proceeded to hand out notices. The system sent out 700,000 debt notices, 470,000 of which turned out to be invalid.[37] This debacle had devastating consequences for both those who incorrectly received notices and the government, which continues to compensate for those errors.

When AI is involved in decision-making, it is easy to blame poor outcomes on the machine. And AI tools can misfire badly. Indeed, the Australian government's use of AI enabled the speed and scale of this disaster. But Robodebt was not an advanced self-learning or unpredictable AI system. Instead, the system connected simple algorithms, many of which were already long in use, with a decision-making algorithm to deduce overpayments using income data.[38] Staff were able to talk people through the system's process and even enter recipients' data on their behalf during a phone call.[39] Policymakers chose to deploy the system and had complete control and oversight of it but failed to supervise or vet the quality of its decisions.

The disaster is not of AI's making, but, according to Australia's Federal Court Justice Bernard Murphy, "a massive failure of public administration."[40] In addition to poor delivery and a negligent appeals process, the system was unlawful because it failed several legal principles, including "innocent until proven guilty."[41] It also used a debt-averaging method the government in Australia has since outlawed.[42] This case does not provide evidence that AI systems should be penalized; rather, it proves the importance of public sector accountability and traceability in government processes. The focus on AI is a dangerous distraction from the systemic problems in public policy and decision-making.[43]

The AI Act, as drafted, penalizes the use of AI in several applications that have substantially improved the lives of citizens: AI traffic systems that have made roads safer and saved police officers from tedious work; intelligent heating systems that reduce both costs for consumers and the burning of fossil fuels; and emergency response triage systems whose improved response times have helped save lives. All these systems are within the scope of "high risk" and are subject to immense compliance costs and transparency and monitoring requirements.

Many AI systems will fall under the approaches listed in Annex I (b): "Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems."[44] These systems should not be in the scope of the final text because they do not pose truly novel risks. Scapegoating the use of these systems merely hampers

innovation while overlooking the risky behavior and decision-making of humans.

Machine Learning

When criticizing the use of AI, legislators are often—even unknowingly—referring to ML systems. Novel concerns over transparency, autonomy, and responsibility are only prescient with ML. Recent proposals indicate that the EU is narrowing the scope to ML.[45] This is a move in the right direction, but policymakers should not stop there.

ML systems improve through observing data, building a model based on the data, and using the model as a hypothesis about the world and as a program to solve problems. In the 2000s, breakthroughs in ML marked the end of the second AI winter.[46] ML is now the current state of the art and drives most of the visible advances in AI.[47]

The introduction of learning systems has brought with it new challenges. Learning systems can misfire by reproducing, revealing, and exacerbating existing human biases when used for recruitment, translation, image recognition, policing and law enforcement, and facial recognition.[48] Systems can also misfire in unpredictable ways unrelated to training sets. For instance, algorithms pushing ads for STEM (science, technical, engineering, and math) jobs target fewer women because the system learned that young women are more expensive to reach.[49] Nefarious actors can also manipulate chatbots to spread disinformation or return misogynistic or racist remarks.[50]

Criticisms of these systems often discount that human processes are also prone—even more, and irremediably so—to discrimination, as well as to irrational decision-making and cognitive biases.[51] The study of avoiding and debiasing AI systems is growing and may be a more promising prospect than that for humans.[52] Research shows that developers can reduce bias in ML systems by simply leaving humans out of the loop.[53]

One of the most important challenges for the EU's economy is the transition to a green economy and fighting climate change. ML systems can help in a myriad of ways, from smart grids and sustainable agriculture to disaster management and extreme-weather prediction.[54] Many of these systems will fall under "high risk," as they relate to the AI Act Annex III's "Management and operation of critical infrastructure" and so firms looking to set up or innovate in the EU will face costly conformity assessments, transparency requirements, and burdensome monitoring obligations. For certain climate innovators, this will be prohibitively expensive; others will prefer to launch their products in AI-friendly jurisdictions such as the United States and the United Kingdom.

Penalizing the learning component would unfairly hold ML to a much higher standard than human processes and undermine the essential element of this "AI spring."

Uninterpretable ML

Due to their complexity, certain ML models (e.g., random forests and neural nets) create parameters unknown to their designer. These models are uninterpretable because the designer cannot decipher the output from the inputs.

Since the 2010s, many of the breakthroughs in AI, including drug discovery, image and voice recognition, game playing, generative art, and self-driving cars, have come from deep learning (DL) systems based on neural nets. While developers can use tools to improve these systems' explainability (i.e., how well the developer can communicate the system's behavior and output in a way that is understandable to the user), these methods are uninterpretable because the developer does not know how the system's features and their weights determine the system's output.[55]

The current draft of the AI Act outlaws DL techniques in "high risk" use cases, together with other uninterpretable systems (e.g., random forests). Article 13 requires of AI systems that "their operation is sufficiently transparent to enable users to interpret the system's output."[56] This requirement is impossible for uninterpretable systems.

Modern robotics, for instance, promises to replace human labor in mundane, dangerous, and deadly tasks, many of which fall under "critical infrastructure" in the "high risk" category. However, because these systems rely on uninterpretable ML, regulators will outlaw such systems from operating in the very sector in which their application is most promising.

The EU should not ban uninterpretable ML systems, but rather remove the interpretability requirement from the AI Act, as the latest Council text does.[57] Parliament should follow suit to avoid needlessly punishing the most effective systems.

Since it is only uninterpretable ML systems—not the more general categories of software, AI, or ML—that pose novel challenges for regulators, policymakers should reduce the scope of the AI Act to include only uninterpretable ML.

CONCLUSION

Some technologies are easy to regulate because they are easily defined, have limited uses, or present extraordinary risks, such as requiring aircraft to meet certain aviation safety standards or prohibiting asbestos because of unacceptable health risks. Yet, other technologies, including AI, have many uses and risks depending on context.

Consider knives. Surgeons use knives to operate on patients, chefs to prepare food, and artists to carve crafts. Knives can also be weapons. Whether knives increase risk, and whether they are desirable, depends on the use case. In some cases (e.g., publicly carrying specific types of blades), the use of a knife is illegal. However, attempts to list every knife that should be subject to regulation are ineffective, as new versions will keep escaping that definition. Likewise, AI is not easily defined and does not pose a single risk or have a single use case. There is no one kind of AI and, it is a technology that can both alleviate and exacerbate risks.

The AI Act should define AI in a way that only captures the genuine, novel risks of the tool and doesn't penalize its use per se.

New Definition

Policymakers should remove Annex I, and the following definition should replace that in Article 3:

> "Artificial intelligence system" (AI system) means a system that, **based on parameters unknown to the provider or user**, infers how to achieve a given set of objectives **using machine learning** and produces system-generated outputs such as content (generative AI systems), predictions, recommendations, and decisions, influencing the real or virtual environments with which it interacts.

This definition removes from scope AI systems that are no longer considered AI, and those in Annex I b and c; places the responsibility back on human operators for systems whose inputs are determined wholly by humans; and regulates only ML that poses truly novel concerns: uninterpretable ML. As mentioned, transparency requirements in Article 13 should be adjusted to ensure the AI Act does not outlaw these systems in "high risk" use cases.

Remaining Questions

There are two ways the scope of the AI Act threatens the EU's innovation ecosystem. The first and most important is the definition of "AI." The second is the definition of "high risk." The Commission created without explanation (spare a hint in Article 7 on how it would add to the list) eight "high risk" categories for which heavy burdens would apply. By burdening each category equally, the Act fails to recognize important differences both between these categories and within them. For instance, AI used to maintain utilities ("Management and operation of critical infrastructure") is treated the same as AI used to evaluate a person's access to public services ("Access to and enjoyment of essential private services and public services and benefits"), despite these categories' different risk profiles. Similarly, within the law enforcement category, AI used to detect deep fakes is treated as equally risky as AI used to assess the length of criminal sentences. This is not a risk-based approach.

The best solution to address these nuances would be implementing or updating sectoral regulations. For instance, policymakers ought to bolster antidiscrimination legislation if outcomes are discriminatory. The priority should be the public sector, where a lack of accountability and liability has led to some of the most pernicious cases of AI misuse, and where misuse inflicts additional (democratic) harm.

Poor legislation is not better than no legislation. If the EU cannot fix the AI Act's scope problem, it should revisit the legislation entirely. Failing to do so would invite a period of grave AI deterrence in Europe.

REFERENCES

1. Europe's Digital Decade: Digital targets for 2030 (2022) European Commission - European Commission, accessed October 21, 2022, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en; A European approach to Artificial Intelligence Shaping Europe's digital future, accessed October 21, 2022, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.
2. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.
3. B.J. Koops, "Should ICT Regulation Be Technology-Neutral?" papers.ssrn.com (2006), accessed January 25, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=918746.
4. Mikolaj Barczentewicz and Benjamin Mueller, "More Than Meets The AI: The Hidden Costs of a European Software Law," Center for Data Innovation (2021), https://www2.datainnovation.org/2021-more-than-meets-the-ai.pdf.
5. "Live data from OECD.AI partners," OECD, accessed October 21, 2022, https://oecd.ai/en/data?selectedArea=investments-in-ai&selectedVisualization=vc-investments-in-ai-by-country; Daniel Castro and Michael Mclaughlin, "Who Is Winning the AI Race: China, the EU, or the United States? -2021 Update," https://www2.datainnovation.org/2021-china-eu-us-ai.pdf.
6. Peter Stone et al, "Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence, *Stanford University* (2016), http://ai100.stanford.edu/2016-report.
7. Azeem Azhar, "State of AI: How did we get here, and where are we going next?" *Medium* (2019), https://medium.com/hackernoon/state-of-ai-how-did-we-get-here-and-where-are-we-going-next-2e2196049547.
8. Proposal for a Regulation on Artificial Intelligence.
9. Barczentewicz and Mueller, "More than Meets the AI"
10. European Commission, "Impact Assessment of the Regulation on Artificial intelligence," European Commission (2021), https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-

intelligence; European Commission Directorate-General for Communications Networks, Content and Technology, "Study supporting the impact assessment of the AI regulation," European Commission (2021), https://digital-strategy.ec.europa.eu/en/library/study-supporting-impact-assessment-ai-regulation.

11. Retail Technology Innovation Hub, "Co-op expands age estimation technology trial with Diebold Nixdorf and Yoti," *Retail Technology Innovation Hub* (2022), https://retailtechinnovationhub.com/home/2022/6/15/co-op-expands-age-estimation-technology-trial-with-diebold-nixdorf-and-yoti; Ivan Mehta, "Instagram expands AI-powered age verification program to India and Brazil," *TechCrunch* (2022), https://techcrunch.com/2022/10/13/instagram-expands-ai-powered-age-verification-program-to-india-and-brazil/?guccounter=1.

12. Viso, "Eye Gaze Detection," viso.ai (2022), accessed December 15 2022, https://viso.ai/application/eye-gaze-detection/.

13. "Active traffic management: adaptive traffic signal control,"AccessScience (2014), https://www.accessscience.com/content/briefing/aBR0106141; Kitae Jang, Koohong Chung, and Hwasoo Yeo, "A Dynamic Pricing Strategy for High Occupancy Toll Lanes," Transportation Research Part A: Policy and Practice 67 (2014): 69–80; A. G. Sims and K. W. Dobinson, "The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits" (1980). IEEE Transactions on Vehicular Technology, 29(2), pp.130–137.

14. Energy Robotics, "Robotic Inspection in Power Utilities,"*Energy Robotics* (2022), https://www.energy-robotics.com/industries/robotic-inspection-power-utilities

15. Jill Burstein et al, "Automated Scoring Using a Hybrid Feature Identification Technique," *Association of Computational Linguistics* (1998), https://www.ets.org/Media/Research/pdf/erater_acl98.pdf; Peter Stone et al, "Artificial Intelligence and Life in 2030."

16. Paul James "3D Mapped HTC Vive Demo Brings Archaeology to Life," Road to VR (2015), https://www.roadtovr.com/3d-mapped-htc-vive-demo-brings-archaeology-to-life/.

17. Trisha Patel et al., "Resume Sorting using Artificial Intelligence," *International Journal of Research in Engineering, Science and Management* (2019), https://www.ijresm.com/Vol.2_2019/Vol2_Iss4_April19/IJRESM_V2_I4_117.pdf; Mohammad Hossein Jarrahi and Will Sutherland, "Algorithmic Management and Algorithmic Competencies: Understanding and Appropriating Algorithms in Gig work," (2019), https://link.springer.com/chapter/10.1007/978-3-030-15742-5_55.

18. Stephen Jones, "I went to a metaverse recruitment fair with 30 companies and 200 attendees. The avatars were creepy but I liked it — take a look around," *Business Insider*, (2022), https://www.businessinsider.com/metaverse-job-fair-avatar-creepy-careers-2022-2?r=US&IR=T#; Hyundai, "New Employees, Come to Metaverse!" *Hyundai Motor Group* (2022), hyundaimotorgroup.com/story/CONT0000000000001842

19. Eurofound, "Impact of digitalisation on social services," Publications (Luxembourg: Office of the European Union, 2020), https://www.eurofound.europa.eu/publications/report/2020/impact-of-digitalisation-on-soci al-services; William Eichler, "Your authority on UK local

government - Pepper the robot to take on social care tasks," Localgov.co.uk (2017), https://www.localgov.co.uk/Pepper-the-robot-to-take-on-social-care-tasks/44080; Kai Chew et al., "Digital health solutions for mental health disorders during COVID-19," *Frontiers in Psychiatry* (2020), 11, 898.

20. Mirjam Lisa Scholz et al., "Artificial intelligence in Emergency Medical Services dispatching: assessing the potential impact of an automatic speech recognition software on stroke detection taking the Capital Region of Denmark as case in point," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* (2022), https://doi.org/10.1186/s13049-022-01020-6.

21. Financier, "Using technology and Big Data to tackle fraud and money laundering," *Financier Worldwide* (2016), https://www.financierworldwide.com/forum-using-technology-and-big-data-to-tackle-fraud-and-money-laundering#; *International Association of Chiefs of Police,* "Tracking Sex Offenders with Electronic Monitoring Technology: Implications and Practical Uses for Law Enforcement," *International association of chiefs of Police* (2020, https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/IACPSexOffenderElecMonitoring.pdf; Rosaline Chan, "The CEO of the secretive big-data startup Palantir, which is looking to IPO this year, says he finds out about a stopped terrorist attack once a week," *Business Insider* (2019), https://www.businessinsider.com/palantir-ceo-alex-karp-interview-stopped-terror-attack-weekly-2019-1.

22. Deloitte, "The Government & Public Services AI Dossier," *Deloitte AI Institute* (2021), https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-goverment-public-services-dossier.

23. Thales, "The electronic passport in 2021 and beyond," *Thales Group* (2021), https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/passport/electronic-passport-trends; Smart Engines, "AI-driven technologies increase the efficiency of passport control with Smart ID Engine," *Smart Engines* (2021), https://smartengines.com/news-events/ai-driven-technologies-increase-the-efficiency-of-passport-control-with-smart-id-engines-installed-software/.

24. Peter Stone et al, "Artificial Intelligence and Life in 2030"

25. John Markoff, "Armies of Expensive Lawyers, Replaced by Cheaper Software," *The New York Times* (2021) https://www.nytimes.com/2011/03/05/science/05legal.html.

26. Deloitte, "The Government & Public Services AI Dossier,"

27. Startup Genome, "The Global Startup Ecosystem Report," *Startup Genome* (2022), https://startupgenome.com/report/gser2022

28. Iqbal Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science* (2021), doi:10.1007/s42979-021-00592-x.

29. Proposal for a Regulation on Artificial Intelligence.

30. Directive 2002/21 on a common regulatory framework for electronic communications networks and services, OJ L108/33. (2002).

31. Joshua New and Daniel Castro, "How Policymakers Can Foster Algorithmic Accountability," (2018), https://www2.datainnovation.org/2018-algorithmic-accountability.pdf.

32. John Billings and Richard Thornton, “Expounded in Its Primitive Purity, Excellence and Loveliness,” *Primitive Expounder, Devoted to Theoretical and Practical Religion* (1845).

33. Lighthouse Reports, “Junk Science Underpins Fraud Scores,” Lighthouse Reports (2022), https://www.lighthousereports.nl/investigation/junk-science-underpins-fraud-scores.

34. VPRO, “Hoe een verboden Excelsheet bepaalt wie verdacht wordt van bijstandsfraude,” VPRO (2022), https://www.vpro.nl/argos/lees/onderwerpen/artikelen/2022/hoe-een-verboden-excelsheet-bepaalt-wie-verdacht-wordt-van-bijstandsfraude.html.

35. VPRO, “Bijstandsfraudebingo,” VPRO (2022), https://www.vpro.nl/argos/media/luister/argos-radio/onderwerpen/2022/bijstandsfraudebingo.html.

36. Tapani Rinta-Kahila et al, “Algorithmic decision-making and system destructiveness: A case of automatic debt recovery,” E*uropean Journal of Information Systems* (2022), 31:3, 313-338, DOI:10.1080/0960085X.2021.1960905

37. Luke Henriques-Gomes “Robodebt: government to refund 470,000 unlawful Centrelink debts worth $721m,” *The Guardian* (2020), https://www.theguardian.com/australia-news/2020/may/29/robodebt-government-to-repay-470000-unlawful-centrelink-debts-worth-721m#.

38. Paul Henman, Paul, “The computer says ‘DEBT’: Towards a critical sociology of algorithms and algorithmic governance,” in *Data for Policy,* London: Zenodo, 2017), https://doi.org/10.5281/zenodo.884117

39. Amie Meers et al, “Commonwealth Ombudsman Lessons learnt about digital transformation and public administration,” Centrelink’s online compliance intervention Commonwealth Ombudsman Office project team (2017), https://www.ombudsman.gov.au/__data/assets/pdf_file/0024/48813/AIAL-OCI-Speech-and-Paper.pdf.

40. Rebecca Turner, “Robodebt condemned as a ‘shameful chapter’ in withering assessment by federal court judge,” *ABC News* (2021), https://www.abc.net.au/news/2021-06-11/robodebt-condemned-by-federal-court-judge-as-shameful-chapter/100207674.

41. Richard Glenn, “Centrelink’s automated debt raising and recovery system: A report about the Department of Human Services’ online compliance intervention system for debt raising and recovery,” (Canberra: Commonwealth Ombudsman, 2017), https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf.

42. Jordan Hayne and Matthew Doran, “Government to pay back $721m in Robodebt, all debts to be waived,” *ABC News* (2020, https://www.abc.net.au/news/2020-05-29/federal-government-refund-robodebt-scheme-repay-debts/12299410.

43. Hodan Omaar, “AI Alarmism is a Distraction,” Center for Data Innovation (2022), https://datainnovation.org/2022/10/ai-alarmism-is-a-distraction/.

44. Proposal on the Regulation on Artificial Intelligence.

45. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending certain

Union legislative acts - General approach (2022); Luca Bertuzzi, "Artificial Intelligence definition, governance on MEPs' menu," *Euractiv* (2022), https://www.euractiv.com/section/digital/news/artificial-intelligence-definition-governance-on-meps-menu/.

46. Sebastian Schuchmann, "Analyzing the Prospect of an Approaching AI Winter," *Research Gate* (2019), https://www.researchgate.net/publication/333039347_Analyzing_the_Prospect_of_an_Approaching_AI_Winter

47. Peter Stone et al, "Artificial Intelligence and Life in 2030"; Michael L. Littman et al, "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report," *Stanford* University (2021), http://ai100.stanford.edu/2021-report.

48. Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters* (2018), https://www.reuters.com/article/amazon-com-jobs-automation/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idINKCN1MK0AH?edition-redirect=in; Parmy Olson, "The Algorithm That Helped Google Translate Become Sexist," *Forbes* (2018), https://www.forbes.com/sites/parmyolson/2018/02/15/the-algorithm-that-helped-google-translate-become-sexist/?sh=6d6cbf127daa; Alex Hern, "Flickr faces complaints over 'offensive' auto-tagging for photos," The Guardian (2015), https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos; GOV.UK, "Report commissioned by CDEI calls for measures to address bias in police use of data analytics," *GOV.UK* (2019), https://www.gov.uk/government/publications/report-commissioned-by-cdei-calls-for-measures-to-address-bias-in-police-use-of-data-analytics; BBC News, "Facial recognition fails on race, government study says," *BBC News* (2019), https://www.bbc.com/news/technology-50865437.

49. Anja Lambrecht and Catherine Tucker, "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," *Management Science* (2019), 65 (7). pp. 2966-2981. ISSN 0025-1909

50. Katyanna Quach, "Meta's AI internet chatbot starts spewing fake news," *The Register* (2022), https://www.theregister.com/2022/08/14/in_brief_ai/; The Guardian, "Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot," *The Guardian* (2016), https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot.

51. Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011).

52. James Manyika and Jake Silberg, "Notes from the AI frontier: Tackling bias in AI (and in humans)," *McKinsey* (2019), https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.pdf.

53. Jon Kleinberg et al, "Human Decisions and Machine Predictions," *The Quarterly Journal of Economics, Volume 133, Issue 1,* (February 2018), 237–293, https://doi.org/10.1093/qje/qjx032.

54. David Rolnick, "Tackling Climate Change with Machine Learning," Arxiv (2019), https://arxiv.org/pdf/1906.05433.pdf; Rohit Sharma, "A

Systematic Literature Review on Machine Learning Applications for Sustainable Agriculture Supply Chain Performance," *Computers & Operations Research (*2020), doi:10.1016/j.cor.2020.104926; Thorsten Kurth, "Exascale Deep Learning for Climate Analytics," *Arxiv* (2018), https://arxiv.org/abs/1810.01993.

55. Sebastian Bordt et al, "Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts," *Arxiv* (2022), https://arxiv.org/abs/2201.10295; Patrick Grady, "The EU Should Clarify the Distinction Between Explainability and Interpretability in the AI Act,". The Center for Data Innovation (2022), https://datainnovation.org/2022/08/the-eu-should-clarify-the-distinction-between-explainability-and-interpretability-in-the-ai-act/.

56. Proposal for a Regulation on Artificial Intelligence.

57. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach (2022).

ABOUT THE AUTHOR

Patrick Grady is a policy analyst at the Center for Data Innovation, focusing on AI and content moderation. Previously, he was the project lead at the Internet Commission and worked in strategy at the European Institute of Innovation and Technology. Patrick holds masters in Philosophy and Political Science.

ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation studies the intersection of data, technology, and public policy. With staff in Washington, London, and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as open data, AI, and the Internet of Things. The Center is part of the Information Technology and Innovation Foundation (ITIF), a nonprofit, nonpartisan think tank.

Contact: info@datainnovation.org

datainnovation.org