Historical antecedents for AI governance

by Jeff Sheng and the CASBS team

As noted in our introductory AI level-set essay, the history of artificial intelligence dates back almost seventy years and includes different approaches that characterize the progress of both philosophical and technical advances in AI. At this point in its evolution, it is no longer a question of *if* AI should be governed, but *how.*

Instead of outlining specific options, this paper uses historical antecedents as frameworks for thinking about AI and Governance. We focus on the following cases: regulatory safety, auditing practices, and medical ethics.

We highlight both similarities and differences between the cases presented and AI today. By no means are the historical examples representative of all the issues at hand, but a snapshot of what we think might be most pertinent. The cases are also drawn from the history of the United States, reflecting the authors' experience and knowledge. But we encourage readers to draw on other historical antecedents that can be useful in framing AI governance.

1) Regulatory Safety

There have been multiple cases in history where, after a period of rapid technological and societal advancements, effective governance systems at a federal level in the United States have taken a long time to implement, particularly with issues regarding safety, such as building codes and infrastructure. Here we focus on two: The Food and Drug Administration (FDA) and the National Highway Traffic Safety Administration.

In many ways, today's conceptions of artificial intelligence mirror these cases in the way that most (including those who use AI in their applications) are not fully aware of the potential harm that can be caused by them. An analogous "black box" currently exists for AI, in the way those in the 19th Century were unsure of what happened in meat processing plants, while also not realizing there was a problem with the science and safety behind new drugs. Our enjoyment of AI algorithms – in the pleasure of our automated Instagram Feeds and seeing only curated comments in Facebook we agree with – is analogous to car manufacturers who downplayed safety concerns to build faster and more attractive cars consumers wanted to drive. Our exuberance for the promise of self-driving cars has allowed certain car manufacturers such as Tesla to bypass standard safety protocols in the testing of autonomous vehicles. Such similar blind spots previously hindered effective governance of food, drugs, and transportation for many decades.

While almost everyone agrees today that the existence of The Food and Drug Administration (FDA) is a necessity, in the 19th Century, the wide scale societal adoption, acceptance, and codification into law, of regulation over the manufacture and sale of food and drugs, took decades to achieve. In the early 19th century, most Americans were unaware of the safety issues involving the distribution of food and drugs that arose as a product of the Industrial Revolution (Young 1989). Initial attempts at reform included the Drug Importation Act of 1848 and the creation of The Board of Health by Congress in 1879. But it was not until 1906, when the Pure Food and Drug Act was signed by President Theodore Roosevelt, that our modern conception of the FDA became law.

Change in public opinion is often credited to journalists working at the time who exposed unsanitary food preparation conditions, with the most famous example being Upton Sinclair's "The Jungle." At the same time, food and drug industries heavily resisted legislation, fearing that new laws would result in a heavy loss in profits (Young 1989). Legislators eventually gave into public pressure, but it still took 27 years to adopt the 1906 laws, after much persuasion by the media and the medical community that food and drug manufacturers needed stronger regulations, overriding the deregulatory wishes of industry.

Car safety regulations similarly took many decades to become accepted societal practices. Although cars became more widely available in America in the 1920s, safety standards, enforcement, and regulations lagged. Ralph Nadar's best-selling book *Unsafe at Any Speed*, first published in 1965, assailed the automobile industry for prioritizing consumer comfort and industry profits and highlighted their disregard for safety efforts, such as the lack of seatbelts in all vehicles. As a response to the book and public outcry, U.S. Senate hearings resulted in the creation of the Department of Transportation and the precursor to the National Highway Traffic Safety Administration in 1966. In 1968, the first Federal Motor Vehicle Safety Standards took effect, which included mandating that all new cars be made with left and right front front-seat shoulder belts. Widespread usage of seatbelts, however, did not occur until each state began mandating them, beginning with New York in 1984, with 48 other states following suit by 1995. Similar to the case with the food and drug industries, public efforts for change had to overcome intense lobbying by the automobile industry who worried such laws would hinder profits.

When we look at both the FDA and the National Highway Traffic Safety Administration, these governance systems took decades to become established, and even more time for the public to understand and accept current laws as beneficial. It should also be noted that in both cases, immense public awareness and pressure was created by consumer advocates, journalists, and the medical profession, who were often ahead of government in addressing the main areas of harm. Notably, in the past few years we have had an increased public awareness of AI and its potential harm thanks to work done by Cathy O'Neil, Safiya Noble, and Kara Swisher, among many others, and similarly, there have been increased calls for investigation by Congress as well as intense resistance by tech companies over regulation.

2) Financial Auditing Practices

A set of accounting principles, standards, and procedures now referred to as Generally Accepted Accounting Principles (GAAP) have in some form or another existed since 1933 and have been widely accepted as norms in the financial world. One reason for their implementation was because of the Great Depression, as many faulted untruthful financial reporting practices as a contributing factor of the stock market crash of 1929.[1] Since then, the accounting standard in the United States has been adopted by the U.S. Securities and Exchange Commission (SEC), with guidance from the Financial Accounting Standards Board (FASB) along with other accounting agencies, which set forth a set of principles and guidelines for companies to follow in their reporting and auditing.

Fundamentally, the aim of GAAP is to create consistent accounting and reporting standards, which allow prospective and existing investors clear evaluative measures in assessing their investments.[2] These include a set of basic accounting principles and specific guidelines pertaining to ethical reporting; for example, companies are expected to accurately state their assets, liabilities, risks, and conflicts of interest. Failure to do so results in harsh penalties by the regulatory agencies in charge of enforcing these standards.

In recent years, the idea of algorithmic auditing has become more widely discussed in both academia and industry (Guszcza 2018). In this work, two methodologies are often used that mean slightly different things: algorithm audits and algorithmic impact assessment (Ada Lovelace Institute 2020). Algorithm audits often include a call for greater regulation and increased awareness into the types of automated systems that are used, while also deploying a so-called "black box algorithmic audit" for commercial AI systems (Centre for Data Ethics and Innovation 2020; German Data Ethics Commission 2019). Algorithmic impact assessments, on the other hand, examine the outcomes associated with the usage of algorithms, while also posing the more philosophical question of "What is fairness?" (Sandvig at al. 2014; Sánchez-Monedero, Dencik, and Edwards 2020). Over the past few years, there has been a significant amount of research using both methodologies, but challenges still remain. These include bias in benchmark metrics (particularly in systems that examine race), access to the systems being audited, a lack of public pressure, and hostile corporate reactions (Buolamwini and Gebru 2018; Mikians et al., 2012; Phillips et al. 2011; Diakopoulos 2011).

The historical case of financial regulations shows how a widely used and accepted auditing system by corporations and industry could be successful. There is a decades-long consensus that companies must follow GAAP protocols, and not doing so incurs significant penalties. One important difference between the two systems that should be noted is that, unlike current auditing approaches to AI, there is less concern with outcomes associated with the financial industry or concern with overall fairness. For example, GAAP generally

[1] https://corporatefinanceinstitute.com/resources/knowledge/accounting/gaap/
[2] https://www.accounting.com/resources/gaap/

does not ask the question of overall fairness (such as whether or not the system discriminates against certain racial groups), the way questions about AI are being raised today.

3) Bioethics Ethics and Harm

A third historical framework that we can use to understand approaches to AI and governance is the example of bioethics and harm. In current usage, it is widely understood that for medical practice and biological research to be "ethical," it respects four general principles: autonomy, justice, beneficence, and non-maleficence. Autonomy means that a patient (or subject) has the right to make a fully informed decision; justice is the idea that unfairness is avoided and benefits are distributed equally in society; beneficence implies a moral obligation to promote the well-being of people, societies, and the planet: and non-maleficence requires that a procedure or technology not harm others (Cook-Deegan 2020; Stanford Encyclopedia of Philosophy 2013).

The application of these principles extends to medical practice and biomedical research, and heavy oversight exists to ensure that these principles are followed. One of the primary ways this is enforced is through Institutional Review Boards (IRBs), that consist of independent review panels and for the last few decades are now codified into US regulation and widely required globally (Grady 2015).

The current applications of AI in technological systems does not have the same level of oversight that we find in the biomedical and academic research space. Each company establishes its own guidelines for such research and the extensive "terms and conditions" that users are forced to opt into, are sometimes used to suggest "autonomy" and willing compliance have been satisfied (Meyer 2014). Moreover, only research that is made publicly available comes under public scrutiny, where most experiments and technological products created in companies are kept private. Because of this, proposed solutions, such as mandating all collaborative academic research with industry receive IRB approval, would only reach a limited subset of AI applications (Relias Media 2017). Another idea would be to create an industry-level IRB that all companies would need to report to, as well as to establish firm ethical standards based on biomedical ethics that could be enforced by laws and governmental regulation.

4) Discussion

We emphasize that none of these are perfect analogies, and that there will always be gaps in logic that make it impossible for the past to predict exactly what the future should be. Moreover, the systems that we highlight have their own flaws: problems arise with our food/drug supply, people are still hurt and injured in car accidents, and despite regulations, large-scale financial crashes such as the Great Recession of 2008 still occur.

There are however some key takeaways. One is that regulations have historically taken a lot of time, public pressure, and trial and error to become accepted practices. Moreover, adoption of new norms does not usually occur at a linear rate, and can also be highly variable and dependent on the attitudes and social networks individuals are in.[3] This being said, academic experts, consumer advocates, and journalists often do have a strong role to play in advancing public awareness of the issue. In many cases, this occurs through negotiation among the interested actors and organizations where there is participatory buy-in from those most affected. Once established in this way, regulatory standards such as GAAP have shown to work, particularly paired with auditing standards and public trust. And finally, codes of biomedical ethics are already in place, and can be extended to and applied towards the governance of AI. We also stress that there are many other historical antecedents not mentioned here that can be applied in various ways towards our thinking of the application and regulation of AI and hope that these ideas presented here help guide our discussion and thought generation processes in this area.

We offer these illustrative questions to prime discussion.

1. Recent books, articles, and documentary films have attempted to help individuals and organizations overcome "blind spots" about the need for regulations or other AI governance arrangements. Have these been sufficiently effective? What other blind spots exist? What other ways might help raise needed awareness?
2. On analogy with the role of chartered accountants in the historical antecedent of financial auditing, is there a societal need for a dedicated profession of algorithm auditors with courses of professionalism, professional standards of practice, continuing education requirements, and so on?

[3] A current example that illustrates this well is the issue of masks and the Covid-19 pandemic. In the United States, the adoption of mask wearing occurred very quickly by certain groups, but it also became politicized by others. Thus, while one part of society took little time to follow new mask wearing regulations, in other areas, regulations were either not established, or heavily resisted by some.

References

Ada Lovelace Institute. 2020. "Tools for Examining the Black Box." Available at: https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81:1–15, 2018. Available at: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Centre for Data Ethics and Innovation. 2020. "Online targeting: final report and recommendations." *Gov.uk.* Available at: www.gov.uk/government/publications/cdei- review-of-online-targeting.

Cook-Deegan, Robert. 2020. "What are the Basic Principles of Medical Ethics?" Stanford in Washington Seminar: How Decisions are Made About Health Research and Health Policy. Available at: https://web.stanford.edu/class/siw198q/

Diakopoulos, Nicholas. 2011. "Accountability in Algorithmic Decision Making." Communications of the ACM, Feb 2016. Vol 59, No. 2. Available at: http://www.nickdiakopoulos.com/wp-content/uploads/2016/03/Accountability-in-algorithmic-decision-making-Final.pdf

German Data Ethics Commission. 2019. "Opinion of the Data Ethics Commission." Bmjv.de. Available at: www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf

Grady, Christine. 2015. "Institutional Review Boards: Purpose and Challenges." U.S. National Library of Medicine National Institutes of Health. Nov; 148(5): 1148-1155. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4631034/

Guszcza, James et al. 2018. "Why we need to audit algorithms." *Harvard Business Review.* Nov 28, 2018. Available at: https://hbr.org/2018/11/why-we-need-to-audit-algorithms

Meyer, Michelle. 2014. "How an IRB could have legitimately approved the Facebook experiment – and why that may be a good thing." Harvard Law Bill of Health post. Available at: https://blog.petrieflom.law.harvard.edu/2014/06/29/how-an-irb-could-have-legitimately-approved-the-facebook-experiment-and-why-that-may-be-a-good-thing/

Mikians et al., 2012. "Detecting price and search discrimination on the Internet." *Hotnets* '12, October 29–30, 2012, Seattle, WA, USA. Available at:

https://www.researchgate.net/publication/232321801_Detecting_price_and_search_discrimination_on_the_Internet/citation/download

Nadar, Ralph. 1965. *Unsafe at Any Speed*. New York: Richard Grossman.

Noble, Sofiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: NYU Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Crown Books.

Phillips et al. 2011. ACM Transactions on Applied Perception, Feb 2011, Article 14. "An other-race effect for face recognition algorithms." Available at: https://dl.acm.org/doi/10.1145/1870076.1870082

Relias Media. 2017. "IRB collaborations with tech companies could mean what to the IRB?" Available at: https://www.reliasmedia.com/articles/141587-irb-collaborations-with-tech-companies-could-mean-what-to-the-irb

Sánchez-Monedero, Javier, Line Dencik, and Lilian Edwards. 2020. "What does it mean to 'solve' the problem of discrimination in hiring?: social, technical and legal perspectives from the UK on automated hiring systems." *Conference on Fairness, Accountability, and Transparency*, p33–44. [online] Barcelona: ACM. Available at: https://arxiv.org/pdf/1910.06144.pdf

Sandvig, Christian et al. 2014. "Auditing algorithms: research methods for detecting discrimination on internet platforms." *Pre-conference on Data and Discrimination at the 64th annual meeting of the International Communication Association*, p1–23. Available at: http://social.cs.uiuc.edu/papers/ pdfs/ICA2014-Sandvig.pdf

Stanford Encyclopedia of Philosophy. 2013. "The Principles of Beneficence in Applied Ethics." Available at: https://plato.stanford.edu/archives/win2013/entries/principle-beneficence/

Young, James H. 1989. *Pure Food: Serving the Federal Food and Drugs Act of 1906.* Princeton: Princeton University Press.