

For Safe AI Tomorrow, Fiduciary Duties for Big Tech Today

By Sebastian Benthall (Information Law Institute, NYU School of Law)

For Safe AI Tomorrow, Fiduciary Duties for Big Tech Today

By Sebastian Benthall (Information Law Institute, NYU School of Law)

The best way to ensure that future artificial intelligences are safe and friendly towards humanity is to regulate technology companies today. Many of the sectors that we consider most trustworthy – such as our legal, health, and financial advisors – are under legal obligations to act with loyalty and care towards their clients. These obligations are known to lawyers as *fiduciary duties*. Now, regulators and scholars are considering whether fiduciary duties should be applied to on-line service providers, digital assistants, and other computational systems. Such a legal rule would be the best way to ensure that the major drivers of AI research – the large technology firms – develop AI in a way that benefits humanity today and on into the future.

AI ethics is a broad research field spanning academic centers, non-profit research and advocacy organizations, corporate research labs, and professional services companies. One significant thread of this research, AI Safety, has resonated with philanthropists that take a long term view. They believe that future human lives are as ethically valuable as present ones, and so the highest priority ethical causes are those that reduce humanity's existential risks. Organizations working from or funding this perspective include the Future of Life Institute, Cambridge University's Centre for the Study of Existential Risk, Oxford University's Future of Humanity Institute and Global Priorities Institute, 80,000 Hours, Open Philanthropy, and the FTX Future Fund. These organizations tend to agree that “advanced AI”, sometimes imagined as an autonomous Artificial General Intelligence (AGI) capable of independent scientific advancement and manipulation of humans, poses a potential existential threat to humanity. The proper response to such an existential threat, the logic goes, is to do something to avert it today.

In the messy field of AI research, this long term philanthropic interest meets and competes with many other immediate and pragmatic priorities. Currently, the leaders in advanced AI are large technology companies. New laws and regulations, especially in major jurisdictions, create powerful incentives for these companies to steer AI development towards compliance. For example, much of AI Fairness research is motivated by the practical problem of compliance with nondiscrimination laws. A self-driving car that causes an accident due to the negligence of its manufacturers will render its manufacturers liable under tort law. When the EU's General Data Protection Regulation

passed in 2016, compliance with its quite vague principles and directives became a major priority for every international tech company.

Neither the reactive, short term, compliance-oriented perspective, nor the futurist, long term perspective on artificial intelligence safety is complete. Rather, the future of artificial intelligence depends on innovation pathways guided by legal rules established today. The corporate organization is a form of artificial general intelligence that is over a century old. The path to autonomous AGI is quite likely through the automation of activities that companies currently perform with a hybrid human-computer supply chain. Hence, AI can be steered through law that binds corporate behavior.

The long term interest in AI has motivated computer science and engineering research into the problem of AI Alignment: how can we make sure artificially intelligent systems understand and serve our interests? How can AI be guaranteed not to deviate from the goals that it has been given? Many of today's consumers have the same questions for their ISPs, smartphones, digital services, and IoT devices. There is a legal solution to this anxiety about the alignment of computational systems with the interests of their users: fiduciary duties.

The European Union has recently passed the Data Governance Act, which creates a new legal category of data intermediaries with fiduciary duties. Fiduciary duties for online services have appeared in recent legislative proposals in the U.S., at the Federal and State levels, most notably the Data Care Act of 2021, a bill that is currently being reviewed by the U.S. Senate Committee on Commerce, Science, and Transportation. This bill would establish fiduciary duties for online service providers, including a duty of loyalty that forbids the company from using personal data in any way that would be a detriment to the end user. While not framed directly in terms of AI, such a law would make it in the interest of every major big tech company to research and develop technologies that were aligned with the interests of their users.

In the United States, outside of a few regulated sectors like health and finance, personal data is regulated via notice and consent: the presentation of a contract to users, on which they then sign off before using a connected service or product. It is well known that most users are unable to read these contracts. Under stronger data protection regimes, such as the European Union's General Data Protection Regulation and the California Consumer Privacy Act, there are stronger requirements on notices which are designed to better inform consumers about the purposes to which their data will be used. However, these "purposes" are often defined so broadly that they

offer consumers little control over whether a company is acting in a way that is aligned by their interests. Fundamentally, the complexity of the technical service, and expertise (not to mention bargaining power) of the service provider mean that contracting will always be either incomplete or unfavorable from the consumer's perspective. Fiduciary duties are used in other sectors that would otherwise have the similar problem. These duties involve an obligation to act in the best interests of the principal that users cannot be deceived or pressured to sign away.

Technology policy is a hotly contested issue today with involvement from all corners. The profits of the largest companies are hanging in the balance, as well as the power of the major political parties and their leaders. Technology law spans issues of privacy and data protection, antitrust and competition, intellectual property, and cybersecurity. Any of these issues may have path-dependent impact on the kind of AI we see in the future. A pragmatic approach to technology policy requires a thorough understanding of the present-day political economy driving new regulations as well as the long term interests of humanity. What's needed is a practical political strategy for steering the future of AI through law. A good start would be the passage of information fiduciary laws that would empower regulators to define and refine what it means for technology to work in the best interests of its users.



Sebastian Benthall
DLI Associate
Research Fellow @ Information Law Institute
NYU School of Law
spb413@nyu.edu
[More Info >](#)