

Data Trusts in AI Governance

14th April 2025

Problem and Scope

Debates about the use and misuse of artificial intelligence constitute a driving force in the tech policy sphere as the technology itself becomes increasingly widely adopted. Beneath promises that AI will save the world, and concerns that AI will destroy it, lies the oft-ignored raw material that powers the whole operation: the millions of articles, photographs, books, paintings, and YouTube comments comprising the terabytes upon terabytes of training data which form the basis of any machine learning model.

The legal status of this data is very much ambiguous, something most recently highlighted by a lawsuit filed by the New York Times against ChatGPT maker OpenAI alleging infringement of their copyrighted materials which were included in the model's training set.¹ However, the essential question of "who knows, who decides, who decides who decides?"² sits at the heart of AI and the tech industry more generally. The status quo is one in which a small number of behemoths, exerting unprecedented influence on their users through AI systems, control data unilaterally. Crucially, it is through the aggregation of data that these firms grew to dominate the AI market: any individual data point has an infinitesimally small effect on the output of a model, but put together, sufficient data points allow for incredibly sophisticated predictive abilities. A resolution to the current state of affairs, therefore, requires a resolution to these two seemingly conflicting truths: that individuals ought to have a say in how their data is used, but that data only has power on a scale beyond any one individual user.

Findings

The main policy proposal put forward by this document is the creation of *data trusts* for AI training data. Generally speaking, trusts are legal entities entrusted to act on behalf of the owner of some asset in order to manage it for some purpose. In the context of data, this involves beneficiaries entrusting an organisation with access or rights to data, where this organisation is legally bound by fiduciary duty to act in the undivided interest of these beneficiaries³. For example, this could involve a group of newspapers (the beneficiaries) all entrusting the rights to their articles (data) to a third-party trustee who negotiates with OpenAI on their behalf. The larger the group, the greater the value of their collective data, and the greater the bargaining power of the trust. Since data exhibits increasing returns to scale, the value of the collected data far exceeds the sum of the values of the constituent data points, giving the trust a significantly better position in any negotiation than any of the individual trustees.

¹ Allyn, B. (2025) ['The New York Times' takes OpenAI to court. ChatGPT's future could be on the line](#), NPR

² Zuboff, S. (2019) *The Age of Surveillance Capitalism*, Profile Books

³ Mozilla Insights et al. (2020) [Shifting Power Through Data Governance](#). Mozilla Foundation.

This framework allows for the creation of a diverse market of public and private data trusts, each with their own goals, values, and priorities, which reflects the diversity in the desire for individuals' use of their data.⁴ One trust might focus on maximising financial returns for users, while another might allow free access to data for causes deemed socially beneficial. As a consequence, data trusts support individuals in their desire to have their data be used in ways they see fit, while taking advantage of the increasing returns to scale for data so that they can more effectively bargain in the interests of their beneficiaries. They have the additional advantage of redressing information asymmetries between individuals and AI companies, who may collect and use data in ways that the individual does not fully understand. This has the potential to allow for technically sophisticated, third-party auditing of AI services, carried out by trusts in the interests of their users, in the same way that a hedge fund might do due diligence on a company on behalf of its investors. Finally, they level the playing field between small and large AI companies, by providing pools of data that are equally accessible to all participants in the market.

Recommendations

Requirement that AI Companies Negotiate with Trusts

For certain classes of data, require that AI companies negotiate with trusts in order to gain access to the data for the purposes of training. The bargaining power of data trusts is premised on their ability to deprive AI companies of large amounts of valuable data if they are unwilling to make certain concessions; if these companies are able to circumvent trusts to use to their beneficiaries' data, the trusts become meaningless. This is most easily accomplished for data with existing legal protections, like copyrighted newspaper articles or medical records (the MIDATA cooperative provides a good example of the latter)⁵.

Creation of Public Data Trusts on a Local Level

Data trusts can be (and have been) created by municipalities to protect data on a local level. The European Commission's DECODE project, for example, has empowered cities across the continent to pool constituents' data to ensure that its use is geared towards social benefit.⁶ The creation of a data trust in a medium-sized city in California could therefore provide a fascinating test-case for this proposal, while also inspiring the creation of similar trusts in the private sector.

Conclusion

Data trusts constitute a way to centre the creators and subjects of data in decisions about their use, simultaneously fostering innovation by providing centralised pools of data that are easily accessible to AI startups, while also creating a diverse market of independent security researchers with the ability to meaningfully investigate AI products and services.

⁴ Delacroix, S., Lawrence, N. D. (2019) [Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance](#). International Data Privacy Law, 2019, Vol. 9, No. 4.

⁵ [Home | MIDATA](#)

⁶ Bass, T., Sutherland, E., Symons, T. (2018) [DECODE: Reclaiming the Smart City](#). Nesta.