Applying Torts to Juridical Persons: Corporate and AI Governance

Aaron D. Tucker [1]

Abstract

In theory, tort law provides a general structure for identifying and addressing a wide variety of concrete harms in society, and could provide a mechanism to address the harms of deployed AI systems. However, even in contemporary non-AI contexts remediations for many torts such as those involving privacy violations are often difficult to obtain in practice. In other domains specially-crafted legislation with specific liabilities and rules succeed at compelling companies to implement specific procedures. This essay draws parallels between problems in extending tort law to corporate governance and problems in extending tort law to AI governance.

1. Introduction

Corporations are the first artificial intelligences, built on top of people and procedures instead of silicon and computation. Common law has gradually developed to handle problems and disputes between people over time, and provides a flexible framework for handling and adjudicating concrete harms. While legal personhood allows companies to own property, contract, and be held liable, corporations differ from natural persons in many ways that make the application of tort law substantially more difficult than its application to natural persons. These dis-analogies make it more challenging to remediate harms even in cases that are already covered by existing torts. As a result, many risks are managed or remediated through special-purpose legislation and rules. This essay draws parallels between the difficulty in extending tort law to corporations and in extending tort law to generative AI.

[1]Department of Computer Science, Cornell University, Ithaca NY, USA. Correspondence to: Aaron Tucker <aarondtucker@cs.cornell.edu>.

Accepted to the *1st Workshop on Generative AI and Law*, colocated with the *International Conference on Machine Learning*, Honolulu, Hawaii, USA. 2023.

2. Cheap Injury, Expensive Trials

Tort law, which includes copyright law, concerns harms which are caused by acts or omissions that the defendant is held liable for. Since tort law provides accountability and liability for realized harms and injuries, it could be a general tool for addressing many of the concrete harms of AI systems. However, many technologies such as generative AI or the internet introduce complications and dis-analogies to interactions solely between people that make it difficult to apply tort law even in cases that it applies.

Consider the example of online harassment. The internet makes it easier and cheaper to communicate, but also removes important contextual information from interactions which can make prosecution considerably harder. For example, online harassment can easily cross into the intentional infliction of emotional distress (IIED) tort, but can be hard to prosecute when people are anonymous (Citron, 2014), or acting as part of a large group (Thomas et al., 2021). In contrast, in-person IIED is not anonymous, and often has witnesses. The release of personally identifying documents (doxxing) or non-consensual intimate imagery (NCII) can fall under the intrusion into seclusion tort (Citron, 2014), but repeated sharing and resharing of offending information or content can make prosecution harder. Social media makes it much easier to join a harassment campaign than would be in person (Thomas et al., 2021), and even though torts cover many of the relevant harms it is often difficult to prosecute (Citron, 2014). More generally, dual-use technologies make it easier to inflict harm without making it easier to conduct trials to remediate them.

If the courts find that common uses of Generative AI violate copyright, then Generative AI makes copyright violation much easier since it also makes it easier to make copyright violating imagery. Further, open-source Generative AI can decontextualize the training processes and data used in creating a model. If someone trains a model and shares the weights, then someone else finetunes it on a dataset and shares the weights, then an end-user may not be able to know what data was used in the training process without further technical advances in attribution.

Generative AI extends trends which are also involved in older technologies such as the internet, by making it easier and cheaper to do things or commit harms without necessar-

ily making it easier to remediate them.

3. Supplementing Torts with Regulation

One alternative to relying on tort law to remediate harms caused by new technologies is to craft special-purpose legislation for corporations which deploy a technology. This can help ameliorate the decontextualization associated with the technologically enabled mediation of interactions between groups of people, by instead consolidating the issues into compliance questions for corporations.

3.1. Liability in Online Payments

Online payment services are a domain where special purpose legislation is used in a setting where traditional torts apply. Similarly to various forms of online harassment, online payment services make it easier and cheaper both to conduct legitimate business as well as illicit purposes such as fraud, in a setting which affords more anonymity than their in-person counterparts. In many situations, these challenges are dealt with largely by assigning liability for harms to corporations involved in the transaction.

The Electronic Fund Transfer Act of 1978 established that US consumers are not liable for any charges that occur if their credit or debit card is stolen, and instead merchants or issuing banks are liable. Even though the tort of conversion of property (theft) is being carried out by a third party, the fact that merchants and issuing banks are responsible for preventing fraudulent transactions has created the industry of antifraud. Critically, this changes the timescale and unit economics of avoiding fraud – online payment systems are typically automated and do not involve a human processing the transaction, and the first line of defenses against fraud are typically automated as well.

Chargebacks are a process in which financial transactions are reversed in response to customer complaints, and are implemented as a form of consumer protection. For example, if a customer buys a product from a fraudulent merchant, then the customer's bank can issue a chargeback to recover the funds from the merchant's bank (Furletti & Smith, 2005). Here, the fraud tort is remediated by the customer's bank recovering the money and returning it to the customer. Online payments have very clear and traceable harms, making it easier to assign liability that incentivizes corporations to take pre-emptive steps to avoid fraudulent transactions, making credit and debit cards much safer.

This reassignment of liability helps to protect consumers even when fraudsters are not prosecuted. In this setting banks and payment processors already know how to track and transfer money between each other, and so it is possible to reverse payments. Instead of customers obtaining remediation through torts, they can instead call their bank or credit card company to get their money back. Disputes between issuing banks and merchant banks are resolved by the payment processor network, effectively operating a high-throughput small claims court rather than relying on the broader legal system for remediation.

3.2. Challenges in Liability

A key dis-analogy between corporate governance and AI governance is the fact that AI systems are much newer and develop at a faster pace, and it can be difficult to know how to implement policies which accomplish our desired goals. For example, while there are many technologies for trying to detect pirated movies online, the question of how to detect substantially similar but AI generated images is new and quite different. It is straightforward for a person to tell when two videos are the same movie, but much harder to decide if they have the same style.

Attempts to implement policies have sometimes led to issues such as large classes of users being excluded from platforms as poor risk/benefit tradeoffs, or simply removing large amounts of content due to difficulty in implementing the desired policy. For example, people in many parts of the world can have difficulty getting access to US-based financial services because of the risks imposed by Know Your Customer and Anti-Money Laundering compliance (Sanderson et al., 2018). When Tumblr wanted to remove adult content from its website, its filters were overly cautious and removed a considerable number of additional innocuous images (Foundation, 2018). How can companies develop the technologies that allow them to follow their legal requirements?

3.3. Avoiding and Remediating Harms by Corporations

Corporations also have factors which make them easier to apply torts to than individuals. For example, the fraud tort requires *intentional* deception, which can be much easier to demonstrate for corporations than for people. Corporations use a variety of internal communications to direct employees, which can lead to a paper trail which makes it easier to prove that deception is intentional. For example, Wells Fargo came under investigation for fraudulently opening bank accounts for customers without their consent, and investigators found internal communications demonstrating that executives were aware of that this was happening, leading to a $3 billion settlement (U.S. Department of Justice, 2020).

While it may be easier to understand corporate intentions than individual intentions, substantially more work is needed to audit and understand AI systems. Worryingly, while one can ask Large Language Models (LLMs) to explain their reasoning, it appears that their answers can be inaccurate (Turpin et al., 2023). One unique advantage of AI systems is that they can be copied and rerun exactly, which

makes it easier to analyze their decision-making. In contrast, a company's policies may be unevenly implemented, which makes it more difficult to understand their impacts.

Corporations are also more able to pursue legal cases than individuals, so that inter-company disputes and compliance can help develop processes and technologies that are more broadly useful. For example, the capabilities necessary to comply with avoiding legal action by well-resourced entities are similar to those needed to to comply with torts more generally. For example, the ability to identify derivative works could be necessary to comply with DMCA takedown requests from large companies which are able to pursue lawsuits to protect their IP. However, once developed, this ability to detect copyright violations could be used to identify derivative works more broadly. Tools developed to help protect a famous actor's likeness could also be used to help protect other individuals from NCII or smaller-scale copyright violations.

4. Regulating Context

Besides having unintended consequences from overly broad application of a policy, it may be difficult to develop policies which have the intended effect at all. This is particularly difficult in cases where the intended goal is difficult to operationalize computationally.

4.1. General Data Protection Regulation

The European Declaration of Human Rights recognizes a right to privacy as part of section. In 2016, the General Data Protection Regulation (GDPR) was introduced to strengthen privacy protections in Europe. The GDPR introduces several requirements, including consent for data processing for specific purposes and the right to be forgotten.

The right to be forgotten adds a very clear feature to websites and platforms – the ability to have your data be deleted as long as doing so does not interfere with other legal requirements (GDPR, 2016). Because this a concrete requirement for a specific *technical* ability to delete user's data, it is much clearer how to comply with the requirement than a less-defined requirement to explain how data is used.

In contrast, consent for a particular type of data processing is considerably more difficult to implement. While adding opt-in windows gives users more influence over their data, it does not always feel like a meaningful increase in control. Implementing consent for specific purposes is more challenging to do for online services than with in-person services, since when using traditional in-person services the customer necessarily interacts with a representative who can answer questions about company policies. Customers can ask representatives about specific use cases that they are worried about, rather than needing to read and understand a webpage which may or may not clearly describe the situation. Even if the representative does not know the answer, it is more plausible that the representative could tell their managers that customers care about the question, than that a web page with no freeform entry could aggregate this information. Information is easier to contextualize amongst people than between people and corporations, or people and AI systems, and checking "essential uses only" does not engender much trust.

This difficulty can be understood through the lens of Contextual Integrity, which analyzes privacy as the conformation of information flows to the specific norms of a given context (Nissenbaum, 2004). In this framework, the uses of information are critical. However, governing the uses of information can be challenging. While it is technically feasible to only include information in select settings, computers make it extremely easy to spread and share information in a way that takes them out of that context. Further, many online services are new and do not have clearly developed norms. Generative AI can make these problems worse, since it is not even clear on a technical level how to use some of the training data for one purpose but not another.

On the other hand, LLMs may make it easier for computers to interface with human norms. For example, Constitutional AI is an approach to finetuning instruction-following LLMs by instructing LLMs to follow human-sounding norms such as "Please choose the response that most supports and encourages freedom, equality, and a sense of brotherhood" (Bai et al., 2022; Anthropic, 2023). The ability to directly guide behavior using natural language is encouraging, but more research is needed in order to understand how well these methods agree with human normative judgments. Again, with more research and comparison to human evaluations, LLMs may make it easier to implement more contextually aware policies.

5. Conclusion

Augmenting tort law with specific procedural requirements can be an effective way to implement consumer protections and improve the ability of corporations to avoid and remediate harms. Many of the challenges involved in applying tort law to Generative AI are similar to challenges faced, though substantially more research into how to implement policies to avoid and remediate torts is needed.

Acknowledgements

This research was supported in part by NSF Awards IIS-1901168, IIS-2008139, and scholarship funding from Open Philanthropy. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

References

European convention on human rights. https://www.echr.coe.int/Documents/Convention_ENG.pdf, 1950. Adopted by the Council of Europe, Rome, 4.XI.1950.

Electronic fund transfer act, 1978. URL https://www.fdic.gov/regulations/laws/rules/6500-2515.html.

Anthropic. Claude's constitution, 2023. URL https://www.anthropic.com/index/claudes-constitution.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., Kerr, J., Mueller, J., Ladish, J., Landau, J., Ndousse, K., Lukosuite, K., Lovitt, L., Sellitto, M., Elhage, N., Schiefer, N., Mercado, N., DasSarma, N., Lasenby, R., Larson, R., Ringer, S., Johnston, S., Kravec, S., Showk, S. E., Fort, S., Lanham, T., Telleen-Lawton, T., Conerly, T., Henighan, T., Hume, T., Bowman, S. R., Hatfield-Dodds, Z., Mann, B., Amodei, D., Joseph, N., McCandlish, S., Brown, T., and Kaplan, J. Constitutional ai: Harmlessness from ai feedback, 2022.

Citron, D. K. *Hate Crimes in Cyberspace*, chapter 5, pp. 120–141. Harvard University Press, 2014. ISBN 9780674368293. URL http://www.jstor.org/stable/j.ctt7zsws7.8.

Foundation, E. F. What tumblr's ban on 'adult content' actually did, December 2018. URL https://www.eff.org/tossedout/tumblr-ban-adult-content. accessed July 7th, 2023.

Furletti, M. J. and Smith, S. The laws, regulations, and industry practices that protect consumers who use electronic payment systems: credit and debit cards. *Federal Reserve Bank of Philadelphia Payment Cards Center Discussion Paper*, 2005.

GDPR. General data protection regulation, Apr 2016. URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN. Came into effect on May 25, 2018.

Nissenbaum, H. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

Sanderson, A., Mutandwa, L., and Le Roux, P. A review of determinants of financial inclusion. *International Journal of Economics and Financial Issues*, 8(3):1, 2018.

Thomas, K., Akhawe, D., Bailey, M., Boneh, D., Bursztein, E., Consolvo, S., Dell, N., Durumeric, Z., Kelley, P. G., Kumar, D., McCoy, D., Meiklejohn, S., Ristenpart, T., and Stringhini, G. (eds.). *SoK: Hate, Harassment, and the Changing Landscape of Online Abuse*, 2021.

Turpin, M., Michael, J., Perez, E., and Bowman, S. R. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *arXiv preprint arXiv:2305.04388*, 2023.

U.S. Department of Justice, O. o. P. A. Wells fargo agrees to pay $3 billion to resolve criminal and civil investigations into sales practices involving the opening of millions of accounts without customer authorization, 2020. URL https://www.justice.gov/opa/pr/wells-fargo-agrees-pay-3-billion-resolve-criminal accessed July 7th, 2023.