Liability and Insurance for Catastrophic Losses: the Nuclear Power Precedent and Lessons for AI

Anonymous Authors[1]

1. Background

With generative AI's sudden deployment across society, there has been a sharp rise in AI related incidents (OECD, 2024). Many experts fear that more advanced AI could cause catastrophic losses (Grace et al., 2024). Agentic AI is arguably of greatest concern (Carlsmith, 2022; Chan et al., 2023), the sparks of which can arguably be seen in today's generative large language models.

With the exception of Weil (2024) which addresses existential risks, previous research on liability regimes for AI has largely ignored the possibility of catastrophic losses. This paper seeks to fill this gap, focusing on catastrophic yet still insurable risks. I reply to Weil and extend my analysis to strictly uninsurable risks elsewhere.

Research on non-catastrophic losses provides a useful baseline. It's been argued that as models become more autonomous, applying *respondeat superior*, a form of strict vicarious liability, is appropriate (Lior, 2019). Strict liability has also been recommended under Fletcher's framework of non-reciprocal harms: parties are to be held strictly liability when they generate "disproportionate, excessive risk of harm, relative to the victim's risk-creating activity" (2020). It's been suggested that AI developers or manufacturers should be the liable parties, given they're better able to absorb or spread the costs (Vladeck, 2014, 146-147).

Under welfare-maximization analyses, researchers find developers to be the least-cost avoiders (Buiten et al., 2021; 2023; Price II & Cohen, 2023).[1] This suggests developers should be assigned greater liability. Concerns of slower adoption of safer technologies, or a possible chilling effect on innovation moderate this suggestion, but scholars express uncertainty on the matter (Buiten et al., 2023, 6-7, 9)(Maliha et al., 2021). Strict liability is also often recommended for its greater legal predictability and reduction in litigation (e.g. (Vladeck, 2014, 147)). Liability would create new barriers to entry, but scholars have noted this industry will produce natural monopolies regardless (Lior, 2020; Vipra & Korinek, 2023); consumer interests are arguably best advanced through stricter regulation and liability (Narechania & Sitaraman, 2024).

2. Liability: Strict, Exclusive, and Limited

Drawing on the precedent set by the nuclear power industry, this paper argues developers of frontier foundation models should be assigned limited, strict, and exclusive third party liability for "Extraordinary AI Occurrences" (EAIO). Criteria for EAIOs might include: monetary thresholds, specific technical failure modes, threat modalities (e.g. CBRN), and persons/objects affected (e.g. critical infrastructure).

Given the baseline set by non-catastrophic losses and the importance of robustly internalizing catastrophic negative externalities, strict liability is a natural choice. Developers are even more clearly the least-cost avoiders, and, should such risks materialize, these systems will certainly have generated starkly "disproportionate, excessive risk of harm, relative to the victim's risk-creating activity." By strict liability it's meant that most defenses relating to claimant conduct or developer fault are waived (defenses based on failure to mitigate damages or wrongful causation by the claimant are not waived).

Capping liability and making it exclusive helps ensure insurability. A cap might appear to generate moral hazard, but *de facto*, liability is typically limited by the solvency of the liable party (Meehan, 355). One thus improves over this by setting a cap well above the expected solvency of liable parties and mandating carrying insurance (Logue, 1993). Furthermore, given the uncertainty and size of the risks in question, premiums would be prohibitive without a cap, running counter to the policy goal of encouraging responsible innovation. Additionally, safety investment for heavy-tail risks is likely inelastic; a cap much greater than e.g. $200B is unlikely to yield greater care taken, but simply lower activity levels. Finally, frontier AI's potential positive externalities justify keeping premiums affordable. Exclusive

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[1]While operators can play a role in mitigating harms, it's found to decrease as model autonomy increases or the information asymmetry between developer and operator increases – precisely the case we're concerned with here.

liability further reduce litigation and insurers' aggregate risk: this keeps premiums down while maximizing capacity insurers can commit to the industry. These elements (and the reasoning behind them) mirror the nuclear power precedent (Commission, 2021, sec. 2.3.8).

3. Mandatory Insurance

Mandating developers carry EAIO insurance is recommended for several reasons. First, as noted above, developers are judgment-proof to an extent: EAIO liability alone will fail to sufficiently incentivize caution. Second, the most successful developers are likely those that most underestimate the risks (a form of the winner's curse). This underestimation is likely substantial given how large and uncertain the risks are. By turning very uncertain *ex post* costs into very certain *ex ante* costs, insurance mitigates this perverse selection effect. (It's been observed that insurers charge a premium for uncertainty, a heuristic likely aimed at avoiding their own winner's curse (Mumpower, 1991)). Third, mandating insurance eliminates adverse selection and further spreads risk, thus keeping premiums down (minimizing barriers to entry). Fourth, insurers are likely to play a significant quasi-regulatory role (possibly more effectively than public institutions presently could), all while providing a testbed for future regulations.

This last bears elaborating. While some scholars tout the virtues of "outsourcing regulation" to (commercial)[2] insurers (Ben-Shahar & Logue, 2012; Baker & Swedloff, 2012), others note limits (Abraham & Schwarcz, 2022). Insurers' standard tools – risk-based pricing, underwriting, monitoring, *ex post* loss control – *mitigate* but don't *eliminate* moral hazard. For a variety of reasons (e.g. the cost of monitoring, the limits of actuarial risk-modeling, the cost of causal risk-modeling), insurers will fail to incent socially optimal levels of care with these tools alone.

However, insurers also often have incentive to reduce aggregate risk in a predictable manner and directly control insured's behavior (Abraham & Schwarcz, 2022, III). Hence we find them funding safety research, lobbying for stricter regulation, counseling on loss prevention, and enforcing private safety codes. When successful, these measures can improve loss prevention over the baseline (liability without insurance). Unfortunately obstacles often impede such efforts: safety research and stricter regulation are public goods, requiring competing insurers to coordinate; private advice and safety codes can be appropriated by insureds and competitors, cause policyholder backlash, and expose insurers to liability.

This paper claims none of these obstacles are present in the context of insuring against heavy tail risks, and we should expect a mix of causal risk-modeling, monitoring, safety research, lobbying for stricter regulation, and private safety guidance from insurers. Causal risk-modeling because actuarial data will be insufficient for such rare events (cf. causal risk-modeling in nuclear insurance underwriting and premium pricing (Gudgel, 2022, ch. 4 sec. VII)(Mustafa, 2017)). Monitoring, again due to a lack of actuarial data and the need to reduce information asymmetries (cf. regular inspections by nuclear insurers with specialized engineers (2022, ch. 4 sec. VI.C)). Safety research and lobbying regulators because insurers will almost certainly have to pool their capacity in order to offer coverage, eliminating competition and with it, coordination problems (cf. American Nuclear Insurer's (ANI) monopoly[3] on third party liability (2022, ch. 4 sec. VII.A)). Loss prevention guidance, because it can't be appropriated or drive away customers here: there will be little competition and the insurance is mandatory (cf. ANI sharing inspection reports and recommendations with policyholders (2022, ch. 4 sec. VII.A.2)).

Thus, in the context of insuring against catastrophic risks from AI, we can expect regulation via insurance to be effective – possibly more effective than what public institutions are currently capable of. (Certainly insurers' involvement is expected to complement efforts by public institutions). Government agencies are not disciplined by the market or a profit motive, and are comparatively under-resourced. Courts are purely *ex post*, and reliant on the costly, ad hoc litigation process (Ben-Shahar & Logue, 2012, 198-199).

4. Conclusion

Advances in AI appear to pose risks as large and uncertain as those that nuclear power did (and largely still does) when it was first introduced. This paper claims we can learn from the successful precedent set by that industry's liability and insurance regime. It argues that in similarly heavy-tail risk contexts, insurers are expected to play a socially beneficial quasi-regulatory role. It's *not* claimed this can *fully substitute* for regulation. Rather, it's claimed that by assigning clear liability and mandating insurance for these emerging risks, lawmakers can leverage the market's strength for aggregating information, and efficiently allocating resources to model risks and develop safe design. Regulating efficiently is easier when risks are clearer and safe design is available.

[2] That *mutuals* play this quasi-regulatory role is less contentious (Abraham & Schwarcz, 2022, sec. I.B.3). See e.g. the effectiveness with which the nuclear utilities mutual Nuclear Electric Insurance Limited (NEIL) and its sister organization, the Institute of Nuclear Power Operations (INPO) have promoted safety (Gudgel, 2022, ch. 4 sec. VII.B)(Rees, 1996).

[3] Concerns of monopolistic practices can be mitigated by regulating premiums and or encouraging self-insurance through a mutual. Cf. the nuclear utilities mutual Nuclear Electric Insurance Limited (NEIL) competing with ANI for certain lines of coverage (Gudgel, 2022, 147-148).

References

Abraham, K. S. and Schwarcz, D. The Limits of Regulation by Insurance. *Indiana Law Journal*, 98:215, 2022. URL https://heinonline.org/HOL/Page?handle=hein.journals/indana98&id=228&div=&collection=.

Baker, T. and Swedloff, R. Regulation by Liability Insurance: From Auto to Lawyers Professional Liability. *UCLA Law Review*, 60:1412, 2012. URL https://heinonline.org/HOL/Page?handle=hein.journals/uclalr60&id=1454&div=&collection=.

Ben-Shahar, O. and Logue, K. D. Outsourcing Regulation: How Insurance Reduces Moral Hazard. *Michigan Law Review*, 111(2):197–248, 2012. ISSN 0026-2234. URL https://www.jstor.org/stable/41703440. Publisher: Michigan Law Review Association.

Buiten, M., de Streel, A., and Peitz, M. EU liability rules for the age of Artificial Intelligence, 2021. URL https://cerre.eu/publications/eu-liability-rules-age-of-artificial-intelligence-ai/.

Buiten, M., de Streel, A., and Peitz, M. The law and economics of AI liability. *Computer Law & Security Review*, 48:105794, April 2023. ISSN 0267-3649. doi: 10.1016/j.clsr.2023.105794. URL https://www.sciencedirect.com/science/article/pii/S0267364923000055.

Carlsmith, J. Is Power-Seeking AI an Existential Risk?, June 2022. URL http://arxiv.org/abs/2206.13353. arXiv:2206.13353 [cs].

Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krasheninnikov, D., Langosco, L., He, Z., Duan, Y., Carroll, M., Lin, M., Mayhew, A., Collins, K., Molamohammadi, M., Burden, J., Zhao, W., Rismani, S., Voudouris, K., Bhatt, U., Weller, A., Krueger, D., and Maharaj, T. Harms from Increasingly Agentic Algorithmic Systems. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pp. 651–666, New York, NY, USA, June 2023. Association for Computing Machinery. ISBN 9798400701924. doi: 10.1145/3593013.3594033. URL https://dl.acm.org/doi/10.1145/3593013.3594033.

Commission, U. S. N. R. The Price-Anderson Act: 2021 Report to Congress, Public Liability Insurance and Indemnity Requirements for an Evolving Commercial Nuclear Industry. 2021. URL https://www.nrc.gov/docs/ML2133/ML21335A064.pdf.

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., and Brauner, J. Thousands of AI Authors on the Future of AI, April 2024. URL http://arxiv.org/abs/2401.02843. arXiv:2401.02843 [cs].

Gudgel, J. E. Insurance as a Private Sector Regulator and Promoter of Security and Safety: Case Studies in Governing Emerging Technological Risk From Commercial Nuclear Power to Health Care Sector Cybersecurity. 2022. URL https://hdl.handle.net/1920/13083.

Lior, A. AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy. *Mitchell Hamline Law Review*, 46:1043, 2019. URL https://heinonline.org/HOL/Page?handle=hein.journals/wmitch46&id=1043&div=&collection=.

Lior, A. AI Strict Liability vis-a-vis AI Monopolization. *Columbia Science and Technology Law Review*, 22:90, 2020. URL https://heinonline.org/HOL/Page?handle=hein.journals/cstlr22&id=90&div=&collection=.

Logue, K. D. Solving the Judgment-Proof Problem. *Texas Law Review*, 72:1375, 1993. URL https://heinonline.org/HOL/Page?handle=hein.journals/tlr72&id=1409&div=&collection=.

Maliha, G., Gerke, S., Cohen, I. G., and Parikh, R. B. Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation. *The Milbank Quarterly*, 99(3): 629–647, September 2021. ISSN 0887-378X. doi: 10.1111/1468-0009.12504. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8452365/.

Meehan, T. Lessons From the Price-Anderson Nuclear Industry Indemnity Act for Future Clean Energy Compensatory Models.

Mumpower, J. L. Risk, Ambiguity, Insurance, and the Winner's Curse. *Risk Analysis*, 11 (3):519–522, 1991. ISSN 1539-6924. doi: 10.1111/j.1539-6924.1991.tb00637.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.1991.tb00637.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.1991.tb00637.x.

Mustafa, R. *Insuring nuclear risk*. PhD thesis, July 2017. Available from INIS: http://inis.iaea.org/search/search.aspx?orig_q=RN:50083099 INIS Reference Number: 50083099.

Narechania, T. N. and Sitaraman, G. An Antimonopoly Approach to Governing Artificial Intelligence, January 2024. URL https://papers.ssrn.com/abstract=4755377.

OECD. OECD AI Incidents Monitor (AIM), 2024. URL https://oecd.ai/en/incidents.

Price II, W. N. and Cohen, I. G. Locating Liability for Medical AI, July 2023. URL https://papers.ssrn.com/abstract=4517740.

Rees, J. V. *Hostages of Each Other: The Transformation of Nuclear Safety since Three Mile Island*. University of Chicago Press, Chicago, IL, June 1996. ISBN 978-0-226-70688-7. URL https://press.uchicago.edu/ucp/books/book/chicago/H/bo3618989.html.

Vipra, J. and Korinek, A. Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT. 2023.

Vladeck, D. C. Machines without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review*, 89:117, 2014. URL https://heinonline.org/HOL/Page?handle=hein.journals/washlr89&id=124&div=&collection=.

Weil, G. Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence, January 2024. URL https://papers.ssrn.com/abstract=4694006.