Care for Chatbots [Extended Abstract]

Anonymous

June 2024

1. Introduction

Individuals will rely on language models (LMs) like ChatGPT to make decisions. Sometimes, due to that reliance, they will get hurt, have their property be damaged, or lose money. If the LM had been a person, they might sue the LM. But LMs are not persons.

This paper analyses whom the individual could sue, and on what facts they can succeed according to the *Hedley Byrne*-inspired doctrine of negligence in Canadian and English tort law. The paper identifies a series of hurdles conventional Canadian and English negligence doctrine poses and how they may be overcome. Such hurdles include identifying who is making a representation or providing a service when an LM generates a statement, determining whether that person can owe a duty of care based on text the LM reacts to, and identifying the proper analytical path for breach and causation.

The primary contribution of this paper is to show how common law courts could address liability by making natural extensions to existing common law doctrine rather than by inventing new doctrines solely for dealing with LMs.

2. What it means to do something in law

In Canadian (but not English) negligence law, a person is liable for pure economic loss only if their conduct fits into certain categories. These categories include "providing a service" and "making a representation". The first question in Canadian law is thus to decide whether someone is providing a service or making a representation.

In English law, meanwhile, the first question is simply whether there is a duty of care. The answer to this question, according to recent UKSC jurisprudence, is that a duty can only arise when someone takes on a task. This answer is probably also the same in Canadian negligence law, albeit phrased in a different manner.

The upshot of these two doctrinal points is that whether someone has taken on a task (or performed a service, or made a representation) is important to whether they owe a duty. LMs present a confusing set of facts for that question, because it is not clear who "does" what an LM "does".

This paper suggests that the way to address this question is to query who "controls" the system that the LM is embedded in.The paper suggests that, in answering this question, courts should prioritise social dimensions of control (for example, who understand how a system works, not merely what it does) over physical dimensions of control (such as on whose hardware a program is running) when assessing control and therefore responsibility.

To sharpen this question, this paper sets out three hypotheticals: *Gmail*, *Outlook*, and *Thunderbird*. All have the same structure. Recipient receives an email from Sender. Recipient's email client automatically identifies that Sender's email is long and suggests it summarise Sender's email. Recipient accepts this suggestion and Recipient is shown a summary of Sender's email. Recipient then loses money due to relying on an incorrect statement in the summary.

In all three cases, Developer's action (putting the code into operation) may far precede and be done on less precise information than Recipient had when Recipient agreed to summarise the email. The questions in each case are whether Developer has made a representation (via the summary) or provided a service (summarisation) and whether Recipient's acceptance of the suggestion redirects the causal attribution.

The difference between the three is how the email is summarised. In *Gmail*, the summarising occurs on the LM developer's computer. In *Outlook*, the summarising occurs via a product that operates on Recipient's computer. In *Thunderbird*, the summarising occurs only if Recipient has installed an LM onto their own computer.

3. What it means to know something in law

The duty analysis is not complete even if someone has taken on a task. The duty-bearer also must have reasonably expected another person to reasonably rely upon it. To reasonably expect something, the duty-bearer must know something about the relying party. But what does knowing something mean? The conventional philosophical answer is that knowledge requires a justified true belief.

LMs make this analysis challenging, because they do not have justified true beliefs. The justified true belief defini-

tion implies human knowledge. Even if one can say that the controller of an LM acts through their tool, they act through their tool without any human having knowledge.

The paper thus reassesses what it means to know something. Extending Samir Chopra and Laurence White's earlier work, it proposes redefining "knowledge" such that:

X knows *p* if:

1. (*X* has ready access to *p*; and
2. *p* is true; and
3. *X* makes use of the informational content of *p* without necessarily accessing *p*; and
4. The purpose of asking whether *X* knows *p* relates to how *X* uses *p*) or
5. *X* has a justified true belief in *p*.

Under this definition, one might say that Amazon (X) actually knows an individual's address (p) because Amazon has ready access to the address, the address is the real address, and Amazon can make use of the address when it ships a package there without any person in Amazon's employ accessing the address and forming a justified true belief about it. But one could not say 'Amazon (actually) knows a review was doxxing me when it included my address', even if it also had my shipment information, because the purposes of 'sending me things' and 'moderating reviews' are distinct.

Adopting this definition would have significant consequences for a duty arising. On the justified true belief definition of knowledge, the Developer in *Gmail* would owe no duty to Recipient for the quality of the summary, because no human person would know the content of the email. On this revised definition, Developer could owe a duty. Developer has ready access to the email and uses the informational content of the summarised email to make the summary. Developer's (actual) knowledge of the contents of the email should make Developer expect the purpose for which Recipient will reasonably rely on the summary.

4. Disclaiming duties

The duty analysis does not end there. A *prima facie* duty may be disclaimed, albeit imperfectly. Both English law (by statute) and Canadian law (at common law) have rules that make disclaimers invalid in certain circumstances. The paper thus analyses whether these circumstances would hold for LM-driven advice.

5. What the standard of care truly requires

The next focus of the paper is on a deep tension running through the breach and causation analyses, relating to how to describe someone who takes an imprudent process when performing an act but whose ultimate act is nonetheless justifiable. According to *Adams v Rhymney Valley*, [2000] EWCA Civ 3035, there is no breach when someone used an imprudent process so long as they got to a justifiable result. This is the conventional view. According to *Lion Nathan Ltd v CC Bottlers Ltd*, [1996] UKPC 9, however, a reasonable result reached via a calculation error still involves a breach. The paper interrogates these two cases, and suggests ways that they can be reconciled.

The paper also addresses the consequence of not-reconciling the cases, and of following *Adams* vs following *Lion Nathan* for LM-generated representations. On the *Adams* approach, one would ask whether the LM-generated representation was one a reasonable human might have made. On the *Lion Nathan* approach, one wopuld instead ask whether one should have answered the question using an LM.

6. Alternative approaches

Finally, the paper identifies alternative approaches to liability for software propounded in the literature and suggests that these approaches are not plainly superior to working within the existing framework that treats software as a tool used by a legal person. It rejects products liability (because pure economic loss is not compensable and because of the limited liability associated with informational goods); negligent supervision (again because of the limitations on pure economic loss, and because supervision focuses on discrete wrongful conduct, not someone's choices about how to design an LM); agency (because it needs the same doctrinal moves as advanced above, but while inviting the incorrect inference that an LM is an agent); and vicarious liability (because LMs are not persons, they have no liability that can be vicariously attributed to their 'principal').

References

Unfair Contract Terms Act, 1977, c 50 (UK)

Consumer Rights Act, 2015, c 15 (UK)

1688782 Ontario Inc v Maple Leaf Foods Inc, 2020 SCC 35

Adams v Rhymney Valley DC, [2000] EWCA Civ 3035

Arab Lawyers Network Co Ltd v Thomson Reuters (Professional) UK Ltd, [2021] EWHC 1728 (Comm)

Armstrong v Strain, [1951] 1 TLR 856

Baden v Societe Generale pour Favoriser le Developpement du Commerce et de l'Industrie en France SA, [1993] 1 WLR 509 (Ch)

Caparo Industries plc v Dickman, [1990] UKHL 2

Deloitte & Touche v Livent Inc (Receiver of), 2017 SCC 63

Dorset Yacht Co Ltd v Home Office, [1970] AC 1004

El Ajou v Dollar Land Holdings Plc (No1), [1993] EWCA Civ 4

Esso Petroleum Co Ltd v Mardon, [1976] QB 801 (EWCA Civ)

Hedley Byrne & Co Ltd v Heller & Partners Ltd, [1963] UKHL 4

Ilott v Wilkes, (1820) 106 ER 674

Lion Nathan Ltd v CC Bottlers Ltd, [1996] UKPC 9

Manchester Building Society v Grant Thornton UK LLP, [2021] UKSC 20, [2022] AC 783

Micron Construction Ltd v Hong Kong Bank of Canada, 2000 BCCA 141 [*Micron*]

Moffatt v Air Canada, 2024 BCCRT 149

NRAM Ltd (formerly NRAM plc) v Steel, [2018] UKSC 13

Playboy Club London Ltd v Banca Nazionale del Lavoro SpA, [2018] UKSC 43

Queen v Cognos Inc [1993] 1 SCR 87

Royal Bank of Scotland International Ltd v JP SPC 4, [2022] UKPC 18

Smith v Eric S Bush, [1990] *UKHL 1*

Uber Technologies Inc v Heller, 2020 SCC 16

Walter v Bauer, [1981] 109 Misc 2d 189 (NY Sup Ct)

Pınar Çağlayan Aksoy, "AI as Agents: Agency Law" in Larry A DiMatteo, Cristina Poncibò & Michel Cannarsa, ed, *The Cambridge Handbook of Artificial Intelligence*, 1st ed (Cambridge, UK: Cambridge University Press, 2022)

Roy W Arnold, "The Persistence of Caveat Emptor: Publisher Immunity from Liability for Inaccurate Factual Information Note" (1992) 53:3 U Pitt L Rev 777

Roderick Bagshaw, "Causing the Behaviour of Others and Other Causal Mixtures" in Richard Goldberg, ed, *Perspectives on Causation* (London, UK: Bloomsbury, 2011), 361

Andrew T Bayman, "Strict Liability for Defective Ideas in Publications Notes" (1989) 42:2 Vand L Rev 557)

Allan Beever, "The Basis of the Hedley Byrne Action" in Kit Barker, Ross Grantham & Warren Swain, ed, *The Law of Misstatements: 50 Years on from Hedley Byrne v Heller* (London, UK: Hart Publishing, 2015), 83

Peter Cane, *The Anatomy of Tort Law* (Oxford: Hart Publishing, 1997)

Karni Chagal-Feferkorn, "Tort Law: Applying A 'Reasonableness' Standard to Algorithms" in Woodrow Barfield, ed, *The Cambridge Handbook of the Law of Algorithms* (Cambridge: Cambridge University Press, 2020),

Alan Chan et al, "Harms from Increasingly Agentic Algorithmic Systems" (Paper delivered at FAccT '23: the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago IL USA, 6 December 2023), ACM 651

Bryan H Choi, "Crashworthy Code" (2019) 94:1 Wash L Rev 39

Bryan H Choi, "Software as a Profession" (2020) 33:2 Harv JL & Tech 557

Samir Chopra & Laurence White, "Attribution of Knowledge to Artificial Agents and their Principals" (Paper delivered at IJCAI'05, Edinburgh, Scotland, 30 July 2005), 1175

Ben Clifford, "Preventing AI Misuse: Current Techniques", (17 December 2023), online (blog): *GovAI Research Blog* ¡www.governance.ai/post/preventing-ai-misuse-current-techniques¿

Ignacio Cofone, "Servers and Waiters: What Matters in the Law of A.I." (2018) 21:2 Stan Tech L Rev 167 at 176

Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction" (2019) 5 Engaging STS 40

Richard A Epstein, "A Theory of Strict Liability" (1973) 2:1 J Leg Stud 151

Hugh Evans, "Negligence and Process" (2013) 29:4 PN 212

John Jay Fossett, 'The Development of Negligence in Computer Law' (1987) 14:2 N Ky L Rev 289

A Michael Froomkin, Ian Kerr & Joelle Pineau, "When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced over-Reliance on Machine Learning" (2019) 61:1 Ariz L Rev 33

Deep Ganguli et al, "Predictability and Surprise in Large Generative Models" (2022), *arXiv* ¡arxiv.org/pdf/2202.07785.pdf¿

Deep Ganguli et al, 'The Capacity for Moral Self-Correction in Large Language Models' (2023), *arXiv* ¡arxiv.org/pdf/arXiv:2302.07459.pdf¿

Michael C Gemignani, "Product Liability and Software"

(1981) 8:2 Rutgers Computer & Tech LJ 173

Stephen G Gilles, "Negligence, Strict Liability, and the Cheapest Cost-Avoider" (1992) 78:6 Va L Rev 1291

James Grimmelmann, "Regulation by Software Note" (2005) 114:7 Yale LJ 1719

James Grimmelmann, "Spyware vs. Spyware: Software Conflicts and User Autonomy" (2020) 16:1 Ohio St Tech LJ 25 at 27–34

Ian Lloyd, "A rose by any other name" (1993) Jan J Bus L 48

Tejas N Narechania, "Machine Learning as Natural Monopoly" (2022) 107:4 Iowa L Rev 1543

Donal Nolan, "Assumption of Responsibility: Four Questions" (2019) 72:1 Current Leg Probs 123

Susan Nycum, "Liability for Malfunction of a Computer Program" (1979) 7:1 Rutgers Computer & Tech LJ 1

Long Ouyang et al, "Training language models to follow instructions with human feedback" (2022), *arXiv* ¡arxiv.org/pdf/2203.02155.pdf¿

Dylan Patel, "Google 'We Have No Moat, And Neither Does OpenAI'", (4 May 2023), online: ¡ www.semianalysis.com/p/google-we-have-no-moat-and-neither¿

Robert S Peck, "The Coming Connected-Products Liability Revolution The Internet and the Law: Legal Challenges in the New Digital Age" (2022) 73:5 Hastings LJ 1305

Paul M Perell, "False Statements" (1996) 18:2 Adv Q 232

Michael Pratt, "What Would the Defendant have Done but for the Wrong?" (2020) 40:1 Oxford J Leg Stud 28

Jennifer Rowley, 'The wisdom hierarchy: representations of the DIKW hierarchy' (2007) 33:2 Journal of Information Science 163

Jane Stapleton, 'Duty of Care Factors: a Selection from the Judicial Menus" in Peter Cane & Jane Stapleton, ed, *The Law of Obligations: Essays in Celeberation of John Fleming* (Oxford: Clarendon Press, 1998), 59

Jane Stapleton, "Software, Information and the Concept of Product" (1989) 9 Tel Aviv U Stud L 147

Sandy Steel, 'Defining causal counterfactuals in negligence' (2014) 130:Oct Law Q Rev 564

Greg Swanson, "Non-Autonomous Artificial Intelligence Programs and Products Liability: How New AI Products Challenge Existing Liability Models and Pose New Financial Burdens Comments" (2019) 42:3 Seattle UL Rev 1201

Niranjan Venkatsen, "Causation in misrepresentation: historical or counterfactual? And 'but for' what?" (2021) 137:Jul Law Q Rev 503

Peter Watts, "Principals' Tortious Liability for Agents' Negligent Statements—Is 'Authority' Necessary?" (2012) 128:260 Apr Law Q Rev 260

Simon Whittaker, "European product liability and intellectual products" (1989) 105: Jan Law Q Rev 125

Cecil A Wright, "Knowledge of an Agent or Principal as Affecting Liability" (1937) 15:9 Can Bar Rev 716

John Zerilli et al, "Algorithmic Decision-Making and the Control Problem' (2019) 29:4 Minds & Machines 555