NORTH EAST JOURNAL OF LEGAL STUDIES

Volume Forty-Seven Fall 2026

ABOUT THE JOURNAL

An official publication
of the
North East Academy of Legal Studies in Business
ISSN: 1545-0597

The North East Journal of Legal Studies is a double-blind refereed journal, published annually. Its purpose is to encourage scholarly research in legal studies, taxation and pedagogy related thereto.

Articles should be submitted by October 1st each year. The review process takes up to 8 weeks. Notice to authors will occur between November 15th and December 1st. Accepted articles must be corrected and revised up to February 1st and submitted in the proper format by email. Articles will be published on or before May 1st on the NEALSB website.

Articles may be submitted simultaneously to this journal and others with the understanding that the author(s) will notify this journal if the article will be published elsewhere. We will not publish an article that will be published in another journal.

Papers should relate to the field of Business Law (including recognized topics within Business Law, Taxation and the Legal Environment of Business) or to Legal Studies Education.

The Journal will consider the submission of articles from those papers presented at the North East Academy of Legal Studies in Business Annual Conference. The paper designated the recipient of the Hoehlein Award for Distinguished Papers at the NEALSB Conference will serve as the lead article of the journal. Up to four articles from resources other than those presented at the NEALSB Conference may be published in the journal.

INFORMATION FOR CONTRIBUTORS

Articles offered for inclusion in the next issue of the journal shall be submitted to the editor by October 1st. There are two procedures for submitting papers for publication consideration:

- two Microsoft Word attachments (one with author-identifying information and one without author-identifying information emailed to nealsbjournal@gmail.com; or

- uploaded via Scholastica

Submission must include a check for $50.00 (non-refundable) payable to the North East Academy of Legal Studies in Business. Article submissions will not be reviewed until checks are received. Checks must be sent to:

NEALSB
c/o John Paul, Editor-in-Chief
15 Gaynor Avenue 2G
Manhasset, New York 11030

Format

1. Papers should be no more than twenty single-spaced pages, including footnotes. Use font 12 pitch, Times New Roman. Skip lines between paragraphs and between section titles and paragraphs. Indent paragraphs 5 spaces. Right-hand justification is desirable, but not necessary.

2. Margins: left- and right-hand margins should be set at 1 ¼ inches, top margin at 1 ½ inches and bottom margin at 1 ¾ inches.

3. Page Setup: Custom size your paper to have 6 ¾ inch width and a height of 10 inches. Your hard copy should be printed on a standard 8 ½" x 11" paper size. This will allow for the proper binding and trimming for printing purposes.

4. Upon acceptance, the first page must have the following format: the title should be centered, in CAPITAL LETTERS. Two lines down center the word "by" and the author's name, followed by an asterisk (*). Begin text three lines under the author's name. Two inches from the

bottom of the page, type a solid line 18 inches in length, beginning from the left margin. On the second line below, type the asterisk and the author's position of title and affiliation.

5. Headings

First Level (caps, flush with left margin)

Second Level (center, italics)

Third Level (flush with left margin, italics, followed by a colon [:],

Fourth Level (flush with left margin, italics, followed by a colon [:], with text immediately following).

6. Endnotes should conform to Uniform System of Citation, current edition, and should begin three lines after the end of the text.

7. All papers must be submitted using Microsoft Word.

8. Email a copy of your paper to the following email address: nealsbjournal@gmail.com

The Journal is listed in Cabells, EBSCO: Index to Legal Periodicals, Legal Source, Omni File Full Text Mega, Omni File Full Text Select, Volumes from 2008-2022. Also listed in Hein Online, Digital Commons/Law Review Commons, the Index to Legal Periodicals and Scholastica.

In recent history, the acceptance rate ranges from 15%-20%; in overall history, the acceptance rate has been 20-30%.

Individual copies of the journal are available to non-members and libraries at $25.00 per copy.

General correspondence, application for membership NEALSB or change of address notice should be addressed to the name above at the address therein stated.

NORTH EAST JOURNAL OF LEGAL STUDIES

VOLUME 47 **Fall 2026**

ARTICLES

SURVIVING STRICT SCRUTINY: POST-JUDGMENT INJUNCTIONS FOR DEEPFAKE ELECTION DEFAMATION

by

Joseph Romano*

"In a republic where the people are sovereign, the ability of the citizenry to make informed choices among candidates for office is essential, for the identities of those who are elected will inevitably shape the course that we follow as a nation."

— *Buckley v. Valeo*, 424 U.S. 1, 14-15 (1976) (per curiam)

INTRODUCTION

The Supreme Court has characterized prior restraints as "the most serious and the least tolerable" speech restriction.[1] But what if the speech is not only false, but synthetically

* J.D. Candidate Penn State Dickinson Law (expected 2027); B.S., Florida State University. Contact: www.linkedin.com/in/joey-romano-4587a7290

[1] Nebraska Press Ass'n v. Stuart, 427 U.S. 539, 559 (1976).

fabricated to convincingly depict events that never occurred—and timed to distort the outcome of an election?

That's a reality with "deepfakes":[2] computer generated images or videos developed with generative artificial intelligence.[3] Unlike traditional computer-generated images, deepfakes are both strikingly realistic and difficult to distinguish from genuine content. Research has concluded that participants could only identify deepfake videos as fake 24.5% of the time.[4] Another study found that even when participants were warned they would see deepfakes, be asked to detect them, and were trained to do so, their accuracy was just 60.7%, barely better than a flip of a coin.[5]

Consequently, this technology may be used to impersonate people, saying and doing things they never said or did, and convincingly depicting them. Given that voters rely more and more on advertisements to become informed on candidates,[6] deepfakes therefore pose a danger in the electoral sphere. Both U.S. and international headlines are filled with instances of the newest deepfake political advertisement.[7] Malicious deepfake users have even distributed content hours before an election, falsely stating that a candidate had dropped out.[8]

As a result, the government response has started to gain traction. Notably, the former Director of National Intelligence, Daniel R. Coats, identified deepfakes as a threat to the United States because they enable the foreign manipulation and disruption of U.S. election

---

[2] *See, e.g.*, Darrel M. West, *Can you spot a fake political Ad? AI is making it harder.*, THE WASHINGTON POST (April 14, 2026); Lindsey Wilkerson, *Still Waters Run Deep(fakes): The Rising Concerns of "Deepfake" Technology and Its Influence on Democracy and the First Amendment*, 86 MO. L. REV. 407 (2021); Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1792 (2019).

[3] *What the heck is a deepfake?*, UNIV. VA. INFO. SEC., https://security.virginia.edu/deepfakes.

[4] *See* Pavvel Korshunov & Sébastien Marcel, *Deepfake Detection: Humans vs. Machines*, ARXIV, 4 (Sept. 7, 2020), https://arxiv.org/pdf/2009.03155.

[5] Klair Somoray & Dan J. Miller, *Providing Detection Strategies to Improve Human Detection of Deepfakes*, 149 COMPUT. HUM. BEHAV. 1, 8 (2023); *see also* Nils C. Köbis, Barbora Doležalová & Ivan Soraperra, *Fooled Twice: People Cannot Detect Deepfakes, but Think They Can*, ISCIENCE, Oct. 29, 2021, at 1, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8602050/pdf/main.pdf (finding that participants consistently failed to detect deepfakes, and their overconfidence in their ability further exacerbated the problem); Alena Birrer and Natascha Just, *What We Know and Don't Know About Deepfakes: An Investigation into the State of the Research and Regulatory Landscape*, NEW MEDIA & SOCIETY (2024), https://doi.org/10.1177/14614448241253138 (finding that human participants could correctly identify deepfakes with an average accuracy of 63.3%, but their accuracy varied depending the image resolution, familiarity with the person depicted, and demographic similarities between the observer and the deepfake subject).

[6] Evan Richman, *Deception in Political Advertising: The Clash Between the First Amendment and Defamation Law*, 16 CARDOZO ARTS & ENT. L.J. 667, 667 (1998).

[7] *See, e.g.*, *The Age of Erie A.I. Political Ads is Here*, N.Y. TIMES, Mar. 13, 2026; Deb Riechmann, *Fears Grow Over Deceptive "Deepfake" Videos Made to Sway Elections*, TALKING POINTS MEMO, July 2, 2018, https://talkingpointsmemo.com/news/deepfake-videos-adversaries-political-campaigns-national-security; Cameron McKay, Inga Trauthig, *Then and Now: How Does AI Electoral Interference Compare in 2025?*, CENTRE FOR INTERNATIONAL GOVERNANCE INTERVENTION, Aug. 7, 2025 (discussing deepfake election advertisements in multiple countries).

[8] Juan Romero, *Deepfake Electoral: el falso video, la denuncia del PRO, la orden de la Justicia a X y el "Macri está hecho un llorón" de Milei*, FORBES ARGENTINA, May 18, 2025, https://www.forbesargentina.com/today/deepfake-electoral-falso-video-denuncia-pro-orden-justicia-x-macri-esta-hecho-lloron-milei-n72431.

systems.[9] In Congress, both the Senate and the House have also started to discuss these dangers[10] and have proposed regulations.[11] Some policymakers propose criminalizing deepfake use.[12] Others propose that the Federal Election Commission should regulate deepfakes when used to solicit funds.[13]

On the criminal and civil litigation side, scholars have suggested actions for deepfake use,[14] including a hybrid action combining defamation and impersonation.[15] However, existing tort theories may already be actionable.

In the realm of tort law, plaintiff-candidates can bring claims for damaging deepfake uses such as the intentional infliction of emotional distress (IIED), false light, the right of publicity, and defamation.[16] Nevertheless, defamation is one of the most likely actions to stick: Deepfake creations that intend to depict factually true events are both inherently false and fabricated, and creations after a court adjudicates them to be defamatory fall outside First Amendment protection to the extent that it satisfies the constitutional defamation standards.[17] Accordingly, a prevailing plaintiff-candidate may request that the court issue an injunction restricting the republication of a defendant's defamatory deepfake advertisement.

But, depending on the jurisdiction, the breadth of these injunctions could raise a prior restraint issue,[18] in which the injunction may be constitutional only if it survives strict

---

[9] *Worldwide Threat Assessment of the U.S. Intelligence Community, Hearing Before the S. Select Comm. on Intel.*, 116th Cong. 7 (2019) (prepared statement of Daniel R. Coats, Dir. of Nat'l Intel.) ("[a]dversaries and strategic competitors . . . may seek to use cyber means to directly manipulate or disrupt election systems - such as by tampering with voter registration or disrupting the vote tallying process - either to alter data or to call into question our voting process."); *see also* Letter from Adam B. Schiff, Stephanie Murphy, and Carlos Curbelo, Representatives, U.S. House of Representatives, to Daniel R. Coats, Dir., Office of Nat'l Intelligence (Sept. 13, 2018), https://schiff.house.gov/imo/media/doc/2018-09%20ODNI%20Deep%20Fakes%20letter.pdf (noting the threats deepfakes pose to domestic election security and asking Daniel Coats to report to Congress on them).

[10] For Congressional statements regarding the threats that deepfakes pose to elections, see: Yvette Clarke, *Deepfakes Will Influence the 2020 Election - and Our Economy, and Our Prison System*, QUARTZ: IDEAS (July 11, 2019), https://qz.com/1660737/deepfakes-will-influence-the-2020-election/ (Congresswoman Yvette Clarke has suggested that "the threat of election interference is perhaps the most menacing and urgent."); *Nomination of William R. Evanina to Be the Director of the National Counterintelligence and Security Center Before the S. Select Comm. on Intel.*, 115th Cong. 12 (2018) (statement of Sen. Marco Rubio, Member, S. Select Comm. on Intel.) (Marco Rubio echoing these concerns).

[11] For an active tracker of the legislation for deepfakes, see: *Tracker: State Legislation on Deepfakes in Elections*, PUB. CITIZEN, https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/.

[12] H.R. 4611, 118th Cong. (2023); H.R. 6088, 116th Cong. (2020).

[13] S. 2770, 118th Cong. (2023), which was later amended in S. 2770, 118th Cong. (2024); *see also* Chesney & Citron, *supra* note 2, at 1807.

[14] Eugene Volokh, *One-to-One Speech Vs. One-to-Many Speech, Criminal Harassment Laws, and "Cyberstalking"*, 107 NW. U. L. REV. 731 (2013).

[15] John Thayer, *Defamation or Impersonation? Working Towards a Legislative Remedy for Deepfake Election Misinformation*, 66 WM. & MARY L. REV. 251 (2024).

[16] Ayelet Gorden-Tapiero, Yotam Kaplan, & Gideon Parchomovsky, *Deepfake Liability*, 104 N.C.L. REV. 377, 408 (2026).

[17] *See Note: Defamatory Political Deepfakes and the First Amendment*, 70 CASE W. RES. 417, 431-32 (2019) [hereinafter *Defamatory Deepfakes*] (discussing how other civil torts such as the right of privacy is unlikely to stand given a politician's status); *see also infra* Part I (discussing the greater likelihood that deepfake related defamation claims prevail compared to other forms of communication).

[18] McCarthy v. Fuller, 810 F.3d 456 (7th Cir. 2015) ("'[p]rior restraint' is just a fancy term for censorship, which means prohibiting speech before the speech is uttered or otherwise disseminated.").

scrutiny. This Article argues that narrowly tailored post-judgment injunctions prohibiting the republication of adjudicated defamatory deepfake content can survive strict scrutiny in jurisdictions that treat such injunctions as prior restraints. It asserts that preventing the dissemination of defamatory deepfakes is a compelling interest and offers considerations for courts when determining whether an injunction is narrowly tailored.

This Article proceeds in six parts. Part I defines deepfakes, examines their current and potential political uses, and considers when political deepfakes become defamatory. Part II provides the relevant background on post-judgment injunctions as a remedy for defamation and explores how defamatory deepfakes fit into this context. Part III discusses how preventing the republication of defamatory deepfakes in close temporal proximity to an election may be a compelling government interest. Part IV outlines the considerations for courts when determining whether an injunction is narrowly tailored. Part V proposes a test that encapsulates all of these considerations and addresses its practical limits. Part VI concludes.

This Article contributes to the emerging scholarship on defamatory deepfakes in electoral politics[19] by addressing how strict scrutiny governs post-judgment injunctions against adjudicated defamatory deepfake content in jurisdictions that treat such relief as a prior restraint.

I. POLITICAL DEEPFAKES AND DEFAMATION

At its core, a deepfake is a computer-generated video that's so advanced it's remarkably difficult to discern whether it's fake. But while that definition is simple, the technology behind it isn't.

A. Deepfake Technology.

Unlike traditional photo and video editing, which requires manual work in editing software, deepfake technology lets computers do the job in a fraction of the time.[20] While one can create deepfakes through many methods, creators often use a "deep" machine learning model, known as a "generative adversarial network" or a "GAN." GANs employ two neural networks that work in conjunction to create deepfake content. First, a network that generates realistic images; second, another network that judges whether those images are real or fake.[21] The network that generates the images essentially tries to fool the network that judges whether they are real. These systems, working in opposition with each other, produce realistic deepfake content as the final result.[22]

At first, the technology was widely used to swap faces in videos and in the late 2010s became popular on the online forum "r/deepfakes."[23] Many used deepfakes for the purposes

---

[19] *See Defamatory Deepfakes*, *supra* note 17, at 420.

[20] *See* Mika Westerlund, *The Emergence of Deepfake Technology: A Review* 40-41, TECH. INNOV. MGMT. REV. (Nov. 2019), https://timreview.ca/article/1282.

[21] Jobit Varughese, *What Are Generative Adversarial Networks (GANs)*?, IBM, https://www.ibm.com/think/topics/generative-adversarial-networks.

[22] Prakash Pandey, *Deep Generative Models*, MEDIUM (Jan. 23, 2018).

[23] Jeevan Biswas, *What Exactly Is Deepfakes and Why Is This AI-Based Creation a Menace*, ANALYTICS INDIA MAG. (Feb. 8, 2018), https://www.analyticsindiamag.com/deepfakes-ai-celebrity-fake-videos/.

of inserting actors' faces into movies and television shows they were not initially in,[24] or for pornographic content.[25] Soon after, deepfakes made their way into the electoral sphere. So far in U.S. courts, only nonpolitical public figures have brought defamation claims over deepfakes,[26] but candidate-plaintiffs have done so abroad.[27] It is inevitable that the U.S. will follow.

B. Political Defamation

While modern defamation law[28] allows plaintiffs to recover for harm caused by false factual statements, courts have struggled to define when recovery is appropriate in elections.[29] The issue requires balancing the First Amendment's protection of speech against the need to deter various harms. On one hand, political speech regulations are subject to strict scrutiny.[30] Yet on the other hand, candidates have an interest in preserving their reputation, and voters have an interest in "the ability to absorb 'real' information from candidates to make informed decisions about who should represent them in government."[31]

---

[24] Sam Haysom, *Nicolas Cage is Being Added to Random Movies Using Face-Swapping Technology*, MASHABLE (Jan. 31, 2018), https://mashable.com/2018/01/31/nicolas-cage-face-swapping-deepfakes/#WGhEd3yKgiqw.

[25] *2023 State of Deepfakes*, HOME SEC. HEROES (2023), https://www.securityhero.io/state-of-deepfakes/ (finding that after examining the 95,820 deepfake videos online, ninetyeight percent of those videos involved swapping the face of a pornographic actress with that of another woman). Deepfake content also has an app for this specific use, "FakeApp." *See* Megan Farokhmanesh, *Deepfakes Are Disappearing from Parts of the Web, But They're Not Going Away*, THE VERGE (Feb. 9, 2018), https://www.theverge.com/2018/2/9/16986602/deepfakes-bannedr-eddit-ai-faceswap-porn.

[26] For example, Grammy award winner Megan Thee Stalion, prevailed on a defamation claim against a defendant after they promoted a deepfake video falsely depicting Megan Thee Stalion in a pornographic context. *See* Steven Yablonski & Ivan Taylor, *Jurors Rule in Favor of Megan Thee Stalion in Miami Deepfake Porn Case, order Milagro Gramz to pay $75,000 in damages*, CBS NEWS (Dec. 1, 2025); *see also* Pete v. Cooper, No. 24-24228-CIV-ALTONAGA/Reid, 2026 U.S. Dist. LEXIS 86289 (S.D. Fla. Apr. 20, 2026) (discussing the remedies to award after including a declaratory injunction).

[27] *See, e.g.*, *Court orders Cong, leaders to take down 'deepfake' Modi-Adani video*, THE TIMES OF INDIA, https://timesofindia.indiatimes.com/city/ahmedabad/court-orders-cong-leaders-to-take-down-deepfake-modi-adani-video/articleshow/126082885.cms (invokving a defamatory deepfake video of the Indian Prime Minister Narendra Modi).

[28] Defamation's inception dates all the way back to early English common law. For the history up until *Sullivan*, see: Lovell, *The "Reception" of Defamation by the Common Law*, 15 VAND. L. REV. 1051, 1052 (1962); Van Vechten Veeder, *The History and Theory of the Law of Defamation*, 3 COLUM. L. REV. 546 (1903).

[29] For an overview, see: Comment: *American Defamation Law: From Sullivan, Through Greenmoss, and Beyond*, 48 OHIO ST. L.J. 513, 516-26 (1987).

[30] Citizens United v. FEC, 558 U.S. 310, 340 (2010) (quotations removed; citations removed); *see also* Meyer v. Grant, 486 U.S. 414, 425 (1988) (noting that, in the political sphere, First Amendment protection is "at its zenith," and that the burden to pierce that protection is "insurmountable.").

[31] Rebecca Green, Counterfeit Campaign Speech, 70 HASTINGS L.J. 1445, 1458 (2019). For an overview of when the First Amendment does *not* protect speech, see: Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc., 472 U.S. 749, 763 (1985) (noting that actual malice only applies to matters of "public concern"); Gertz v. Robert Welch, Inc., 418 U.S. 323, 346 (1974) (actual malice does not apply to standard individuals); Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (showing that the First Amendment does not protect speech that incites "imminent lawless action"); Chaplinsky v. State of New Hampshire, 315 U.S. 568, 572 (1942) (showing that the First Amendment does not protect "fighting words"); Watts v. United States, 394 U.S. 705, 708 (1969) (noting that the First Amendment permits the government to ban "true threats"); Virginia v. Black, 538 U.S. 343, 360 (2003) (describing "intimidation" as a subset of the unprotected category of "true threats," "where a speaker directs a threat to a person or group of persons with the intent of placing the victim in fear of bodily harm or death"). For an overview on the

Soon after *N.Y. Times Co. v. Sullivan*,[32] the Supreme Court calibrated these interests to their current state in both *Monitor Patriot Co. v. Roy*,[33] and *Ocala Star-Banner Co. v. Damron*.[34] The Court noted that candidate-plaintiffs qualify as "public figures,"[35] and are subject to the "actual malice" standard, which permits liability to withstand constitutional scrutiny[36] only when clear and convincing evidence shows that the defendant knew or recklessly disregarded the truth or falsity of their statement.[37] As such, a defendant's statements touching on "conduct relevant to [the candidate-plaintiff's] fitness for office"[38] are subject to the actual malice standard, such as a candidate's past criminal history,[39] mental illness,[40] and the candidate's professional skill.[41]

---

history of election defamation since the U.S's founding, see: Michael J. Klarman, *The Framers' Coup: The Making of the United States Constitution* 404, 409 (2016) (discussing how on "[b]oth sides" of the debate to ratify the Constitution distributed defamatory content) ("[i]n Pennsylvania, Federalist publishers went so far as to deliberately distort the published account of the state ratifying convention's debates to make it appear as if the Constitution had been unopposed there."); Jack Winsbro, Comment, *Misrepresentation in Political Advertising: The Role of Legal Sanctions*, 36 EMORY L.J. 853, 853-54 (1987). (tracing the history of misleading political advertising).

[32] 376 U.S. 254 (1964).

[33] 401 U.S. 265, 274 (1971).

[34] 401 U.S. 295, 299 (1971).

[35] Courts differ on whether a political candidate is either a "public figure" or "public official." Some courts even use the terms interchangeably. What is clear, however, is that the actual malice standard extends to political candaidates regardless of whether they are public figures, officials, or both. *Roy*, 401 U.S. at 271-72 ("the question is of no importance so far as the standard of liability in this case is concerned, for it is abundantly clear that, whichever term is applied, publications concerning candidates must be accorded at least as much protection under the First and Fourteenth Amendments as those concerning occupants of public office."). The only times court have not recogized a candidate as a public figure was when a plaintiff was a candidate for Haitian presidency, Rosenblatt v. Baer, 383 U.S. 75 (1966), and when plaintiff was a candidate for election as a delegate to the Maryland Constitutional Convention. A. S. Abell Co. v. Barnes, 258 Md. 56 (1970). *But see* Pendleton v. City of Haverhill, 156 F.3d 57, 69 (1st Cir. 1998) (applying the actual malice standard to a plaintiff who was not running for an elected position but nevertheless a "public post.").

[36] *See Richard L. Hasen, A Constitutional Right to Lie in Campaigns and Elections?*, 74 MONTANA L. REV. 53, 77 (2013) ("[t]o survive constitutional review, any false campaign speech law would have to be narrow, targeted only at false speech made with actual malice."); *Sullivan*, 376 U.S. at 279 (1964). The Court justified its limitation based on "a profound national commitment to the principle that debate on public issues should be uninhibited, robust and wide-open, and that it may well include vehement, caustic and sometimes unpleasantly sharp attacks on government and public officials." *Id.* at 271; *accord* Restatement (Second) of Torts § Scope for Defamation cmt. a (AM. L. INST. 1979).

[37] St. Amant v. Thompson, 390 U.S. 727, 732 (1968) (noting that actual malice can be inferred "where a story is fabricated by the defendant, is the product of his imagination, or is based wholly on an unverified anonymous telephone call[,]...[or] when the publisher's allegations are so inherently improbable that only a reckless man would have put them in circulation[, or] where there are obvious reasons to doubt the veracity of the informant or accuracy of his reports.").

[38] *Roy*, 401 U.S. at 273. In *Garrison v. Louisiana*, the Court defined what may be "relevant" as "dishonesty, malfeasance, or improper motivation, even though these characteristics may also affect the official's private character." 379 U.S. 64, 77 (1964).

[39] *See, e.g.*, *Roy*, 401 U.S. at 275-76 (holding that actual malice applies to a public figure candidate for a statement relating to a past criminal charge "no matter how remote in time or place"); *Garrison*, 379 U.S. at 77 (finding that a candidate's perjury indictment in a civil rights suit constituted).

[40] Goldwater v. Ginzburg, 414 F.2d 324 (2d Cir. 1969).

[41] *E.g.*, Dyer v. Davis, 189 So. 2d 678 (La. Ct. App. 1966).

In actual litigation, defendants frequently prevail at early dismissal or summary judgment for written or televised advertisements because of the stringent actual malice threshold.[42] But deepfakes differ. Because the events that they depict are fabricated, deepfake content is more susceptible to being considered factually false.[43] Given this nature and the fact that actual malice only requires knowledge of that falsity rather than the intended effect to cause harm,[44] deepfake uses may more readily support scienter inferences when compared to regular televised ads or written publications.[45]

Defendants who make content with a disclaimer may argue that the content is not within the bounds of defamation because the expression would not "reasonably appear to state or imply assertions of objective fact."[46]

Legal scholar Kathleen Ross compares this scenario to *Byrd v. Hustler Magazine, Inc.*, where the court rejected the claim that a magazine photo was defamatory on the theory that it implied the plaintiff posed for it, because the caption expressly stated it had been "retouched." [47]

However, as Kareem Gibson notes, disclaimers for explicit deepfakes may still not be enough, because a defendant could reasonably foresee[48] that the deepfake could be

---

[42] Note: *Malice, Lies, and Videotape: Revisiting New York Times v. Sullivan in the Modern Age of Political Campaigns*, 30 RUTGERS L. J. 755, 780 n.168, 783 (1999); Debra T. Landis, *Criticism or disparagement of character, competence, or conduct of candidate for office as defamation*, 37 A.L.R.4th 1088. Courts are reluctant to prohibit speech even for false statements. *See, e.g.*, *Sullivan*, 376 U.S. at 279, n. 19 ("even a false statement may be deemed to make a valuable contribution to public debate, since it brings about 'the clearer perception and livelier impression of truth, produced by its collision with error.'"); U.S. v. Alvarez, 567 U.S. 709, 733 (2012) (noting that "false factual statements can serve useful human objectives, for example: in social contexts, where they may prevent embarrassment, protect privacy, shield a person from prejudice, provide the sick with comfort, or preserve a child's innocence; in public contexts, where they may stop a panic or otherwise preserve calm in the face of danger; and even in technical, philosophical, and scientific contexts, where (as Socrates' methods suggest) examination of a false statement (even if made deliberately to mislead) can promote a form of thought that ultimately helps realize the truth.").

[43] *See Defamatory Deepfakes*, *supra* note 17, at 433-34 ("[u]nlike an article where an individual might not be sure about one particular fact regarding a political figure, creators of deepfakes know that the content they are creating is false") ("[e]ven as technology makes it easier to create deepfakes, it will still be hard to prove that an individual 'accidentally' created a video of a public figure in a compromising situation. . . . A deepfake creator will not only know that what she is creating is false, but since she is creating a *video*, many more persons are likely to believe that what the video depicts is true"); *see also* Marc Jonathan Blitz, *Deepfakes and Other Non-testimonial Falsehoods: When Is Belief Manipulation (Not) First Amendment Speech?*, 23 YALE J.L. & TECH. 160, 220 (2020) (noting that deepfakes give rise to defamation claims because of their false nature); Nina I. Brown, Deepfakes and the Weaponization of Disinformation, 23 Va. J.L. & Tech. 1, 39-40 (2020) (same).

[44] Abigail George, *Defamation in the Time of Deepfakes*, 45(1) COLUMBIA J. GENDER & L. 122, 158-59 (2024); Prosser and Keeton on the Law of Torts 809 (W. Page Keeton et al. eds., 5th ed. 1984).

[45] *See* Russel Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 367 ("[a]ll deepfakes, by definition, rise to the level of actual malice, should that standard apply.") (citing Ashby v. Hustler Mag., Inc., 802 F.2d 856, 860 (6th Cir. 1986)), 373-74 (discussing how deepfakes may be defamatory depending on their content and how they add a new dimension to defamation law).

[46] Takieh v. O'Meara, 497 P.3d 1000, 1006 (Ariz. Ct. App. 2021).

[47] Kathleen Ross, *Pornographic Deepfakes and Ugly Social Facts: The Costs of a Normative Approach to Defamation*, 124 MICH. L. REV. 838, 847 n.68 (2026).

[48] *See, e.g.*, Oparaugo v. Watts, 884 A.2d 63, 73 (D.C. 2005) (noting that "[t]he original publisher of a defamatory statement may be liable for republication if the republication is reasonably foreseeable."); Schneider v. United Airlines, Inc., 208 Cal. App. 3d 71, 75 (1989) (same). *But see* Fashion Boutique of Short Hills, Inc. v. Fendi USA, Inc., 314 F.3d 48, 60 (2d Cir. 2002) (rejecting this idea).

disseminated by others absent a disclaimer.[49] Gibson also suggests that the distributors of the deepfake altered to remove the disclaimer may even incur liability themselves as subsequent distributors or disseminators of the content.[50]

Consequently, courts will more often entertain awarding remedies for prevailing candidate-plaintiffs, with one of such being an injunction to cease the republication of the defamatory deepfake.

II. POST-JUDGMENT INJUNCTIONS FOR DEFAMATION

Generally, courts award prevailing plaintiffs with monetary damages to compensate for the damage to the plaintiff's reputation.[51] Yet, when defendants air defamatory deepfake advertisements in close temporal proximity to an election, monetary damages may be inadequate. After an adjudication, voters may still see the misleading and defamatory advertisement, which can harm their perception of a candidate even after the correction.[52] Further, the collective harm cannot be monetized. While plaintiff-candidates may be the "victims," inadvertently, both voters and the democratic process are too. Monetary remedies cannot compensate for the distortion of democratic choice. Nevertheless, courts may issue post-judgment injunctions prohibiting further republication of the adjudicated defamatory material.[53]

The law is relatively unsettled as to when these post-judgment speech suppressing injunctions are constitutional. The Supreme Court faced this question in *Tory v. Cochran*, noting that such injunctions may be overbroad. But the Court vacated the opinion because one of the parties had died.[54] Courts have been split since.

---

[49] Kareem Gibson, Note, *Deepfakes and Involuntary Pornography: Can Our Current Legal Framework Address This Technology?*, 66 WAYNE L. REV. 259, 272-73 (2020).

[50] *Id.* at 273-276. This scenario is similar to a rumor, where "a person repeats a slanderous charge, even though identifying the source or indicating it is merely a rumor, this constitutes republication and has the same effect as the original publication of the slander." Ringler Assocs. Inc. v. Maryland Cas. Co., 80 Cal. App. 4th 1165, 1180 (2000).

[51] Prevailing plaintiffs are entitled to many remedies outside of monetary damages. *See* Restatement (Second) of Torts § 620 (general damages), 621 (nominal damages). In terms of public officials, such as politicians, "a public image is a valuable asset. A favorable public image enables a public figure to earn large fees for lecturing or for endorsing products. It is a source of influence in politics, entertainment, sports, religion, education, or other fields. It may be an important source of self-esteem and personal satisfaction. A person who enjoys a positive public image thus may be injured by defamation, even if there is no harm to his existing or future personal relations." David A. Anderson, *Reputation, Compensation and Proof*, 25 WM. & MARY L. REV. 747, 766 (1984).

[52] *Defamatory Deepfakes*, *supra* note 17, at 437 ("[d]espite the attraction of monetary damages, they do nothing to stop the ongoing reputational loss caused by the deepfake's continued existence on the Internet") ("[p]olitical figures are likely to favor injunctions because removing the original video from the Internet will remedy the imminent issue of reputational loss caused by a defamatory political deepfake"); *accord* David S. Ardia, *Freedom of Speech, Defamation, and Injunctions*, 55 WM. & MARY L. REV. 1, 9, 15-16 (2013); *see also* Eugene Volokh, *Anti-Libel Injunctions*, 168 U. PA. L. REV. 73, 3 (2019) (noting how damages are inadequate for internet defamation because "[t]he Internet lets speakers publish libels at little cost to a potentially broad audience, and these libels can cause enduring damage. Every time someone types a plaintiff's name into Google, the libels pop up again. Moreover, 47 U.S.C. § 230(c)(1) generally immunizes intermediaries, such as search engines or online service providers, that do have money.").

[53] *See* Restatement (Second) of Torts § 623 (remedies outside of monetary damages).

[54] No. B159437, 2003 WL 22451378 (Cal. Ct. App. Oct. 29, 2003), *vacated*, 544 U.S. 734, 738 (2005).

In some jurisdictions, such an injunction may be constitutionally permissible if narrowly tailored to restrain only the defamatory speech.[55] These jurisdictions reason that a post-judgment injunction is not a prior restraint if narrowly tailored because the suppressed speech is defamatory and therefore not protected under the First Amendment.[56]

However, a minority of jurisdictions (including the Fourth Circuit, various federal district courts, and the Texas Supreme Court) treat orders restricting the republication of defamatory speech as prior restraints.[57] These jurisdictions reason that speech suppression injunctions act as prior restraints regardless of whether that speech is protected.[58] Others explain that defamation should not give remedies outside of equity because it is a tort.[59]

---

[55] For jurisdictions that do not treat post-judgment injunctions prohibiting defamatory speech as prior restraints, see: Brown v. Petrolite Corp., 965 F.2d 38 (5th Cir. 1992); Lothschuetz v. Carpenter, 898 F.2d 1200 (6th Cir. 1990); San Antonio Cmty. Hosp. v. S. Cal. Dist. Council of Carpenters, 125 F.3d 1230 (9th Cir. 1997); Wagner Equip. Co. v. Wood, 893 F. Supp. 2d 1157, 1164 (D.N.M. 2012) (District of New Mexico); Balboa Island Village Inn, Inc. v. Lemen, 156 P.3d 339 (Cal. 2007), *as modified* (Apr. 26, 2007) (iterating that a post-judgment injunction "does no more than prohibit the defendant from repeating the defamation.") (California); Retail Credit Co. v. Russell, 218 S.E.2d 54 (Ga. 1975); Advanced Training Sys., Inc. v. Caswell Equip. Co., 352 N.W.2d 1 (Minn. 1984); O'Brien v. Univ. Comty. Tenants Union, Inc., 327 N.E.2d 753 (Ohio 1975). Courts generally trend towards this line of thinking. *See* In re Conservatorship of Turner, No. 2013-01665, 2014 Tenn. App. LEXIS 278, 2014 WL 1901115, at *19 (Tenn. App. Mar. 19, 2014) (discussing the purported development of "a modern, superseding rule" that, once a trial court or a jury has made a final determination that speech is defamatory, the speech determined to be false may be enjoined); Hill v. Petrotech Resources Corp., 325 S.W.3d 302, 308 (Ky. 2010) (same); *see also* McCarthy v. Fuller, 810 F.3d 456, 464 (7th Cir. 2015) (J., Skyes, concurring) ("an emerging modern trend, however, acknowledges the general rule but allows for the possibility of narrowly tailored permanent injunctive relief as a remedy for defamation as long as the injunction prohibits only the repetition of the specific statements found at trial to be false and defamatory.").

[56] Counterman v. Colorado, 600 U.S. 66, 75-76 (2023) (iterating that defamatory speech is *not* protected under the First Amendment).

[57] *See, e.g.*, Alberti v. Cruise, 383 F.2d 268 (4th Cir. 1967); Oakley, Inc. v. McWilliams, 879 F. Supp. 2d 1087, 1090-92 (C.D. Cal. 2012); *see also* Kinney v. Barnes, 443 S.W.3d 87, 101 (Tex. 2014) (holding that a mere injunction to remove posted speech that is adjudicated to be defamatory is not a prior restraint, but an injunction to prohibit future speech, even if it is adjudicated to be defamatory, is a prior restraint); Kreimer, § 10.5(a)(1), at 311 ("[i]njunctions interfere with the dissemination of information on the basis of potentially exaggerated threats of possible future harm, rather than on the basis of the results of abuse proven before a jury.").

[58] Erwin Chemerinsky, *Injunctions in Defamation Cases*, 57 SYRACUSE L. REV. 157, 165 (2007) ("[i]njunctions are treated as prior restraints because that is exactly what they are: a prohibition on future expression."). Many of the minority jurisdictions rely on *Alexander v. United States*, which iterated that judicial orders "forbidding certain communications" that are "issued in advance of the time that such communications are to occur" are prior restraints. 509 U.S. 544, 550 (1993) (citing M. Nimmer, *Nimmer on Freedom of Speech* § 4.03, p. 4-14 (1984) (adding italics)). *But see* Madsen v. Women's Health Ctr., Inc., 512 U.S. 753, 764 n.2 (1994) (holding that certain content-neutral injunctions are not prior restraints); John Calvin Jeffries, Jr., *Rethinking Prior Restraint*, 92 YALE L.J. 409, 416-19 (1983) (questioning the rationale for prior restraints in applying to all speech regardless of its First Amendment protection). Legal scholar Erwin Chemerinsky also notes that even if a court limits an injunction to only suppressing unprotected defamatory speech, that speech may later become protected, rendering the restriction a prior restraint, such as when defamatory speech later becomes true and is therefore no longer defamatory. *See* Chemerinsky, *supra* note 58, at 171-72.

[59] *See, e.g.*, Kramer v. Thompson, 947 F.2d 666, 677 (3d Cir. 1991); Cmty. for Creative Non-Violence v. Pierce, 814 F.2d 663, 672 (D.C. Cir. 1987); United Sanitation Servs. of Hillsborough, Inc., v. City of Tampa, 302 So. 2d 435, 439 (Fla. 2d DCA 1974) ("even if a defamation had in fact taken place, "there would be no basis for a failure to follow the well-established rule that equity will not enjoin either an actual or a threatened defamation."). *But see* DOUGLAS LAYCOCK, THE DEATH OF THE IRREPARABLE INJURY RULE 165 (1991) ("damages are a seriously inadequate remedy for defamation.").

In these jurisdictions, such injunctions are only constitutional if they survive strict scrutiny. This Article does not discuss what jurisdiction is or should be constitutionally "correct." But instead assumes the minority framework and asserts that an injunction restraining the republication of a defamatory deepfake advertisement may survive strict scrutiny when the circumstances are ripe.

In the minority jurisdictions, prior restraints carry a "heavy presumption against [their] constitutional validity,"[60] which the proponent of the restraining injunction may only rebut if the injunction survives strict scrutiny.[61] To do so, the restrained activity must address either a serious or imminent threat to a "compelling" state interest and be "narrowly drawn," meaning that no less speech-restrictive alternative would achieve the interest as effectively.[62]

While a successful rebuttal is rare, the Court in *Near v. Minnesota* suggested in dicta that threats to interests such as national security, military recruiting, and obscenity[63] may be sufficient grounds.[64] Decades later, the Court saw some of these hypotheticals come to life.

For instance, in 1971, the Court in *New York Times Co. v. United States*[65] most famously faced the question of when national security concerns rebut the prior restraint presumption when the U.S. government attempted to restrain the distribution of the "Pentagon Papers," a controversial Department of Defense study that outlined relations between the U.S. and Vietnam during the Vietnam War.[66] In a 6-3 decision, the Court determined that the restraint did not rebut the presumption because the threat to national security, specifically the president's ability to conduct foreign affairs, was unpersuasive.[67]

But, eight years later, the Western District of Wisconsin in *United States v. Progressive, Inc.*, found "that publication of the technical information on the hydrogen bomb contained in the article is analogous to publication of troop movements or locations in time of war and

---

[60] Southeastern Promotions, Ltd. v. Conrad, 420 U.S. 546, 558 (1975); Bantam Books, Inc. v. Sullivan, 372 U.S. 58, 70 (1963); New York Times Co. v. United States, 403 U.S. 713, 714 (1971); Organization for a Better Austin v. Keefe, 402 U.S. 415 (1971).

[61] *Stuart*, 427 U.S. at 559; *Sindi v. El-Moslimany*, 896 F.3d 1, 32 (1st Cir. 2018) (quoting Carroll v. President & Comm'rs of Princess Anne, 393 U.S. 175, 183 (1968)); *see also* CBS, Inc. v. Davis, 510 U.S. 1315, 1317 (1994) (Blackmun, J., in chambers); Goode, 821 F.3d 553, 559 (5th Cir. 2016); Cty. Sec. Agency v. Ohio Dep't of Commerce, 296 F.3d 477, 485 (6th Cir. 2002); Levine v. U.S. Dist. Ct., 764 F.2d 590, 595 (9th Cir. 1985); New York Times Co. v. United States, 403 U.S. 713, 714 (1971); Shuttlesworth v. Birmingham, 394 U.S. 147 (1969); Staub v. City of Baxley, 355 U.S. 313 (1958); Kunz v. New York, 340 U.S. 290 (1951); Schneider v. State, 308 U.S. 147 (1939); Lovell v. Griffin, 303 U.S. 444 (1938).

[62] *See* Perry Educ. Ass'n v. Perry Loc. Educators' Ass'n, 460 U.S. 37, 45 (1983); Carey v. Brown, 447 U.S. 455, 461 (1980).

[63] *See, e.g.*, Kingsley v. Brown, 354 U.S. 436 (1957) (upholding a New York criminal provision that prevented the distribution of "obscene" rebutted the prior restraint presumption) (finding that such prior restraint was constitutional, specifically for a party that distributed erotic fetish materials); Times Film Corp. v. Chicago, 365 U.S. 43, 47-50 (1961) (upholding a a Chicago ordinance that prohibited a film's public exhibition unless the government approved it) (finding that the prior restraint presumption applied because the plaintiff's film depicted a sexual relationship between an adult woman and a teenage boy).While cases involving obscenity are societally outdated, they demonstrate that the Court is willing to entertain rebuttals in instances beyond those involving national security and Sixth Amendment rights.

[64] 283 U.S. at 716.

[65] 403 U.S. at 713.

[66] Major William D. Toronto, *Fake News and Kill-Switches: The U.S. Government's Fight to Respond to and Prevent Fake News*, 79 A.F. L. REV. 167, 193 (2018).

[67] 403 U.S. at 713.

falls within the extremely narrow exception to the rule against prior restraint."[68] Taken together, these cases demonstrate that national security concerns can rebut the presumption, depending on the severity of the threat that the distributed information poses.[69]

Similarly, courts have found that a party has rebutted the prior restraint presumption for "gag orders" that preserve a criminal defendant's Sixth Amendment right to a fair trial. A "gag order," or a restriction on speech regarding a high-profile criminal case before trial, also "exhibit[s] the characteristics of prior restraints."[70] In *Nebraska Press Ass'n v. Stuart*, the Court first grappled with this issue.[71] Although the Court found that the restriction was not narrowly tailored,[72] a more narrowly tailored order could have sufficed.[73] To illustrate, in *United States v. Davis*, the Eastern District of Louisiana upheld a partial gag order restricting extrajudicial comments of trial participants in a capital police corruption case, finding that alternatives, including venue change, postponement, intensive voir dire, and jury sequestration, could not effectively guard the defendants' Sixth Amendment right to a fair trial.[74]

Prior restraint injunctions against defamation in the election context have not been explicitly addressed, but a constitutional tension remains.[75] In terms of political speech in conflict with the First Amendment, the Court has noted that political speech is at the heart of protected speech.[76] In *Citizens United v. Federal Election Commission*, the Court repeatedly emphasized the importance of political speech while striking down bans on independent political expenditures.[77] Yet at the same time, the Court has also observed that the very fabric of democracy turns on protecting the electoral process.[78] As the Court noted in *Garrison v. Louisiana*, the First Amendment does not automatically protect speech because it is used in

---

[68] 467 F. Supp. 990, 996 (W.D. Wis. 1979), *appeal dismissed*, 610 F.2d 819 (7th Cir. 1979).

[69] *See generally* Bank Julius Baer & Co. v. Wikileaks, 535 F. Supp. 2d 980, 982 (N.D. Cal. 2008); L.A. Powe, Jr., *The H-Bomb Injunction*, 61 U. COLO. L. REV. 55, 71 (1990) (citing Howard Morland, THE SECRET THAT EXPLODED 202 (1981)).

[70] U.S. v. Brown, 218 F.3d 415, 424 (5th Cir. 2000); *accord* In re Dow Jones, 842 F.2d 603, 609 (2d Cir. 1988); Levine v. United States District Court, 764 F.2d 590, 595 (9th Cir. 1985).

[71] 427 U.S. at 562.

[72] *Id.* at 563-64.

[73] *See, e.g.*, U.S. v. Brown, 218 F.3d 415 (5th Cir. 2000); *see also* James C. Goodale, *The Press Ungagged: The Practical Effect on Gag Order Litigation of Nebraska Press Association v. Stuart*, 29 STAN. L. REV. 497 (1977) (discussing the narrowly tailoring requirements for a criminal gag order).

[74] 904 F. Supp. 564, 568-69 (E.D. La. 1995).

[75] Chemerinsky, *supra* note 58, at 173.

[76] Mills v. Alabama, 384 U.S. 214, 218 (1966); Baumgartner v. United States, 322 U.S. 665, 673-74 (1944).

[77] Citizens United v. FEC, 130 S. Ct. 876, 898-99 (2010) (holding that the government "may not suppress political speech on the basis of the speaker's corporate identity"); First Nat'l Bank v. Bellotti, 435 U.S. 765, 798 (1978) (holding that the First Amendment protects corporate speech).

[78] *See, e.g.*, Eu v. S.F. Cnty. Democratic Cent. Comm., 489 U.S. 214, 223 (1989); Burson v. Freeman, 504 U.S. 191, 204-06 (1992); Purcell v. Gonzalez, 549 U.S. 1, 4 (2006); "[c]onfidence in the integrity of our electoral processes is essential to the functioning of our participatory democracy"); McIntyre v. Ohio Elections Commission, 514 U.S. 334, 379 (1995) (Scalia, J., dissenting) ("[t]he State has a compelling interest in preserving the integrity of its election process. So significant have we found the interest in protecting the electoral process to be that we have approved the prohibition of political speech entirely in areas that would impede that process.").

the political sphere, because intentionally lying is against the premise of democratic government.[79]

Thus, given the increasing prevalence of deepfakes, courts will soon face a familiar constitutional tension:[80] whether, and under what circumstances, the democratic choice can outweigh First Amendment speech protection.[81] The Court has tackled this issue many times, often permitting false speech,[82] yet drawing the line in the context of campaign finances.[83]

Defamatory deepfake ads in the context of elections may provide a ripe vessel for the court to consider. Such content could "be as corrosive as the worst campaign finance abuses."[84]

III. A COMPELLING INTEREST: PREVENTING A DEFAMATORY DEEPFAKE'S REPUBLICATION WHEN CLOSE IN TIME TO AN ELECTION

Unlike the false statements the courts have previously grappled with, defamatory deepfake statements are not only false but also are believably true. Thus, when timed close to an election, a court may find that restraining the republication of defamatory deepfakes is a compelling state interest. This Section provides reasons that a court may justify this interest.

A. Preventing Voter Deception and Preserving an Election's Integrity.

When close to an election, a widely circulated defamatory deepfake may create a perfect storm of voter deception. Legal scholars Bobby Chesney and Danielle Citron point out how this storm would form. A deepfake's dissemination with a large "enough window for the fake to circulate but not enough window for the victim to debunk it effectively," could influence an election's outcome by creating a "[a] narrow window[] of time during which irrevocable

---

[79] Garrison v. Louisiana, 379 U.S. 64, 75 (1964).

[80] Thomas M. Franck & James J. Eisen, *Balancing National Security and Free Speech*, 14 N.Y.U. J. INT'L L. & POL. 339, 343 (1982) ("[a] diligent court would ask whether there are weightier countervailing interests."). *But see* Dennis v. United States, 341 U.S. 494, 524-25 (1951) (Frankfurter, J., concurring) ("the demands of free speech in a democratic society as well as the interest in national security are better served by candid and informed weighing of competing interests, within the confines of judicial process, than by announcing dogmas too inflexible for the non-Euclidian problems to be solved.").

[81] *Compare* Garrison v. Louisiana, 379 U.S. 64, 74–75 (1964) ("speech concerning public affairs is more than self-expression; it is the essence of selfgovernment"), *with* Ashutosh Bhagwat, *Details: Specific Facts and the First Amendment*, 86 S. CAL. L. REV. 1, 35 (2012) (citing Borough of Duryea v. Guarnieri, 564 U.S. 379 (2011)). ("[i]ndeed, the votes and statements of the Justices in Guarnieri indicate that all of the current Justices accept the basic premise that the First Amendment's Free Speech Clause is preeminently concerned with the democratic process, and that speech relevant to self-governance receives greater protection than other forms of speech.").

[82] *See Alvarez*, 567 U.S. at 709 ("falsity alone may not suffice to bring speech outside the First Amendment."); *see also* Harry Kalven, JR., A WORTHY TRADITION: FREEDOM OF SPEECH IN AMERICA 10 (1988) ("the state is not to umpire the truth or falsity of doctrine; it is to remain neutral.").

[83] *See* 52 U.S.C.A. § 30104 (West 2026) (setting out campaign financing reporting requirements); McConnell v. Fed. Election Comm'n, 540 U.S. 93 (2003) (holding that the Bipartisan Campaign Reform Act is constitutional).

[84] William P. Marshall, *New Issues in the Law of Democracy: False Campaign Speech and the First Amendmen*t, 153 U. PA. L. REV. 285, 285 (2004).

decisions are made, and during which the circulation of false information therefore may have irremediable effects."[85]

Depending on where the defendant disseminates the content, 47 U.S.C. § 230 or the "Communications Decency Act" may widen this window. § 230 can provide immunity to interactive computer service providers from a declaratory injunction even when a claim is not directly against them.[86] Thus, the defendants themselves would have to depublish, which, in close proximity to an election, may take longer if they do not control the platform on which they publish.

Within this temporal window, the effect on voter perception emerges with voters now more susceptible to "confusion" and "undue influence."[87] Psychological research explains that individuals accept false information as true; subsequent corrections often fail to undo its effects fully.[88] And this persistence is not limited to lay audiences; even experts can remain influenced by debunked claims.[89] In the political context, studies of the 2006 Senate elections suggest that the resulting distortions in voter belief can be effectively irreversible, underscoring the durable harm that deepfake-driven misinformation can inflict.[90]

This, in turn, undermines a direct democracy's legitimacy.[91] A misinformed vote would not truly represent the will of the people. Such speech would instead "turn on rumors, innuendo, and outright fabrication, in effect defeating the entire electoral process."[92] Thus, as legal scholar Rebecca Green notes, "[e]nsuring faked political speech is not circulated without disclaimers clearly identifying it as such, like requiring campaign finance disclosure, fulfills an important government interest of providing accurate information to voters."[93]

In *The Babylon Bee, LLC v. Lopez*, the Hawaii District Court found this to be the case when applying strict scrutiny to Haw. Rev. Stat. Ann. § 11-302, which creates a civil action

---

[85] Chesney & Citron, *supra* note 2, at 1778; *accord* Jack Langa, *Deepfakes, Real Consequences: Crafting Legislation to Combat Threats Posed by Deepfakes*, 101 B.U.L. REV. 761, 765 (2021).

[86] For example, see: Hassell v. Bird, 5 Cal. 5th 522 (2018); *see also* Ben Ezra, Weinstein, & Co. v. Am.Online Inc., 206 F.3d 980, 983-86 (10th Cir. 2000); Medytox Sols., Inc. v. Investorshub.com, Inc., 152 So. 3d 727, 731 (Fla. Dist. Ct. App. 2014); Noah v. AOL Time Warner, Inc., 261 F.Supp.2d 532, 540 (E.D. Va. 2003).

[87] *See Burson*, 504 U.S. at 199 (finding that states have "a compelling interest in protecting voters from confusion and undue influence.").

[88] *See, e.g.*, Craig A. Anderson et al., *Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information*, 39 J. PERSONALITY & SOC. PSYCHOL. 1037 (1980) (confirming, through a study, that social theories can survive the total discrediting of initial evidential base); Tobias Greitemeyer, *Article Retracted, but the Message Lives on*, 21 PSYCHONOMIC BULL. REV. 557, 557 (2014) (confirming, through a study, that "individuals still believe in the findings of an article even though they were later told that the data were fabricated and that the article was retracted."); Charles G. Lord, Lee Ross, & Mark R. Lepper, *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*, 37 J. PERSONALITY & SOC. PSYCHOL. 2098 (1979) (same); *see also* Chip Heath & Dan Heath, *Made to Stick: Why Some Ideas Survive and Others Die* (2007) (discussing the reasons behind why some stories or messages persist in people's memories).

[89] Rachel C. Vreeman & Aaron E. Carroll, Medical Myths, 335 BRIT. MED. J. 1288 (2007).

[90] John G. Bullock, *The Enduring Importance of False Political Beliefs*, at 3 (2006); *see also* Michael Cobb et al., *Beliefs Don't Always Persevere: How Political Figures Are Punished When Positive Information About Them Is Discredited*, 34 POL. PSYCHOL. 307, 307 (2013).

[91] *See* Becky Kruse, Comment, *The Truth Masquerade: Regulating False Ballot Proposition Ads Through State Anti-False Speech Statutes*, 89 CALIF. L. REV. 129, 150 (2001).

[92] Alvin I. Goldman & Daniel Baker, *Free Speech, Fake News, and Democracy*, 18 FIRST AMEND. L. REV. 66, 136-37 (2019); *see also* Winsbro, *supra* note 31, at 863.

[93] Green, *supra* note 31, at 1462.

for "materially deceptive" deepfake political ads.[94] While the court held that the statute was content-based, not narrowly tailored, and failed strict scrutiny, it still found that the statute served a compelling state interest in "protecting the State's electoral integrity—an essential democratic function," crediting evidence that deepfakes erode trust in social media news, weaken democratic norms, and can cause people to dismiss authentic evidence.[95]

Likewise, in *Kohls v. Bonta*, the Eastern District of California found that Cal. Elec. Code §§ 20012(b)(1), 20012(d), a near mirror image to § 11-302 failed strict scrutiny but served the compelling interest of protecting a ballot's integrity.[96]

Nevertheless, scholarship has suggested that prospective speech restrictions that prevent voters from being led astray are not a significant enough threat to democracy because "those who hear the statements . . . are too lazy or dim-witted to sort out truth from falsehood."[97]

However, this assertion assumes that all voters will fail to sort facts from falsity or at least a large enough portion for a speech regulation not to affect an election's integrity. Alternatively, even if voters do not choose to actively sort, they will do so passively and subconsciously by absorbing political messages through repetition. Psychological studies suggest that the more times a message is repeated, the more likely it is to be believed, regardless of its truth.[98]

Thus, because democracy depends on an informed electorate,[99] an election's integrity is tainted even if voters make no effort to fact-check. As such, if widespread enough, a court may plausibly find that preventing voter deception and preserving democratic choice is a valid reason to find that a restraint is a compelling interest.

B. Preserving a Functional Marketplace of Ideas.

Courts may also confirm their interest in preventing the dissemination of defamatory deepfakes by preserving a functional and free "marketplace of ideas," untainted by defamatory campaign speech. This concept dates back to Justice Holmes's 1919 dissent in *Abrams v. United States*, which noted that "The theory of our Constitution is 'that the best test of truth is the power of the thought to get itself accepted in the competition of the market.'"[100] Since Holmes' dissent, courts and scholarship alike have embraced this metaphor[101] regarding speech suppression issues.

---

[94] The Babylon Bee, LLC v. Lopez, No. 25-00234 SASP-KJM, 2026 U.S. Dist. LEXIS 41476 (D. Haw. Jan. 30, 2026).

[95] *Id.* at 36.

[96] Kohls v. Bonta, 797 F. Supp. 3d 1177, 1185-86 (E.D. Cal. 2025).

[97] Steven G. Gey, *The First Amendment and the Dissemination of Socially Worthless Untruths*, 36 FLA. ST. U. L. Rev. 1, 21 (2008); Winsbro, *supra* note 31, at 859.

[98] *See* Jessica Udry & Sarah J Barber, *The illusory truth effect: A review of how repetition increases belief in misinformation*, CURRENT OPINION IN PSYCHOL. (2024); Aumyo Hassan & Sarah J. Barber, *The effects of repetition frequency on the illusory truth effect*, 6 COGNITIVE RESEARCH (2021); Alice Dechêne, et al., *The truth about the truth: a meta-analytic review of the truth effect*, 14(2) PERSONALITY & SOC. PSYCHOL. REV. 238 (2009).

[99] Marcia Clemmitt, *Lies and Politics: Do Politicians Lie More Today?*, 21 CQ RESEARCHER 147, 148 (Feb. 18, 2011).

[100] Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

[101] Although courts treat the "marketplace of ideas" as a metaphor, scholars have argued that it operates as an economic market—and, like any market, is vulnerable to failure. *See* Alvan I. Goodman, *Knowledge in a Social World* 196 (1991). In terms of an economic "marketplace," all speech—whether false or true—acts as a product in

Generally, a free marketplace is functional when it allows for multiple views to occupy an open discourse.[102] However, as legal scholar Rebecca Green notes, a functional "'marketplace of ideas' depends on our ability to validate and invalidate competing claims";[103] thus, if voters cannot readily verify defamatory speech as real or fake, the marketplace is impaired.[104] Given that deepfakes are widely accessible and exploit cognitive biases,[105] they may undermine the conditions necessary for a functioning marketplace of ideas, just as they threaten the integrity of elections.

Some may greet this contention with the fact that even if this were the case, legal suppression is still unnecessary because individuals such as informed individuals would engage in the sorting process regardless of government regulation.[106] But in a litigation context, a court may find this too risky given the high stakes of an election where the window for meaningful correction is extremely narrow and any post hoc debunking may arrive only after voters have already formed—and acted on—misleading impressions. In that setting, the inability of ordinary fact-checking mechanisms to restore the status quo ante before ballots are cast may weigh in favor of treating narrowly tailored injunctive relief as necessary to prevent irreversible democratic harm.

Furthermore, defamatory campaign speech may offer little value to a free marketplace because it "lower[s] the quality of campaign discourse and debate."[107] As the Court noted in *Garrison*, false and calculated falsehoods "are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality."[108]

---

which true speech is a better product. Speakers act as producers, and consumers are only those who believe the speech. Over time, a consumers' preference for the better product will cause true speech production to outweigh false speech, thereby creating an optimal market. However, as social epistemologist Alvin Goodman notes, that model assumes consumers prefer true speech to false speech, even when the latter is the better product. *Id.* at 97-98. In the context of elections, false speech may be more aligned with a consumer-voter's beliefs. Thus, the free market fails to create optimality. Regarding the economics of defamatory speech, regulation may even create more optimality. One may argue that regulation comes at the cost of chilling speech production, and market mechanisms such as corrections and rebuttals may achieve the same outcome that a defamed plaintiff seeks without chilling production. However, as Goodman notes, corrections to defamatory speech and rebuttals cannot be effective if they do not reach the targeted consumers and may not succeed in correcting beliefs. *Id.* Consequently, the functional free marketplace breaks down. As such, regulation may be a compelling state interest under this interpretation of a "marketplace."

[102] Goodman, *supra* note 101, at 209.

[103] Green, *supra* note 31, at 1459; *see also* Robert Post, Participatory Democracy and Free Speech, 97 Va. L. Rev. 477, 479, (2011) ("[t]he creation of knowledge . . . depends upon practices that continually separate the true from the false, the better from the worse.").

[104] *Id.*; *see also* Lang, *supra* note 85, at 781 ("persuasive deepfakes are tantamount to 'false statements of fact' and therefore "interfere with the truth-seeking function of the marketplace of ideas.") (citation removed).

[105] Lang, *supra* note 85, at 766.

[106] *Cf.* Eugene Volokh, Response, *In Defense of the Marketplace of Ideas /Search for Truth as a Theory of Free Speech Protection*, 97 VA. L. REV. 595, 598 (2011) (""University professors, think tank researchers, informed citizens, and others are [already] constantly engaging in a process through which truth and falsehood are separated."

[107] Marshall, *supra* note 84, at 294

[108] *Garrison*, 379 U.S. at 75 (quoting Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942)).

Likewise, the permission of such speech forces the defamed candidates to respond with their own defamation, creating cycles of defamatory attacks.[109] The noise generated by this cycle can undermine public confidence in the electoral process. As one political consultant put it, "If . . . every carrier in the airline industry ran commercials about how many people were killed in competitors' plane crashes—and the competition responded in kind—nobody would feel safe driving or flying anywhere."[110]

As such, when close to an election, a court may plausibly justify a restraint based on its lack of impact on the metaphorical free marketplace, or even on preserving the functional nature of that marketplace, in addition to preventing voter deception. As such, regulation may be a compelling state interest. But even assuming so, a prior restraint must still be narrowly tailored to survive constitutional scrutiny. The next Part explores the need for narrow tailoring and how courts may do so.

IV. A NARROW TAILORING: CONSIDERATIONS FOR COURTS

Expanding post-judgment injunctions in the election context risks creating a pathway for broader speech suppression, particularly when parties ask courts to enjoin categories of false or misleading political speech. But that risk arises only if courts detach such injunctions from the limiting principles that make them constitutional. Hence, the need for a narrowly tailored injunction.[111]

Yet existing First Amendment doctrine offers courts little guidance on how to conduct that inquiry in the context of defamatory deepfake election speech. Traditional strict scrutiny does not account for the unique combination of factors present here: adjudicated falsity, audiovisual realism, and the compressed temporal window in which such speech can distort voter decision-making. This Part sets out considerations for courts to determine whether an injunction will survive strict scrutiny.

A. Temporal Considerations.

Timing is critical. The harm is not just repetition, but the risk of distorting votes before corrections reach voters. However, an injunction that lasts beyond election day or too far in advance could be overbroad. Currently, deepfake regulations like the Texas' Deepfake Act and California's Elections: Deceptive Audio or Visual Media Act,[112] provide explicit guidelines from the time of voting day.

---

[109] Marshall, *supra* note 84, at 294; Shanto Iyengar & Jennifer McGrady, MEDIA POLITICS 150, 168 (2007) ("[t]he most compelling explanation of negative campaigning is that one attack invites a counterattack, thus setting in motion a spiral of negativity.").

[110] Ed Rollins with Tom DeFrank, BARE KNUCKLES & BACK ROOMS: MY LIFE IN AMERICAN POLITICS 350 (1996).

[111] Sypniewski v. Warren Hills Reg'l Bd. of Educ., 307 F.3d 243, 266 (3d Cir. 2002) (quoting Baggett v. Bullitt, 377 U.S. 360, 372 (1964)) (A prior restraint cannot withstand strict scrutiny unless it is "narrowly confined,"otherwise "[speakers] are left without 'fair notice' of the regulation's [or injunction's] reach. Commonly, this uncertainty will lead them to 'steer far wider of the unlawful zone than if the boundaries of the forbidden areas were clearly marked.'").

[112] TEX. ELEC. CODE § 255.004 (LexisNexis 2026); CAL. ELEC. CODE § 20010 (Deering 2026)

Nevertheless, as one scholar notes, "an effective solution must allow for flexibility by considering the context in which a particular deepfake is created and disseminated."[113] Accordingly, courts should not hesitate to consider the modern realities of both early voting and absentee ballots, in which electoral decisions are often made before traditional counterspeech mechanisms can meaningfully operate.

To strike a balance between flexibility and rigidity, courts may limit the duration of the injunction to the relevant electoral cycle, beginning upon final adjudication and terminating once the election has concluded and voters can no longer cast ballots or alter them.[114] This temporal limitation reflects the reality that the constitutional harm at issue is not the mere existence of false speech, but its capacity to influence voter decision-making during the period when electoral choices remain unsettled. Because post-judgment relief is typically issued only after expedited litigation in close proximity to an election, such a limitation ensures that the injunction operates only during the narrow window when corrective measures such as damages or counterspeech are least effective. In doing so, the restriction minimizes the duration of speech suppression while directly targeting the period in which the risk of electoral distortion is most acute.

B. Protected Speech Considerations.

To ensure that protected speech is not inadvertently swept within the scope of an injunction, courts must confine relief to the specific adjudicated defamatory deepfake content and materially identical audiovisual replications. Even "[a] clear and precise enactment [or injunction] may nevertheless be 'overbroad' if in its reach it prohibits constitutionally protected conduct."[115]

Accordingly, the injunction should not extend to expressive speech that merely references, critiques, or materially alters the underlying subject matter, including true speech[116] and parody,[117] or to expressive works that do not constitute adjudicated false audiovisual impersonations of the plaintiff. This limitation is necessary because First Amendment protections attach not to the perceived realism of expression,[118] but to its function. A narrowly tailored injunction must therefore distinguish between a false

---

[113] Langa, *supra* note 85, at 791.

[114] Although elections are often treated as discrete endpoints for remedial purposes, courts should recognize that electoral processes may extend beyond election day through early voting, absentee ballots, recounts, and runoff elections, which might warrant a reinstatement of the injunction.

[115] Grayned v. City of Rockford, 408 U.S. 104, 114 (1972).

[116] Phila. Newspapers v. Hepps, 475 U.S. 767, 778 (1986).

[117] "Parodies," or statements that do not depict false factual statements are inherently excluded from defamation claims because they are not false statements of *fact*. *See Hustler Magazine*, 485 U.S. at 57 (1988); *see generally* Milkovich v. Lorain Journal Co., 497 U.S. 1 (1990); Greenbelt Coop. Publ'g Ass'n v. Bresler, 398 U.S. 6 (1970).

[118] *See Hustler Magazine*, 485 U.S. at 54; Farah v. Esquire Magazine, 736 F.3d 528, 536 (D.C. Cir. 2013). In terms of deepfakes, "Prohibiting demonstrably fake videos that do not actually persuade or deceive viewers or those that are made to parody or satire would chill free speech and violate existing precedent. However, a highly persuasive deepfake that actually deceives viewers and is intended to influence a voter's decision or undermine national security would likely fall outside the bounds of protected speech." Langa, *supra* note 85, at 785; *see generally* X Corp. v. Bonta, 116 F.4th 888 (9th Cir. 2024) ("[w]hen a state 'compel[s] individuals to speak a particular message,' the state 'alter[s] the content of their speech,' and engages in content-based regulation" such that, "[e]ven a pure 'transparency' measure, if it compels non-commercial speech, is subject to strict scrutiny") (citations omitted).

impersonation of reality and independently protected expressive uses of similar or derivative content.

C. Least Intrusive Means Considerations.

A narrowly tailored injunction must be the least restrictive effective means of achieving the government's interest. Accordingly, courts should evaluate whether less speech-restrictive alternatives, such as ordering disclaimers, would adequately mitigate the harm caused by adjudicated defamatory deepfakes.

For example, in *Babylon Bee*, the court found that a Hawaii state law that created a civil action for deceptive deepfake political advertisements was not narrowly tailored because instead of suppression, the state could have increased literacy on identifying deepfakes, "'counter[ed] deceptive speech with factual speech of its own,' or it could [have] start[ed] a government database or committee dedicated to tracking and flagging materially deceptive content."[119]

However, when an ad is disseminated in close temporal proximity to an election, ordering disclaimers may fail to cure the harm when the election is too close in time to reach voters meaningfully. When the record demonstrates that these alternatives cannot meaningfully prevent continued electoral distortion during the relevant voting period, courts may find that injunctive relief constitutes the least restrictive effective means of remedying the harm.

V. TYING IT ALL TOGETHER: A PROPOSED TEST TO DETERMINE WHEN AN INJUNCTION SURVIVES STRICT SCRUTINY

When applying these considerations to an instant case, courts may create a test that incorporates the factors to determine a narrow tailoring that resembles the following:

A post-judgment injunction that prohibits the republication of a computer-generated political advertisement that may reasonably mislead the public that the speech falsely depicts statements that the plaintiff made shall only survive constitutional speech scrutiny when:

(1) The court that orders the injunction has issued a final adjudication on the merits that the suppressed speech is defamatory.
(2) The suppressed speech relates to an upcoming election
(3) The injunction only begins upon final adjudication that the speech is defamatory.
(4) The injunction terminates once the election concludes and voters can no longer cast or alter ballots, and shall not extend beyond that point.
(5) The voting day is so close in time that no other reasonable method to reduce confusion will prevent voter deception. Such methods include modifying the speech's dissemination to include an explicit written disclaimer before the speech begins that the speech is false and in no way reflects the views of the plaintiff, publicized methods to identify that the suppressed speech is false, or disseminating the plaintiff's own speech that may notify voters that the suppressed speech is both false and does not reflect their views.

---

[119] *Babylon Bee*, 2026 U.S. Dist. LEXIS 41476, at *40-43.

(6) The injunction suppresses only the republication of the specific adjudicated defamatory deepfake content and materially identical audiovisual reproductions, and is narrowly confined so as not to suppress any other speech that has not been adjudicated defamatory.

If an injunction passes this test, a court may reasonably find that it survives strict scrutiny.

However, the instance in which these injunctions are appropriate may be limited. Even assuming an injunction passes, the recourse may be inadequate. Defamation proceedings take time. A court may only be able to issue and injunction after and election and the damage is already done. Further, a candidate would have little to no success if they were to ask the court to issue a preliminary injunction. Courts are hostile towards them, with some not even recognizing them as a possibility.[120]

Nevertheless, if ordered in time, post-judgment injunctions may still pass strict scrutiny, serve to mitigate harm and, despite their limits, may preserve the fabric of our democracy.

VI. CONCLUSION

As artificial intelligence continues to blur the boundary between fabrication and reality, defamatory deepfake election content poses a novel challenge to prior restraint jurisprudence. But if narrowly tailored, post-judgment injunctions prohibiting republication of defamatory deepfake election content can satisfy strict scrutiny in jurisdictions that treat such orders as prior restraints. When properly constrained, these injunctions are not censorship; they are remedies that prevent proven falsehoods from distorting democratic choice.

[120] For examples of courts refusing to issue preliminary injunctions suppressing speech, see: Studiorotan LLC v. Howell, No. 2:25-CV-20-GMB, 2025 WL 3041940, at *11 (N.D. Ala. Oct. 31, 2025); Banks v. Jackson, No. 120CV02074DDDKMT, 2020 WL 6870739, at *2 (D. Colo. Oct. 2, 2020); McCarthy v. Fuller, 810 F.3d 456, 462 (7th Cir. 2015); Auburn Police Union v. Carpenter, 8 F.3d 886, 903 (1st Cir. 1993).

Validity of the Case Method in Undergraduate Legal Education: An Empirical Study at a Japanese University

Masamichi Yamamoto*, Chika Y. Rosenbaum** and Katsunobu Sasanuma***

* Professor of Law, Nagoya University of Commerce and Business, Japan; S.J.D., The University of Iowa College of Law; J.D., Vanderbilt University Law School.
ORCID ID: 0009-0005-3323-9696
** Associate Professor of Political Science, Chukyo University, Japan; Ph.D., University of Missouri.
ORCID ID: 0000-0003-4641-050X, 0009-0000-3330-7736
*** Professor of Management, Chuo University, Japan; Ph.D., Carnegie Mellon University.
ORCID ID: 0000-0001-5926-9045

Abstract

Since Professor Christopher Columbus Langdell first developed the Case Method in the late nineteenth century, many have written whether the Case Method is still valid at today's law school in the United States: some argue that the use of the Case Method is declining, and the others argue that it is still the dominant method. In contrast, lecture is the mainstream teaching method in Japan where the main source of law is statutory law. Our survey demonstrates that the Case Method is effective for undergraduate non-law students to obtain legal knowledge and vital meta-skills. Based on the ordered logit analysis, we propose improvements in undergraduate law course design, such as choosing appropriate and interesting cases, mitigating stress that students may have about learning, and giving students active assignments. Such modified Case Method is also useful for students who plan to go to graduate law schools to prepare for rigorous Socratic teaching.

Keywords: Case Method, Legal Education, Undergraduate Education, Japan

Table of Contents

I. Introduction

While practicing as a lawyer, one of the authors, Masamichi Yamamoto, has taught international business law at a Japanese graduate law school[1] as an adjunct instructor for more than ten years, using the case method (the "Case Method").[2] Much like how law is taught at U.S. law schools, before the class Yamamoto gives students assignments to read, during the class he asks a student to state the facts of a case, and then asks another student to provide the plaintiff's arguments . . . . Though it is oftentimes challenging for Japanese students to speak up, the Case Method at the Japanese *graduate* law school would usually work well because most students, who aspire to become a lawyer, come to class fully prepared for the Socratic discussion.[3]

Three years ago, Yamamoto became a full professor at a Japanese university, Nagoya University of Commerce and Business ("NUCB") to teach law courses as liberal arts subjects for *undergraduate* students. As we explain later, the mainstream method to teach law at Japanese

[1] Unlike the universities in the United States, many Japanese universities have an undergraduate law department, but did not have any graduate law school until 2004. *See* Masamichi Yamamoto, Note, *How Can Japanese Corporations Protect Confidential Information in U.S. Courts? Recognition of the Attorney-Client Privilege for Japanese Non-Bengoshi In-House Lawyers in the Development of a New Legal System*, 40 VAND. J. TRANSNAT'L L. 503, 518 (2007) ("In April 2004, sixty-eight law schools modeled on the U.S. legal education system opened for the first time in Japan's history.").

[2] The term "case method" is ambiguous, and it may refer to either the Law School Case Method, the Business School Case Method, the Socratic Method, or the Problem Method. *See* Gregory J. Marsden & Soledad Atienza, *Teaching During a Changing Administration: Doing Law School Wrong: Case Teaching and an Integrated Legal Practice Method*, 66 ST. LOUIS L.J. 543, 547-48 (2022) (describing the differences of four methods). For the purpose of this article, to include all of the four methods above, we define the Case Method as "a teaching method in which a professor calls on students and engage in a dialogue over a case, which includes not only court decisions but also actual or fictious problems."

[3] The Socratic discussion means a teaching method used by law professors under which "the professor asks a student a series of questions designed to elicit information about the reading material, expose weaknesses in the student's thinking, and lead the student to the 'right' answer." Beth H. Wilensky, *Dethroning Langdell*, 107 MINN. L. REV. 2701, 2708 (2023).

universities is still the lecture, in which interactions between a professor and students are rare.[4] Yamamoto, however, had to use the Case Method because at NUCB, all professors are "required" to use the Case Method, ideally using at least one case per class (here, a "case" is not limited to a court decision, but includes any real or fictious problems). At first, he wondered how the Case Method would work to teach undergraduate students who have no legal background and not much intent to become a lawyer. It eventually seemed to work to some degree, but he could not explain how and why it worked. This Article tries to give some empirical insight to answer how the Case Method is effective to teach law in undergraduate classrooms. Although we collected data from a survey conducted at a Japanese undergraduate school, we believe our findings and recommendations are generally applicable to undergraduate legal education worldwide.

Since Professor Christopher Columbus Langdell first developed the Case Method in the late nineteenth century,[5] many have written whether the Case Method is still valid at today's U.S. law school: some argue that the use of the Case Method is declining[6] and the others argue that it is still the dominant method.[7] Some studies also have shown empirical evidence to support their particular argument: for example, Shadel et al. demonstrated that women were less likely to speak up through the Case Method.[8] To the best of our knowledge, however, systematic research to show the relationship between the Case Method and students' learning has been very limited, even by business school scholars.[9]

This Article, therefore, tries to demonstrate that the Case Method is effective for undergraduate students in law courses to gain legal knowledge and vital meta-skills,[10] such as the

---

[4] *See infra* Part II.B (describing Japanese legal system and teaching method).
[5] *See infra* notes 14-17 and accompanying text (describing how and when Langdell developed the Case Method).
[6] *See e.g.*, Orin S. Kerr, *The Decline of the Socratic Method at Harvard*, 78 NEB. L. REV. 113, 113 (1999) (arguing that "the Socratic Method as it was known in the 1950s and 1960s is nearly extinct").
[7] *See e.g.*, Beth Hirschfelder Wilensky, *Dethroning Langdell,* 107 MINN. L. REV. 2701, 2701 (2023) (observing that the case method "is still the dominant approach to pedagogy in many law school classrooms").
[8] Molly B. Shadel, et al., *Gender Differences in Law School Classroom Participation: The Key Role of Social Context*, 108 VA. L. REV. ONLINE 30 (2022); *see also* Jane B. Grise and Dorothy Evensen, *Getting it Right from the Start*, 91 TENN. L. REV 53 (2023) (demonstrating with empirical evidence that a theory-based program will enhance students' ability to lean through the Case Method).
[9] *See* George Rosier, *The Case Method Evaluation in Terms of Higher Education Research: A Pilot Study,* 20 (3) INT'L J. MGMT. 100660, 1 (2022) ("Case method teaching is used extensively in business schools around the world, and has become almost a hallmark of the top business schools. It is therefore surprising to find that there is so little empirical evidence to support its continued use."); *see also* Hiroshi Ito & Shinichi Takeuchi, *The demise of active learning even before its implementation? Instructors' understandings and application of this approach within Japanese higher education*, EDUC. INQUIRY 1 (2020) ("[T]o date, no large-scale empirical studies have been conducted to elicit their understanding and practice of active learning.")
[10] Professor Nohria, the former Dean of the Harvard Business School, argues that students learning through the Case Method should be able to obtain a "group of long-lasting abilities that allow

abilities to prepare, judge, and collaborate, based on the survey taken at NUCB.[11] Part II of this Article discusses how the Case Method has been developed and is used today at U.S. law schools and explains how law courses are taught through the Case Method at NUCB. Part III then provides how we designed the survey and what inferences we drew from the results. Part IV concludes this Article by summarizing our findings and making recommendations for improving undergraduate law courses.

II. The Case Method in History and Today

Although there have been lively debates over the effectiveness of the Case Method,[12] the Case Method is still the main teaching method at most U.S. law schools.[13] In this part, we first explore the history of the Case Method and introduce pros and cons of using the Case Method at law school, and show that, while the use of the Case Method is somewhat declining and the Case Method has been modified by many teachers until today, it is still active. Next, we describe how law is taught in Japan, which has a completely different legal system from that of the United States. Finally, we explain how Yamamoto, a law professor, modifies and utilizes the Case Method for undergraduate students at NUCB.

A History of and Reasons for Using the Case Method

Until 1870 when Professor Langdell was appointed to chair of the Harvard Law School, the lecture and treatise memorization had been the mainstream of law teaching methods in the United States.[14] Under the "text-book method", students were assigned a part of a textbook to memorize rules and then quizzed in class to "cram young lawyers."[15] Langdell, however, changed the teaching method dramatically. He proposed that students learn by reading cases as a source of law and engaging

---

someone to learn new things more quickly" including the abilities to prepare, discern, recognize bias, judge, collaborate, become more curious, and be confident. Nitin Nohria, *What the Case Study Method Really Teaches*, HARV. BUS. REV. (2021).

[11] The authors disclose that we have published a separate article, based on the same survey and the larger data, focused on the relationship between the Case Method and students' learning in politics and law. *See* Chika Y. Rosenbaum et al., *Acquisition of Knowledge and Meta-Skills through the Case Method in Politics and Law Classrooms: New Empirical Insight from Japan*, J. POL. EDUC. (2024). In contrast, this Article focuses on how the Case Method has been developed and modified in legal education and whether it is valid in today's undergraduate legal education.

[12] *See* Grise & Evensen, *supra* note 8, at 56 n. 7 (listing articles arguing the effectiveness of the Case Method).

[13] *Id.* at 56; *see also* Jeannie S, Gersen, Essay, *The Socratic Method in the Age of Trauma*, 130 HARV. L. REV. 2320, 2347 (2017) ("Despite the well-developed consensus that legal education must change to become more practical, the appeal and relevance of Socratic pedagogy lies still in what Langdell first understood.").

[14] Grise & Evensen, *supra* note 8, at 58.

[15] *Id.*

in discussions about the cases.[16] He also changed the dynamics of the classroom by using the Socratic dialogue where he would ask students to state the facts of a case, make the plaintiff's argument, and solve problems.[17]

Langdell's method consists of three separate pedagogical techniques.[18] First, it involves the "case method" under which the professor assigns "appellate court opinions from which students discern aspects of legal doctrine, analyze that doctrine, and apply it to different scenarios."[19] Second, through the "Socratic method . . . the professor asks a student a series of questions designed to elicit information about the reading material, expose weaknesses in the student's thinking, and lead the student to the 'right' answer."[20] Third, the professor uses "cold calling" to "select and call on students to answer questions out of the blue instead of seeking volunteers."[21]

In this Article, we define the Case Method as "a teaching method in which a professor calls on students and engage in a dialogue over a case, which includes not only court decisions but also actual or fictious problems" to include all of the above three pedagogical techniques (i.e., the case method, the Socratic method, and the cold calling). A teaching method used by business schools is also called "case method."[22] At business schools, a "case is a narrative, written by business school faculty and ideally based on an actual problem faced by a real business enterprise"[23] and "cases are action-oriented and business school students are placed in the role of a manager who must make decisions that will impact the success of the enterprise."[24] We understand that Langdell's method and business school's case method are arguably different, but for the purpose of this Article, we define the Case Method as above to include both.[25]

Many scholars have argued for the benefits of using the Case Method at law school. First, for example, the Case Method is useful to teach students how to read and analyze cases, which skills

---

[16] *Id.* at 59.
[17] *Id.* at 60.
[18] Beth H. Wilensky, *Dethroning Langdell*, 107 MINN. L. REV. 2701, 2708 (2023). Kerr defines "traditional" Socratic method as "a teaching style in which the professor selects a single student without warning and questions the student about a particular judicial opinion that has been assigned for class. Orin S. Kerr, *The Decline of the Socratic Method ad Harvard*, 78 Neb. L. Rev. 113, 114 n.3 (1999). This definition includes the three pedagogical techniques.
[19] Wilensky, *supr*a note 18, at 2708.
[20] *Id.*
[21] *Id.*
[22] Gregory J. Marsden & Soledad Atienza, *Teaching During a Changing Administration: Doing Law School Wrong: Case Teaching and an Integrated Legal Practice Method*, 66 ST. LOUIS L.J. 543, 544-45 (2022).
[23] *Id.* at 546.
[24] *Id.* at 547 (citing George J. Siedel, *Legal Complexity in Cross-Border Subsidiary Management*, 36 TEXAS INT'L L.J. 611, 614 (2001).)
[25] As we explain later, the Case Method also needs to be somewhat modified to accommodate undergraduate non-law students. *See infra* Part II.D (explaining how one of the authors, Yamamoto, is not purely following Langdell's method to teach at NUCB).

are vital as a lawyer.[26] Second, the Case Method will increase interest in students in leaning law because students will be attracted to the stories of human behavior described in a legal case.[27] Third, the Case Method helps students to acquire critical thinking skill, which is sometimes referred to as "thinking like a lawyer."[28] The Case Method is also praised as a method to help students learn the law through precedents, understand the legal process, cultivate moral imagination, and develop mental toughness.[29]

B Case Method at Today's Law School

On the other hand, many other scholars expressed concern that the Case Method has multiple shortcomings.[30] First, for example, the Case Method could result in students' obedience to the selected decisions and inadequate attention to the responsibility and ethics of lawyers.[31] Second, the Case Method emphasizes principles and doctrines but omits attention to fact finding, thus making students lose opportunities to face with reality.[32] Third, although "successful lawyering needs skills in various aspects including listening, fact investigation, interest clarification, negotiation and planning" with clients, the Case Method fails to teach such lawyering skills.[33]

Partly because of these shortcomings, around late 1990s, the Case Method had showed decline even at Harvard Law School where the Case Method was born.[34] Kerr interviewed a diverse group of twelve professors who teach first-year courses such as contracts and torts at Harvard Law School in 1997.[35] He found that there were three rough categories of approaches to the Case Method.[36]

Frist, five professors are "traditionalists" who teach "almost exclusively by the case method, calling on students without prior warning to have them discuss assigned cases in a one-on-one dialogue with professor."[37] Second, three professors are "quasi-traditionalists" who "mix elements of the

---

[26] Kuan-Chun Chang, *The Teaching of Law in the United States: Studies on the Case and Socratic Methods in Comparison with Traditional Taiwanese Pedagogy,* 4 NAT'L TAIWAN U. L. REV. 1, 13-14 (2009).
[27] *Id.* at 13.
[28] *Id.* at 14.
[29] *Id.* at 15-16.
[30] *See e.g.*, Edward Rubin, *What's Wrong with Langdell's Method, and What to Do about It*, 60 VAND. L. REV. 609, 616 (2007) (arguing that "the Langdell's model is severely out-of-date on many different fronts").
[31] Chang, *supra* note 26, at 17.
[32] *Id.* at 17-18.
[33] *Id.* at 18. Dickenson, however, argues that "genuine dialogue pedagogy [ ] instills essential professional qualities and skills in those law students who participate in its process." Joseph A Dickinson, *Understanding the Socratic Method in Law School Teaching after the Carnegie Foundation's Educating Lawyers,* 31 W. NEW ENG. L. REV. 97, 113 (2009).
[34] Kerr, *supra* note 18.
[35] *Id.* at 114, 122.
[36] *Id.* at 122.
[37] *Id.*

Socratic dialogue with alternative methods" such as warm calling (i.e., calling with prior warning) to reduce the "authoritarian" nature of the traditional method.[38] Third, four professors are "counter-traditionalists" who reject the Case Method and "substitute a variety of methods such as panel systems, lectures, and group problems, to create a classroom atmosphere designed to be less intimidating, less pressured, and more informative."[39] He concluded that, though its use is declining, the Case Method coexisted with various teaching method because "yesterday's students have become today's professors, and have carried with their perspectives and attitudes toward legal education from their student days."[40]

Although the Case Method is still the main teaching method at most U.S. law schools,[41] many scholars have proposed and utilized modified versions of the Case Method. For example, Abrams argues for "inclusive Socratic teaching" by changing "power-centered and professor-centered" method to "student-centered, client-centered, and community centered."[42] Similarly, Gersen argues that the "Socratic method works best as a form of cooperation and collaboration" and professors can make use of it "by putting students in active and productive dialogue with each other."[43] Drawing inference form business school case teaching, Marsden and Atienza recommend the "use of practical 'case' problems to teach law and legal skills [ ] to provide students not only with substantive adjective legal knowledge, but also with the skills necessary to begin the practice of law."[44]

C Japanese Legal System and Case Method

Before describing Yamamoto's teaching method at NUCB, we need to explain the Japanese legal system, based in civil law, which is completely different from the U.S. legal system, based in common law, to understand the modification of the Case Method necessary to teach law in Japan.[45]

---

[38] *Id.* at 123
[39] *Id.* at 124.
[40] *Id.* at 131.
[41] Grise & Evensen, *supra* note 8, at 56.
[42] Jamie R. Abrams, *Legal Education's Curricular Tipping Point Toward Inclusive Socratic Teaching*, 49 HOFSTRA L. REV. 897, 897 (2021).
[43] Gersen, *supra* note 13, at 2345.
[44] Marsden & Atienza, *supra* note 22, at 543; *see also* Cynthia G. Hawkins-Leon, *The Socratic Method-Problem Method Dichotomy: The Debate over Teaching Method Continues*, 1998 BYU EDUC. & L.J. 1, 2 (1998) (argues that the most appropriate teaching method is a combination of the Socratic method and the Problem Method, which asks students to apply rules of law to written fact patterns and discern a "correct" answer); Sherri L. Keene & Susan A. McMahon, *The Contextual Case Method: Moving Beyond Opinions to Spark Students' Legal Imaginations*, 108 VA. L. REV. ONLINE 72, 73 (2022) (arguing for the "contextual" Case Method under which students should not "read opinions in isolation, but in the broader context in which they arose"); Grise & Evensen, *supra* note 8, at 54 (arguing that because "the majority of students enter law school unprepared to lean via the case method," it is important to have a theory-based program "to enhance the cognitive and metacognitive capabilities of students to lean through the case method")
[45] *See* Chang, *supra* note 26 at 6 & n. 6 (comparing American legal system and Taiwanese legal

The major source of law is case law in the United States, so it is important to understand how to interpret and apply case law to relevant facts.[46] In contrast, the major source of law in Japan is statutory law, so it is important to understand how to interpret and apply statutory law to relevant facts.[47] This inherent difference in legal systems significantly affects how to educate students to become a lawyer.[48]

The Case Method may not work as it is at law schools in civil law countries such as Japan and Taiwan. Statutory law is not "merely a collection of statutes" but instead "a highly sophisticated as well as organized and systematic treatment of an entire body of law."[49] To understand statutes, it "is not enough to merely read and interpret the text" but "also necessary to have a solid perception of the underlying conceptual issues that the code is meant to address."[50] "The legal culture of codification and emphasis on conceptual issues" shape the legal education at Japanese universities where professors still give formal lectures and typically offer "an organized, abstract, one-way presentation."[51] This traditional teaching method "may be found helpful in memorizing interpretation of codes, scholarly theories, and legal knowledge, [but] it fails to elevate students' analytical or communication skills to a professional standard."[52]

D How Yamamoto Teaches Law at NUCB

This part explains how one of the authors and a law professor, Yamamoto, modifies the traditional Case Method to teach undergraduate students at NUCB. His methodology is affected by a couple of factors. First, he took law classes as a J.D. student at Vanderbilt University Law School in the United States about twenty years ago, so he has naturally "carried with [his perspective and attitude] toward legal education from [his] student days."[53] Second, as explained above, the main source of law in Japan is statutory law, so undergraduate students need to learn underlying conceptual issues that the code is meant to address.[54] Third, even though students at NUCB are generally used to take classes through the Case Method because NUCB requires all professors to use the Case Method, they are not prepared to take law courses and have almost no legal background and knowledge.[55]

---

system).
[46] *Id.* at 6.
[47] *See* Takahiro Saito, *International Conference on Legal Education Reform: The Tragedy of Japanese Legal Education: Japanese "American" Law Schools*, 24 WIS. INT'L L.J. 197, 198-99 ((explaining civil law and statutory system in Japan); *see also* Chang, *supra* note 26, at 6.
[48] *See* Chang, *supra* note 26, at 6.
[49] *Id.* at 29.
[50] *Id.*
[51] *Id.*
[52] Chang at 29.
[53] Kerr, *supra* note 18, at 131.
[54] *See supra* Part II.C (describing Japanese legal system and teaching method).
[55] There is an introductory law course called Introduction to Law, which is taught by Yamamoto and

We explained that there were three pedagogical techniques in the Case Method, namely, the case method, the Socratic method, and the cold-calling.[56] In Yamamoto's class, a case does not mean a court decision but rather, a narrative based on actual or fictious problem faced by real business enterprise or students. Yamamoto believes that because Japan does not have common law system, court's reasoning and facts of the case are not very important, at least for undergraduate non-law students. Students at NUCB are generally more business oriented, and tend to show more interest in cases that they feel closer to their life as a student or future businessperson.

Yamamoto keeps the Socratic method in his class because he believes that the "Socratic method works best as a form of cooperation and collaboration" and professors can make use of it "by putting students in active and productive dialogue with each other."[57] Yamamoto, however, doesn't cold call on students, but instead he asks volunteers to raise hands or asks small groups to discuss problems and have them present their opinions. These modifications partially come from his own experience of humiliation when he was "Kingsfielded" on the very first day at a U.S. law school.[58] He also gives credits to "wrong" answers if they contribute to the discussion of the class so that students will not be afraid of speaking up and making mistakes, by cultivating less intimidating atmosphere.

III. Research Design and Result

One of the authors, Yamamoto, has taught international business law at a Japanese *graduate law school* through the Case Method, but he wondered how the Case Method would work to teach *undergraduate* Japanese students who have no legal background and very little intent to become a lawyer. This Article aims to provide empirical insight into the effectiveness of the Case Method to teach law in undergraduate classrooms. This part explains how the research was designed, what results we gained, and what inferences we can draw from the results. We also discuss some limitations in our research and propose improvements for further research.

A Research Design

another law professor, but it is not mandatory to have taken Introduction to Law for students to register in Civil Law or Corporate Law.

[56] *See supra* notes 18-21 and accompanying text (explaining three pedagogical techniques).

[57] He also uses lots of visual material such as movies to increase the interest of students.

[58] In the movie Paper Chase (1973), Professor Kingsfield started the first day of Contracts by cold calling a student (protagonist). Although the student confessed that he didn't read the case, Professor Kingsfield continued drilling the student. Yamamoto's first class was also Contracts. He read the case well before the class and his professor was much softer than Kingsfield, but his mind became a complete blank when he was called on and could not say a word.

In May 2024, we asked 300 undergraduate students at NUCB to answer a set of questions about their academic status, attitudes towards leaning, course preview, and personal preferences.[59] The online survey was distributed to students taking Civil Law I and Corporate Law I courses taught by Yamamoto. The survey was voluntary and anonymous. Both courses have a total of fourteen class sessions, each of which takes 100 minutes, and at least one case was assigned and discussed in each session. Yamamoto wrote all cases, based on either actual or fictious problems, which have been published by the Case Center Japan.[60]

There are 148 students registered for Civil Law I and 152 students registered for Corporate Law I. It is possible for students to register for both courses, so we asked students to take the survey only once if they registered for both. To ensure that students actually read and answer questions, we omitted all of the student's responses when a student chose an incorrect answer to the dummy question.[61] Some students did not take the survey because they were absent or refused to answer voluntarily. We successfully collected valid answers from a total of 155 students.

1 Dependent Variables

We first explain eight dependent variables—knowledge and seven meta-skills. We asked students to self-evaluate whether they acquired knowledge and seven vital meta-skills through the Case Method, by rating from one to four, based on how much they agree with the statements about their acquisition of knowledge and meta-skills (1: disagree, 2: somewhat disagree, 3: somewhat agree, and 4: agree).

We included the self-evaluated level of acquired knowledge because it is one of the most emphasized leaning outcomes through the Case Method.[62] We also included the following seven meta-skills that students are expected to acquire from the Case Method: (1) preparedness—an ability to be able to prepare, (2) discernment—an ability to identify and focus on what's essential, (3) bias recognition—an ability to recognize personal bias or listen to others with different viewpoints, (4) judgement—an ability to make and defend a decision, (5) collaboration—an ability to collaborate with other students, (6) curiosity—an ability to be curious and know what excites them, and finally (7) self-

---

[59] The Ethical Review Board of NUCB has approved the survey as complying with the ethical standard established by NUCB.

[60] Cases are available at https://casecenter.jp/ (last visited on Feb. 10, 2025).

[61] Some students might choose the same answer, for example choice 1, for all questions, without reading the question at all. To eliminate responses from such students, we asked students just to choose "5" in the second to the last question.

[62] *See* Jaime A. Bayona & Duran F. William, *A Meta-Analysis of the Influence of Case Method and Lecture Teaching on Cognitive and Affective Learning Outcomes*, 22 INT'L J. MGMT. EDUC. 3 (2024) ("Cognitive outcomes refer to the acquisition of knowledge and the understanding of information when exposed to new settings or novel data.").

confidence—an ability to be confident about their decisions.[63] The eight dependent variables are summarized in table 1 below.

Table 1: Dependent Variables and Descriptions

	Dependent Variables	Description (Students are asked to rate 4= if agree; 3= if somewhat agree; 2= if somewhat disagree; and 1= if disagree with the following statement.)
1	Level of knowledge	"After taking the course, I gained new knowledge about the subject."
2	Preparedness	"After taking the course, I gained the greater ability to prepare."
3	Discernment	"After taking the course, I gained the greater ability to identify what's important."
4	Bias Recognition	"After taking the course, I gained the greater ability to listen to others."[64]
5	Judgement	"After taking the course, I gained the greater ability to judge."
6	Collaboration	"After taking the course, I gained the greater ability to collaborate with others."
7	Curiosity	"After taking the course, I gained more curiosity about the subject."
8	Self-Confidence	"After taking the course, I gained more confidence in learning the subject."

2 Independent Variables

There are four categories of independent variables—experience, learning attitude, course preview, and personality. The first category "Experience" includes three independent variables. The first set of variables reflects students' year of study (1: freshman, 2: sophomore, 3: junior, 4: senior), which represents the students' level of experience in the Case Method. Because NUCB requires the Case Method in every single class offered, senior students should have had more opportunities in the past to acquire knowledge as well as the vital meta-skills through the Case Method. Two other independent variables in this category represent students' experience in discussion and presentation, which could be used to facilitate the Case Method. With more experience in these activities, we expected students would be better equipped to take on the Case Method and more likely to self-evaluate highly the acquisition of knowledge and vital meta-skills. Therefore, we also expected a positive correlation between students' acquisition of knowledge and meta-skills and the experience in

[63] *See generally* Nohria, *supra* note 10 (listing seven vital meta-skills that students gain from the Case Method).

[64] The "bias recognition" skill is measured with the students' ability to listen to others with different viewpoints. Nohria argues that through the Case Method, students should be able to notice their biased opinions when thinking of case stories, but he adds that this means students "learn to listen more carefully to classmates" with different viewpoints. *Id*.

discussion and presentation. In the survey, students were asked to rate from one to four about their experience in discussion and presentation (1: no experience, 2: little experience, 3: some experience, 4: substantial experience). These experiences could come from pre-college schooling and extracurricular activities.

The second category of independent variables represents students' attitude towards learning, which is repeatedly found to influence students' learning outcomes.[65] Students' learning attitude is measured with three independent variables. First, we included an explanatory variable measuring the level of students' interest in the course subject. In the survey, students were asked to rate from 1 to 4 about their interest level (1: not at all interested, 2: somewhat uninterested, 3: somewhat interested, 4: very much interested).[66] The second variable measures the level of stress students feel about learning and/or coursework. In the survey, students were asked to rate from 1 to 4 about their stress level (1: very much feeling stressed, 2: somewhat feeling stressed, 3: somewhat feeling not stressed, 4: not at all feeling stressed). The third variable gauges the level of focus students allocate for their coursework as part of their college life. Just like the other variables, students were asked to rate from 1 to 4 about their focus level (1: not at all focused or focused entirely on other things such as club activities and part-time jobs, 2: somewhat not focused or focused more on other things, 3: somewhat focused or focused less on other things, 4: very much focused in coursework).

The third category of independent variables represents "course preview", the level of preparation for the course students are enrolled in. We measure the level of preparation in terms of three different activities (reading cases, conducting relevant research, and submitting homework such as a short paper), all of which are indispensable parts of the particular learning process used in the Case Method.[67] In the survey, students were asked to rate their preparatory activities with the values ranging from one to four (1: no preparation, 2: little preparation, 3: some preparation, 4: substantial preparation).

The fourth and final category of independent variables represents students' personal characteristics regarding coursework. We included it because studies have shown that students' personalities affect their performances in classrooms.[68] In the survey, students were asked whether

---

[65] *See generally* Angela Lumpkin et al., *Student Perceptions of Active Learning*, 49 COLL. STUDENT J. 121, 129 (2015) (finding active and collaborative activities "engaged students and positively impacted leaning").

[66] Both Civil Law and Corporate Law are not "mandatory", but "elective" courses, so we expect students who took these courses should have some interest in the subjects. In reality, there are, however, other reasons to take these courses such as their belief that the courses are easy to get good grades or just filling a gap in their schedule.

[67] *See* Hiroshi Ito & Shinichi Takeuchi, *Instructors' Understanding, Practices, and Issues Regarding the Use of the Case Method in Higher Education*, 45 J. FURTHER & HIGHER EDUC. 1, 2 (2021) (describing students' preparation as one of the three essential components of the Case Method).

[68] *See generally* Fitri M. Hayati, *A Study on the Distinction Between Extrovert vs. Introvert in Learning English*, 14 (2) ENGLISH EDUC.: JURNAL TADRIS BAHASA INGGRIS 159 (2021) (comparing effect of extrovert and introvert personalities in learning English); Maureen A. Conard, *Aptitude Is*

they like to speak in public, study for exams, read, and do groupwork (1: entirely dislike it, 2: somewhat dislike it, 3: somewhat like it, 4: like it very much).

Table 2 below lists the categories and descriptions of each independent variable.

Table 2: Independent Variables Categories and Descriptions

	Independent Variable Category	Variables and Descriptions (The values of all variables range from 1 to 4)
1	Experience	(1) Academic status (Freshman, Sophomore, Junior, or Senior) (2) Level of experience in discussion (3) Level of experience in presentation
2	Learning Attitude	(4) Level of students' interest in subject (5) Level of stress that students feel about learning (6) Level of focus that students have for their coursework in their college life
3	Course Preview	(7) Level of preparatory reading (8) Level of preparatory research (9) Level of preparatory homework
4	Personality	(10) How much students like to speak in public (11) How much students like to study for exams (12) How much students like reading (13) How much students like doing groupwork

B Empirical Results

Figure 1: Percentage of students who felt that they obtained knowledge or the seven meta-skills

Not Enough: How Personality and Behavior Predict Academic Performance, 40 (3) J. OF RSCH. IN PERSONALITY 339 (2006) (studying how personality affects academic performance).

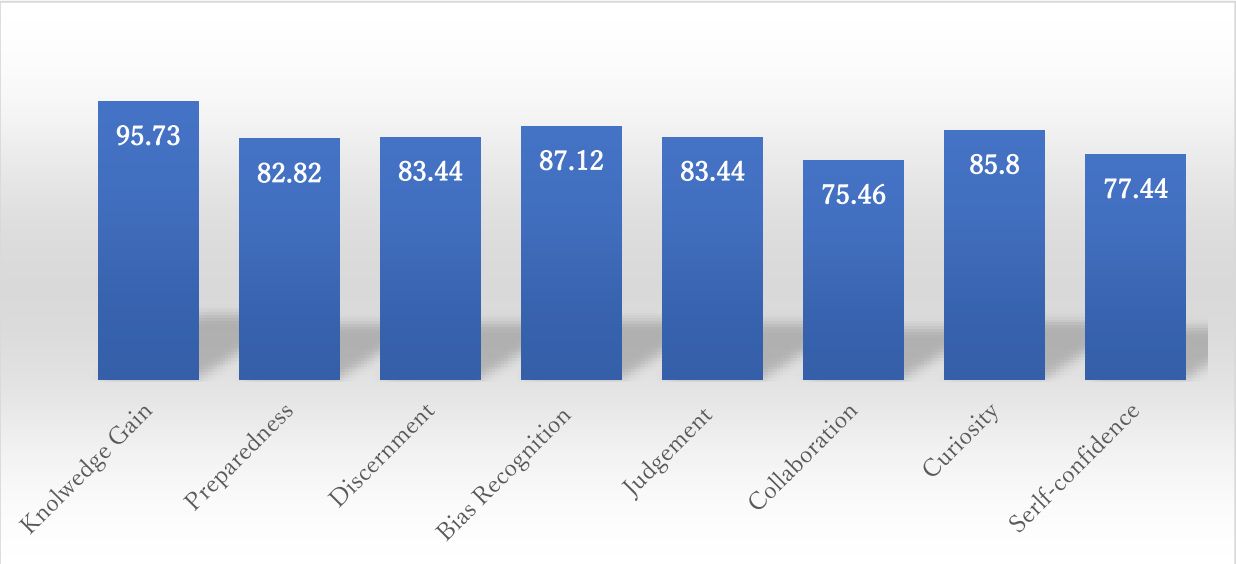

Figure 1 shows total percentages of students who agreed or somewhat agreed with the statement that they acquired each dependent variables—knowledge and seven meta-skills through the Case Method in Civil Law I and Corporate Law I. You can see that over 75% of the students feel that they acquired knowledge and all vital meta-skills; in particular, over 95% of the students feel that they acquired knowledge. This figure, however, does not show which students are more likely to acquire knowledge and meta-skills through the Case Method. We, therefore, move onto the results from running ordered logit models,[69] which examined the effects of various factors associated with students on their acquisition of knowledge and meta-skills.

Table 3 below exhibits the results of running eight different ordered logit models. Each model shows estimated coefficients associated with each independent variable, standard errors in parentheses, and significance levels shown with asterisk symbols.[70]

Table 3: Ordered Logit Model Results

		Model 1 Knowledge	Model 2 Preparedness	Model 3 Discernment	Model 4 Bias Recognition	Model 5 Judgment	Model 6 Collaboration	Model 7 Curiosity	Model 8 Confidence
Experience	Years of study	-0.11 (0.28)	0.21 (0.27)	-0.006 (0.24)	0.49* (0.23)	-0.21 (0.28)	0.16 (0.27)	0.02 (0.26)	0.33 (0.24)
	Discussion	0.21	-0.6*	-0.23	0.06	0.33	-0.32	-0.57*	-0.8**

[69] Ordered logit models are used "when the dependent variables are ordinal rather than continuous" Richard Williams, *Understanding and Interpreting Generalized Ordered Logit Models*, 40 (1) J. MATHEMATICAL SOCIOLOGY 7, 7 (2016).

[70] For the detailed explanation of these statistical terms, see generally Alan O. Sykes, *An Introduction to Regression Analysis*, CHICAGO WORKING PAPER IN L. & ECON. (1993).

		(0.27)	(0.24)	(0.23)	(0.22)	(0.23)	(0.22)	(0.22)	(0.28)
	Presentation	0.005	0.11	0.41	-0.07	0.50	0.03	0.35	0.34
		(0.29)	(0.28)	(0.22)	(0.22)	(0.25)	(0.25)	(0.22)	(0.26)
Learning Attitude	Interest level	0.74*	0.08	0.74**	0.54	0.55	0.47	0.61*	0.34
		(0.32)	(0.31)	(0.26)	(0.28)	(0.31)	(0.26)	(0.25)	(0.26)
	Stress level	0.50	1.02**	0.84**	0.28	0.33	0.27	0.49	0.99**
		(0.30)	(0.25)	(0.34)	(0.27)	(0.25)	(0.22)	(0.25)	(0.28)
	Focus level	1.09**	0.36	-0.04	0.45	0.24	-0.01	0.53	0.27
		(0.38)	(0.35)	(0.34)	(0.39)	(0.41)	(0.36)	(0.35)	(0.38)
Course Preview	Reading	0.28	-0.18	0.21	-0.30	-0.40	-0.43	0.08	-0.16
		(0.29)	(0.31)	(0.36)	(0.32)	(0.30)	(0.33)	(0.26)	(0.32)
	Research	0.06	0.42	0.20	0.26	0.14	0.76**	0.51	0.74**
		(0.27)	(0.25)	(0.20)	(0.25)	(0.24)	(0.22)	(0.28)	(0.26)
	Homework	0.99**	1.14**	0.47	0.61*	0.65*	0.33	0.51	0.58*
		(0.30)	(0.33)	(0.29)	(0.26)	(0.30)	(0.32)	(0.28)	(0.226)
Personality	Talking	0.10	-0.10	-0.13	-0.31	-0.09	-0.13	0.11	-0.25
		(0.32)	(0.28)	(0.30)	(0.27)	(0.27)	(0.26)	(0.25)	(0.23)
	Studying	-0.24	0.26	-0.41	-0.43	-0.13	-0.36	0.20	0.09
		(0.29)	(0.26)	(0.25)	(0.26)	(0.29)	(0.27)	(0.28)	(0.28)
	Reading	0.43	-0.04	0.12	0.53*	0.49*	0.35	0.17	0.68*
		(0.29)	(0.27)	(0.25)	(0.25)	(0.23)	(0.26)	(0.23)	(0.27)
	Groupwork	-0.04	0.24	0.32	0.53	0.16	0.88**	0.29	-0.08
		(0.32)	(0.26)	(0.27)	(0.28)	(0.26)	(0.31)	(0.27)	(0.26)
	Pseudo R-Squared	0.19	0.15	0.12	0.10	0.08	0.13	0.13	0.19
	N	155	154	155	154	154	155	153	155

Robust standard errors in parentheses / **p<.01, * p<.05

C Discussion

Overall, the results shown in Part III.B above illustrate that various factors significantly impact the likelihood of students gaining knowledge and seven vital meta-skills. Here, we discuss material findings in terms of the four categories of independent variables—experience, learning attitude, course preview, and personality.

1 Experience

In Model 4 (Bias Recognition), the estimated coefficient for the variable representing students' academic status is positive and statistically significant at 95 % confidence level. This demonstrates that senior students are more likely to acquire the "bias recognition" skill, which is measured with the ability to listen to others with different points of views.[71]

In addition, we calculated the predicted probabilities to demonstrate the magnitude of the impact that independent variables have on the dependent variables. In our calculations, we focused on comparing only the highest and lowest values. For example, when illustrating the impact of academic year, we only provided the predicted probabilities for seniors and freshmen. This is because the difference between these groups is most pronounced. The predicted probability of senior students self-evaluating the acquisition of the meta-skill is 0.3595, which is over three times higher than the likelihood of freshmen students doing so (0.1138). We argue that, because all courses are taught through the Case Method at NUCB, the more senior a student is, the more the student is exposed to and get used to the learning environment created through the Case Method, which encourages students to be open to different views from classmates.

Next, we turn to the estimated coefficient of the independent variable representing students' experience in discussion. It is statistically significant and negative at 95 % or higher confidence level in Model 2 (Preparedness), 7 (Curiosity), and 8 (Confidence). These results indicate that students with less experience in discussion are more likely to acquire the abilities to prepare and become more curious and confident about learning. The predicted probabilities of students with little experience in discussion self-evaluating the acquisition of the abilities to prepare, become more curious, and confident about learning are 0.4038, 0.4853, and 0.4039 respectively.[72] These probabilities are nearly or way over four times higher than that of students with substantial experience in discussions feeling the same way about gaining these meta-skills (0.1025, 0.1446, and 0.0595 respectively). We argue that this relationship holds because the Case Method constantly makes students engage in discussion. When students enter the world of such an active learning without much experience in discussion, they might feel challenged, but they might also feel rewarded with a greater gain of those abilities in the end.

The estimated coefficient of the explanatory variable reflecting students' experience in presentation is, however, statistically insignificant across all models.[73]

---

[71] *See supra* note 64 (providing Nohria's argument about the bias recognition skill.

[72] As explained above, we calculated the predicted probabilities to demonstrate the magnitude of the impact that independent variables have on the dependent variables and focused on comparing only the highest and lowest values. *See supra* para. 2 of Part III.C.1.

[73] We speculate that experience in presentation may not have much effect on students' learning through the Case Method because communication required in the Case Method is typically a one-to-one dialogue between the teacher and a student or between students, not a presentation in front of the audience.

2 Learning Attitudes

Here, we analyze how the second category of the independent variables, leaning attitudes, influence learning outcomes. First, the estimated coefficient of the level of students' interests in the course subject is positive and statistically significant in Model 1 (Knowledge), 3 (Discernment), and 7 (Curiosity) at 95 % or higher confidence level. This shows that the more interest students have in the course subject, the more likely they acquire knowledge as well as the discernment and curiosity meta-skills. The predicted probabilities of students with the highest interests in course subject self-evaluating the acquisition of knowledge as well as the abilities to identify what is important and become more curious about learning are 0.5590, 0.4855, and 0.3725 respectively. These probabilities are nearly five times higher than those of students with the least interest feeling the same way about gaining these meta-skills (0.1186, 0.0924, and 0.0872 respectively). We argue that students who are interested in the law courses are naturally more focused and eager to grasp what is important during the class discussion, thus acquiring knowledge and becoming more capable of identifying important points. Also, once they acquire basic legal knowledge, they could become interested in learning more complicated legal issues.

In addition, the estimated coefficient of the independent variable representing the level of stress students feel about learning is statistically significant and positive in Model 2 (Preparedness), 3 (Discernment), and 8 (Confidence) at 99 % confidence level. This demonstrates that the less stress students feel about learning, the more likely that students acquire the preparedness, discernment, and confidence meta-skills. The predicted probabilities of students with least stress self-evaluating the acquisition of the abilities to prepare, identify what is important, and become more confident about learning are 0.2801, 0.4156, and 0.2080 respectively. These probabilities are at least twenty times higher than that of students with highest stress feeling the same way about gaining these meta-skills (0.0177, 0.0541, and 0.0132 respectively). We argue that the students who feel less stress about learning may be more open to new ideas and ways to think, thus gaining ability to prepare for difficult situation and find what is important even when they face a new issue. On the other hand, if students feel very stressed about leaning, their participation to the class discussion would become limited and they may have fewer opportunities to gain confidence.

Furthermore, the estimated coefficient of the independent variable reflecting the level of students' focus on coursework is statistically significant and positive in Model 1 (knowledge) at 99% confidence level, meaning that students who are more focused on coursework as part of their college life are likely to be able to attain new legal knowledge through the Case Method. The predicted probability of students with the strong focus self-evaluating the acquisition of knowledge is 0.9698, which is much higher than the probability of students with the least focus feeling the same way about the acquisition of knowledge (0.5488). We argue that, if students spend more time of their college life

in events other than the course, they may be able to acquire personal skills such as a communication skill, but they would lose opportunities to acquire new knowledge in the course.

3 Course Preview

As to the third category of independent variables, course preview, the level of preparedness by simply reading cases does not have statistically significant bearing on students' acquisition of knowledge and all vital meta-skills. We argue this would likely occur because giving a passive assignment, such as just reading a case without further assignments would not make students actually "prepare" for the course. Whether they read the material or not may have little impact on students' learning. We therefore suggest that professors give students more active assignments, such as conducting research to find a solution or write a paper thar requires students to express ideas in their own words as discussed below.[74]

The estimated coefficient of the independent variable measuring the level of students' course preparation by conducting research is statistically significant and positive in Model 6 (Collaboration), and 8 (Confidence) at 99% confidence level. This shows that students who conduct research prior to class are more likely to feel that they gained the abilities to collaborate with others and be confident. The predicted probabilities of students who did enough research prior to class self-evaluating the acquisition of the abilities to collaborate and become more confident are 0.3312 and 0.3512 respectively. These values are at least three times higher when compared to the same predicted probabilities of students who did not do research at all (=0.11901 and 0.0552 respectively). The result could be explained as follows: because by doing research, not just reading a case, students took extra time to gain greater insights about cases, they might have felt that their collaboration with others went well, which ultimately might have made them feel more confident.

The estimated coefficient of the independent variable measuring students' preparedness for course by doing homework is statistically significant and positive in Model 1 (Knowledge), 2 (Preparedness), 4 (Bias Recognition), and 8 (Confidence) at 95% or higher confidence level. This shows that students who do homework are more likely to obtain not only knowledge but also the meta-skills of preparedness, bias recognition, judgment, and confident. The predicted probabilities of students who did homework well self-evaluating the acquisition of the abilities to prepare, listen to others, judge, and become more confident are 0.3425, 0.2827, 0.2454, and 0.1708 respectively. These values are noticeably higher when compared to the same predicted probabilities of students who did not do homework at all (=0.0764, 0.0126, 0.0444, and 0.0342 respectively). Here, homework usually means to write a short paper to answer one or some of the assignment questions in a case, so they need to actively find issues and prepare solutions by themselves. We argue that this process of actively

[74] *See infra* Part IV (recommending teachers give active assignments).

learning "before" the class should give them more opportunity to learn various skills through the Case Method.

4 Personalities

Finally, we move onto the fourth category. The estimated coefficient of the independent variable measuring students' personal preferences in reading is positive and statistically significant in Model 4 (Bias Recognition, 5 (Judgment) and 8 (Confidence) at 95% confidence level. Although we did not find any statistically significant bearing of the level of course preview by reading in relation to learning,[75] the results show that students who like reading are more likely to feel that they obtained the greater abilities to listen to others with different viewpoints, judge, and be confident about learning. The predicted probabilities of students self-evaluating the acquisition of the bias recognition, judgment as well as confidence when students like reading are as follows: 0.4491, 0.4966, and 0.3382. By contrast, the same predicted probabilities when students dislike reading are as follows: 0.1664, 0.1127, and 0.0620. Hence, students who like reading are more than three times likely to feel that they have gained these meta-skills than those who dislike reading.

These results may be relevant to what this personal preference indicates. Students who like reading may not be comfortable with speaking with other people. Hayati, for example, conducts a comparative analysis on how extrovert and introvert students score differently in the different sections of English tests and finds that introverted students are better at reading.[76] As such, more introverted students who like to read may feel that they gained the greater abilities to listen to others, judge, and be confident because the Case Method challenges them to constantly speak with others and make their own argument but simultaneously gives them a sense that their efforts are paid off.

Similarly, the estimated coefficient of the independent variable measuring students' personal preference in doing groupwork is statistically significant and positive in Model 6 (Collaboration) at 99% confidence level. This result shows that students who like to work as a group are more likely to feel that they have obtained the greater ability to collaborate. The predicted probabilities of students self-evaluating the acquisition of the collaboration skill are 0.4384 when students like groupwork and 0.0516 when students dislike the activity. We argue that this holds because students who like groupwork would participate in group discussions more actively through the Case Method, thus feeling more accomplishment as a group.

One the other hand, the level of preference in talking and studying did not have statistically significant bearing on students' acquisition of and all vital meta-skills.[77]

---

[75] *See supra* para. 1, Part III.C.3 (finding the level of preparedness by simply reading cases does not have statistically significant bearing).
[76] *See generally* Hayati, *supra* note 68.
[77] Students who like to talk may be more willing to raise hands in class and those who like studying

D Limitations

We admit that this study is subject to some limitations and there are rooms for improvement to conduct further research. First, the courses subject to this study, Civil Law I and Corporate Law I, are taught by only one teacher. Teachers' philosophies and skills would influence students' learning outcomes in a number of ways.[78] Although the mainstream teaching method for teaching law to undergraduate students is lecture,[79] the Case Method requires a different set of skills to facilitate discussions. Teachers also play an important role in choosing the appropriate cases and are sometimes expected to develop cases for their own courses, but those cases may not align with the intended learning outcomes.[80] Given the understanding of faculty's role in affecting students' learning, further research could use data taken in the same courses taught by different teachers to compare and analyze the effects of teachers' experience, skills and quality of cases.[81]

Second, another important point to consider is to examine acquisition of knowledge and meta-skills through more objective data. This Article utilized the survey in which students self-evaluated acquisition of knowledge and skills. We, however, argue that this type of research is still very insightful for the following reasons. Self-evaluation is an important aspect of active learning.[82] We also found that many previous studies regarding Political Science education and active learning utilized student surveys.[83] Nevertheless, further research, by conducting a comparative analysis on

---

may prepare for and participate in class more actively, but our results did now show any significant estimated coefficient in any models.

[78] *See* Ito & Takeuchi (Instructors), *supra* note 67, at 8-9 (listing teacher's unfamiliarity with and indifference to the Case Method as issues preventing the use of the Case Method).

[79] *See supra* Part II.C (describing Japanese legal system and teaching method).

[80] *See* Ito & Takeuchi (Instructors), *supra* note 67, at 9 (listing lack of adequate case materials as an issue preventing the use of the Case Method).

[81] Rosenbaum et al., *supra* note 11, analyzes data taken in law courses taught by Yamamoto, political science courses taught by Rosenbaum, and data science courses taught by Sasanuma, the authors of this Article. In that article, however, we treated teachers' skills and quality of cases as constants so that we could focus more on student-focused factors in relation to the acquisition of knowledge and meta-skills.

[82] Smith and Cardaciotto argue that since students are expected to learn by "do[ing] meaningful activities and think[ing] about what they are doing" through active learning, "there must be an opportunity for students to reflect" and evaluate what they have done. Veronica C. Smith & Lee A. Cardaciotto, *Is Active Learning Like Broccoli? Student Perceptions of Active Learning in Large Lecture Classes*, 11 (1) J. OF THE SCHOLARSHIP OF TEACHING AND LEARNING 53, 54 (2011).

[83] *See e.g.*, Erik Lundberg, *Can Participation in Mock Elections Boost Civic Competence Among Students?*, 20 (2) J. OF POL. SCIENCE EDUC. 274 (2024); Chika Y. Rosenbaum & Weston Jamison, *Does Political Science Education Improve Electoral Knowledge? An Analysis on U.S. Presidential and Texas Gubernatorial Elections*, 6 (1) J. OF SOCIAL STUDIES AND HISTORY EDUC. 6 (1) (2022); Angela Lumpkin et al. , *Student Perceptions of Active Learning*, 49 (1) COLL. STUDENT J. 121 (2015); Michael Cavanagh, *Students' Experiences of Active Engagement Through Cooperative Learning Activities in Lecture*, 12 (1) ACTIVE LEARNING IN HIGHER EDU. 23 (2011); Juan C. Huerta & Joseph Jozwiak, *Developing Civic Engagement in General Education Political Science*, 4 (1) J. OF POL. SCIENCE EDUC. 42 (2008) (utilizing students' self-evaluation for research). Ito and Takeuchi

students' learning outcomes based on both subjective and objective data about their acquisition of meta-skills and knowledge, should be able to mitigate the concern of using self-evaluation and address the gap between instructors and students' views of the learning process through the Case Method.

Third, even though we asked students to evaluate the effectiveness of the Case Method, the courses subject to the survey are taught not by the "pure" or "traditional" Case Method, but a "modified" Case Method. As explained in Part II.D, Yamamoto uses the Socratic method to facilitate the dialogue between him and a student or between students, but he doesn't cold call, and rather, asks a volunteer to speak up. His teaching methods also include other methods than the Case Method, such as some lecture and group discussion to accommodate undergraduate non-law students. Therefore, the students' self-evaluation may not be solely based on the Case Method, and some may be based on other teaching methods. For example, some of the knowledge that a student believes that the student acquired may come from a short lecture after the case discussion. It would be interesting to see what results we could get if we conduct a control experiment: one class by using only the pure Case Method and another by only lecture, by the same teacher, for the same subject and same students.

IV. Conclusion

While the literature shows the effectiveness of the Case Method generally, this Article adds that the level of effectiveness, in terms of acquiring knowledge and the vital meta-skills, depends on students' experience, learning attitude, course preview, and personalities. Our analysis based on empirical evidence indicates that many of the independent variables—the level of experience in the Case Method and discussion, the level of interest in the course subject, stress about learning, focus on the coursework, the level of course preview by doing research and homework, and their personal preference in reading and groupwork—significantly affect the acquisition of knowledge and various meta-skills. On the other hand, we could not find significant bearing in some of the independent variables—the level of experience in presentation, the level of course preview by reading, and personal preference in talking and studying.

In addition, for each independent variable that showed statistical significance, we calculated the predicted probabilities to demonstrate the magnitude of the impact that these factors have on students' self-evaluated acquisition of knowledge and vital meta-skills. In particular, the predicted probabilities of students with least stress who self-evaluated the acquisition of the abilities to prepare, identify what is important, and become more confident about learning are at least twenty times higher than those of students with highest stress.[84]

---

also show that based on the survey of 400 instructors in Japan, there is no consensus among the instructors as to how students' performance should be evaluated when active learning is used. Ito & Takeuchi (instructors), *supra* note 67.

[84] *See supra* para. 2, Part III.C.2 (showing the predicted probabilities regarding leaning attitudes).

Lecture is still the mainstream teaching method in Japan where the main source of law is statutory law.[85] Our survey, however, demonstrates that the Case Method is effective for undergraduate non-law students to obtain legal knowledge and some vital meta-skills. Given the analysis of estimated coefficients and predicted probabilities of variables, we propose some modifications or improvements in law course design as follows. First, teachers should be careful in choosing appropriate and interesting material, in particular, cases so that students become more interested in the course subject and more focused on coursework.[86] Second, teachers should mitigate stress that students may have about learning by, for example, replacing cold-calling with warn-calling or group discussions.[87] Third, teachers should give students not just a reading material but also active assignments such as making them do their own research to find a solution or write a paper to express what they think.[88] Such a modified Case Method may also be helpful for some, if not all, undergraduate students who plan to go to graduate law schools to prepare for rigorous Socratic teaching. By following these recommendations, undergraduate non-law students will be more likely to acquire legal knowledge and vital meta-skills.

[85] *See supra* Part II.C. (describing Japanese legal system and teaching method).
[86] *See supra* para. 1, Part III.C.2 (arguing that the more interest students have, the more likely they acquire knowledge and certain skills).
[87] *See supra* para. 2, Part III.C.2 (arguing that the less stress students feel, the more likely they acquire certain skills).
[88] *See supra* para. 3, Part III.C.3 (arguing active learning before the class should give students more opportunity to learn).

The Cheapest Cost Avoider Is Dead—Long Live the Best Algorithmic Risk Governor

Boaz Segal*

Abstract

Artificial intelligence is destabilizing one of tort law's most influential organizing principles: the Cheapest Cost Avoider ("CCA"). Classical CCA analysis emerged in relatively bounded accident settings, where the actor best positioned to prevent harm was often also best positioned to evaluate the relevant cost–benefit tradeoffs. Algorithmic systems disrupt this alignment. In AI-driven environments, harms increasingly emerge from layered socio-technical ecosystems involving developers, platform providers, institutional deployers, and opaque model architectures. Under these conditions, the most visible human actor at the point of injury—the physician, driver, or loan officer—often appears to be the CCA, yet lacks meaningful control over the system-level risks that precipitated the harm.

This Article argues that tort law must move beyond the CCA paradigm and reorient liability around a new organizing concept: the Best Algorithmic Risk Governor ("BARG"). Building on Calabresi and Hirschoff's best decision maker ("BDM") framework and the economic logic of the Hand formula, the Article contends that liability should concentrate on the actor best positioned to observe, evaluate, and govern systemic algorithmic risk at scale. The Article develops functional markers for identifying BARGs, including control over training data, model architecture, monitoring infrastructure, and population-level risk observability.

Using medical AI, autonomous vehicles, and algorithmic credit systems as case studies, the Article demonstrates how contemporary tort doctrine systematically misallocates responsibility by focusing on downstream human discretion rather than upstream governance power. It concludes that tort law's efficacy in the algorithmic age lies not in assigning blame for isolated accidents, but in structuring incentives for continuous institutional risk governance.

I. Three Rooms, One Question—Tort Law at the Point of Algorithmic Harm

In a single mid-sized American city, three quiet rooms become sites of algorithmic fate.

In the first room, a radiologist stares at a screen in the city's main hospital. The AI diagnostic system flashes "low malignancy risk" next to a lung nodule the physician's instincts find troubling. A cursor blinks beside a green confidence score; a second monitor shows the hospital's risk-management dashboard, reassuringly clean. In the next room, a patient signs a consent form that mentions "decision-support software" in one buried line of boilerplate, indistinguishable from the noise of other clauses. Months later, the patient dies from a cancer the radiologist would likely have caught without the AI-generated reassurance.

Across town, on the same gray afternoon, a driver "supervises" an almost-autonomous vehicle gliding through familiar streets. The system promises to handle everything "except rare edge cases". The driver, succumbing to the predictable lull of automation bias after hundreds of uneventful miles and a flawless marketing campaign, glances at a text message just as the car's vision system misclassifies a pedestrian. The impact occurs faster than any human could reasonably react, faster than any ordinary negligence narrative can comfortably describe.

A few blocks away, in a glass-front bank branch on the city's main street, a loan officer clicks "approve" or "deny" based on a credit score generated by a proprietary machine-learning model—trained, tuned, and periodically updated by a vendor the bank barely understands. The officer has been instructed not to "second-guess the algorithm," and the bank's compliance manual treats the score as an objective, neutral fact. The applicant denied that afternoon will later default on high-cost informal credit, triggering a cascade of financial and emotional harms that no one in the room believes they "chose."

In each of these "rooms", tort law is about to be asked its oldest question in a radically new setting: *who should pay for these harms—and why?*

For over half a century, the canonical answer has leaned on the figure of the Cheapest Cost Avoider ("CCA"). But when harms emerge from multi-layered, opaque, and constantly updating socio-technical systems, that familiar figure begins to blur. This article argues that tort law must

be re-designed for the age of AI by shifting its focus from locating the CCA to identifying—and legally cultivating—the Best Algorithmic Risk Governor ("BARG").

Artificial intelligence now permeates high-stakes domains: *Medicine*—diagnostic and triage algorithms, radiology tools, robotic surgery, and, increasingly, large language models for clinical decision support;[1] *Mobility*—autonomous and semi-autonomous vehicles, driver assistance systems, and mixed traffic platooning;[2] *Finance and work*—credit scoring, fraud detection, underwriting, algorithmic hiring, and automated performance management;[3] *Generative AI*—large models that produce text, images, and code, with sophisticated but fragile guardrails.[4]

In each of these environments, the AI "decision" is the visible tip of a long chain of design choices, data curation, model training, deployment, updating, and institutional integration. Yet traditional tort reasoning often gravitates toward the last human in the loop – the doctor, the driver, the clerk, or the content poster – as the CCA or primary bearer of negligence-based duties.

The central question of this article is: *On which actor(s) in the AI ecosystem should tort law concentrate liability if it wants to actually minimize the social costs of accidents and promote safe design?*

The article advances three claims:

---

[*] Doctor of Law, Visiting Scholar UC Berkeley School of Law and Vice Dean, School of Law, Sapir Academic College.

[1] David O. Shumway & Hayes J. Hartman, *Medical Malpractice Liability in Large Language Model Artificial Intelligence: Legal Review and Policy Recommendations*, 124 J. OSTEOPATH. MED. 287, 287–288 (2024); Clara Cestonaro et al., *Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review*, 10 FRONT. MED. 1, 1–3 (2023); George Maliha et al., *Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation*, 99 MILBANK Q. 629, 629–632 (2021).

[2] Shi Rui, *Research on Tort Liability of Autonomous Vehicles in Traffic Accidents*, 19 BCP SOC. SCI. & HUMAN. 157, 157–160 (2022); Muhammad Uzair, *Who Is Liable When a Driverless Car Crashes?*, 12 WORLD ELECTR. VEH. J. 1, 1–3 (2021); Xu Chen & Xuan Di, *Legal Framework for Rear-End Crashes in Mixed-Traffic Platooning: A Matrix Game Approach*, 3 FUTURE TRANSP. 417, 417–419 (2023).

[3] Xukang Wang et al., *Algorithmic Discrimination: Examining Its Types and Regulatory Measures with Emphasis on US Legal Practices*, 7 FRONT. ARTIF. INTELL. 1, 2–5 (2024); Lauri Kai, *Machine-Learning Credit Scores and Disparate Impact Theory* 2–8 (2018) (unpublished manuscript) (available at SSRN 3166562); Natalie Sheard, *Employment Discrimination by Algorithm: Can Anyone Be Held Accountable?*, 45 UNSW L.J. 617, 622–630 (2022); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 679–691 (2016).

[4] Justin D. Weisz et al., *Toward General Design Principles for Generative AI Applications*, ARXIV:2301.05578, 2–3, 8–10 (2023); Susan Hao et al., *Safety and Fairness for Content Moderation in Generative Models* 1, 1–3 (2023).

1. *In AI-driven environments, reliance on the classical CCA test alone systematically misallocates responsibility*—Economic logic that once worked tolerably well in simple, bilateral interactions breaks down in multi-layered algorithmic ecosystems.[5]

2. *Courts should pivot toward a "best cost–benefit decision maker" perspective, adapted to algorithmic systems*—This builds on Calabresi and Hirschoff's insight that the core question is who is best positioned to perform and act upon the cost–benefit calculus, not merely who can physically prevent a particular accident.[6]

3. *To operationalize this shift in AI, tort law needs a new organizing concept: the "Best Algorithmic Risk Governor"*—This is the actor best positioned to gather information about algorithmic risks, evaluate cost–benefit trade-offs, and implement system-level changes that affect many users at once.

The payoff is not merely theoretical. Properly used, tort law can function as an architect of algorithmic reality, aligning private incentives toward safer AI design and governance while avoiding the scapegoating of frontline human actors for systemic failures beyond their control.[7]

II. The Best Decision Maker Move—From CCA to BDM to BARG

This chapter develops the economic foundations of the argument by returning to three canonical frameworks in tort theory: Calabresi's CCA, the Calabresi–Hirschoff best decision-maker ("BDM") test, and Judge Learned Hand's negligence formula. It unpacks each of these in turn, showing how they conceptualize the allocation of accident costs, the locus of cost–benefit analysis, and the role of courts in specifying reasonable care. Together, these three theories provide the analytic vocabulary that the rest of the article adapts and extends to AI-driven, algorithmic environments.

---

[5] Roberto Pardolesi & Bruno Tassone, *Guido Calabresi on Torts: Italian Courts and the Cheapest Cost Avoider*, 1 ERASMUS L. REV. 7, 17–18, 34–35 (2008); See generally John C.P. Goldberg, *History, Theory, and Tort: Four Theses*, 11 J. Tort L. 17 (2018); Miriam Buiten, Alexandre de Streel & Martin Peitz, *The Law and Economics of AI Liability*, 48 COMPUT. L. & SEC. REV. 1, 8–12 (2023).
[6] John C.P. Goldberg, *Id*; Helmut Koziol ed., BASIC QUESTIONS OF TORT LAW FROM A COMPARATIVE PERSPECTIVE 453–456 (Jan Sramek Verlag 2015).
[7] George Maliha et al, *supra* note 1, at 634–639; Madalina Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account*, 81 PUB. ADMIN. REV. 825, 832–834 (2021).

A. Move One: Avoidance—The CCA Logic

Calabresi's project in The Costs of Accidents is not only to define the CCA, but to embed that figure in a broader optimization of primary, secondary, and tertiary accident costs in a world of positive transaction costs. Primary costs are the expected losses from accidents plus the resources invested in avoiding them; secondary costs are the costs of spreading or insuring against those losses; tertiary costs are the administrative expenses of operating the liability system itself. The cheapest cost avoider is therefore the actor who, taking all three cost categories into account, can most efficiently combine precautions, activity-level adjustments, and risk-spreading.[8]

This is why Calabresi famously associates the CCA idea with targeted strict liability rather than with negligence: concentrating liability on the CCA is supposed to induce optimal investments in safety while also exploiting that actor's superior capacity to buy insurance and pass residual accident costs along in prices. Subsequent law and economics work has elaborated and refined this insight, for example by asking when loss-sharing or mixed liability rules outperform all-or-nothing allocation, and by testing whether real courts actually succeed in identifying CCAs in complex multi-party accidents. These developments reinforce the core lesson for AI: if courts keep focusing liability on the last human actor, they risk ignoring those institutions that are in fact the cheapest cost avoiders at the systemic level.[9]

B. Move Two: Decision—From Prevention to BDM

In a later article on strict liability, Calabresi and Hirschoff proposed a refinement: liability should be assigned to the party "in the best position to make the cost–benefit analysis between accident costs and accident-avoidance costs and to act on that decision." This formulation shifts the focus From "who can physically prevent the accident?" To "who should be making, and acting on, the optimization decision?"

---

[8] G. Calabresi THE COSTS OF ACCIDENTS – A LEGAL AND ECONOMIC PERSPECTIVE, 26–31, 135–173 (1970); Roberto Pardolesi & Bruno Tassone, *supra* note 5, at 11–12; Guido Calabresi, *The Decision for Accidents: An Approach to Nonfault Allocation of Costs*, 78 Harv. L. Rev. 713, 714–715 (1965).

[9] Calabresi *Id*, 719–730; Emanuela Carbonara, Alice Guerra & Francesco Parisi, *Sharing Residual Liability: The Cheapest Cost Avoider Revisited*, 45 J. LEGAL STUD. 173, 178–180, 191–194 (2016).

Subsequent scholarship has highlighted this as the move from a pure CCA perspective to a "best positioned decision-maker" test, especially relevant when prevention requires information-rich, technical, or systemic judgments rather than simple, observable acts of care.[10]

The Calabresi–Hirschoff refinement pushes the analysis from physical control over a specific accident to informational control over a class of accidents. Their BDM test asks which actor is best placed to gather information about risks and precautions, to perform the cost–benefit comparison, and to translate that comparison into changes in activity levels, technologies, and contractual arrangements. In many simple, bilateral accidents the CCA and the BDM will coincide, but Calabresi and Hirschoff already anticipated settings where design choices, information flows, and institutional structures are so complex that the party who could have grabbed the last clear chance is not the one who should be internalizing the full optimization problem.[11]

Later scholars have used this perspective to support selective strict liability or hybrid regimes in which responsibility is shifted "upstream" to manufacturers, platform operators, or other institutional actors that can redesign processes and technologies for many users at once. This literature also highlights a tension that is central for AI: sometimes the actor best placed to identify the efficient combination of precautions is not the actor best placed to implement them, which forces courts to choose between deterring bad decisions and inducing concrete behavioral change. The *Best Algorithmic Risk Governor* framework can be presented as an explicit attempt to operationalize the BDM idea for algorithmic ecosystems characterized by opacity, continuous updating, and multi-layered value chains.[12]

---

[10] G. Calabresi & J. T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE LAW JOURNAL 1055, 1060–1064 (1972); For further analysis, see Helmut Koziol ed., BASIC QUESTIONS OF TORT LAW FROM A COMPARATIVE PERSPECTIVE 438–456 (Jan Sramek Verlag 2015); See also Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 11–12.

[11] Yuval Sinai & Benjamin Shmueli, *Calabresi's and Maimonides's Tort Law Theories – A Comparative Analysis and a Preliminary Sketch of a Modern Model of Differential Pluralistic Tort Liability Based on the Two Theories*, 26 YALE J.L. & HUMAN. 101 (2013). This article argues that by uncovering utilitarian-economic foundations in Maimonides's tort theory and placing them in dialogue with Calabresi's cheapest cost avoider and best decision maker doctrines (and with Posner's negligence theory), one can develop a modern, differential and pluralistic model of tort liability that integrates efficiency and justice by allocating strict or fault-based liability according to the type of risk-creating activity.

[12] For a critique of the theory, see Megan L. Richardson, *Revisiting Strict Product Liability: Taking Law and Economics Further*, 35 OSGOODE HALL L.J. 195 (1997). This essay argues that Dewees, Duff and Trebilcock's empirical and economic critique of strict product liability does not preclude a strong case for such liability, because a refined version of Calabresi and Hirschoff's cheaper cost avoider test – one that incorporates information costs, loss-

C. Move Three: Calculation—Hand, Posner, and the Economics of Negligence

Learned Hand's formula in *Carroll Towing*[13]—negligence where $B < PL$—has become the canonical translation of negligence into marginal cost–benefit terms. Law-and-economics scholars, above all Posner[14], read the formula as an implicit efficiency test: a defendant is negligent when an additional unit of precaution would have reduced the expected social loss by more than its cost. On this view, negligence liability induces actors to choose an efficient level of care at which marginal prevention costs equal marginal reductions in expected accident losses. Crucially, however, because a non-negligent actor escapes liability, a negligence rule fails to optimally regulate activity levels—a distortion that, as economic theory notes, is typically cured only under strict liability, where residual losses are internalized and reflected in market prices.[15]

At the same time, a rich critical literature emphasizes that courts rarely apply a literal Hand calculus and that real-world fact-finders often treat B, P, and L in moral or distributive terms rather than as purely economic inputs. This suggests that the Hand formula is best understood as a family of standards rather than a precise algorithm: it authorizes judges and juries to reason in terms of avoidable risk and reasonable precautions, but it leaves open which risks count and how to weigh catastrophic low-probability harms, dignitary interests, or structural inequalities. For AI, this opens space to argue that Hand-style negligence analysis should be anchored at the level of BARGs, not frontline users, because the most meaningful choices about B, P, and L are made when models are designed, trained, and integrated—ong before a physician runs a scan or a driver presses "engage".[16]

D. Move Four: Decoupling—Why AI Splits CCA from BDM (and Makes Room for BARG)

spreading and corrective justice – may yield a more certain and efficient liability rule than the negligence regime they favour.

[13] *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947).

[14] R. A. POSNER, ECONOMIC ANALYSIS OF LAW (5th ed. 1998); W. M. LANDES & R. A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW (1987).

[15] Eugênio Battesini, *Incremental Learned Hand Standard, Degrees of Negligence and Allocation of Damages: A Comparative Tort Law and Economics Approach*, 8 RJLB 1249, 1250–1251, 1258–1260 (2022); Yotam Kaplan & Maytal Gilboa, *The Other Hand Formula: Explaining Gain-Based Liability* 1, 9–12 (2021); Richard W. Wright, *Hand, Posner, and the Myth of the "Hand Formula"*, 4 THEORETICAL INQUIRIES IN L. 145, 146–150 (2003).

[16] Christopher Brett Jaeger, *The Hand Formula's Unequal Inputs*, 135 YALE L.J. 461, 465–481 (2025); See also generally Jeonghyun Kim, *Revisiting the Learned Hand Formula and Economic Analysis of Negligence*, 169 J. INST. & THEORETICAL ECON. 407 (2013).

In many classic accident settings—two drivers, a manufacturer and a consumer, a landowner and a visitor—the CCA and the "best cost–benefit decision-maker" often coincide. The party who can cheaply prevent the accident (by driving carefully, installing guards, or designing a safer product) is usually also the party well placed to evaluate the costs and benefits of precautions. In these relatively simple, bilateral interactions, the same actor typically controls both the relevant behavior and the relevant information: the driver knows how much effort safe driving requires, the manufacturer knows the marginal cost of redesigning a product, and the landowner can estimate the expense of making premises safer. As a result, assigning liability to the CCA effectively channels incentives to the actor who is also best situated to perform the cost–benefit analysis that tort law implicitly demands.

AI decision-making environments break this convenient alignment. Algorithmic systems are developed, trained, deployed, updated, and monitored by different actors who operate at different points in time and at different levels of abstraction. The human end-user—whether a physician relying on a diagnostic tool, a driver "supervising" an automated vehicle, or a caseworker using a risk-scoring system – often appears, at the level of a single incident, to be the CCA: she can double-check the output, apply common sense, or refuse to follow an AI recommendation. Yet she is not the actor best placed to govern systemic algorithmic risk across cases. She does not design the model architecture, set the decision thresholds, choose the training data, determine the feedback loop, or decide how errors are distributed across groups and contexts.

Conversely, the entities that are in fact best positioned to engage in meaningful cost–benefit analysis of algorithmic precautions—AI developers, platform operators, large institutional deployers, and sometimes regulators – frequently do not look like the CCA in any particular accident. Their contribution to a specific harm is distributed, probabilistic, and mediated through code, updates, and data pipelines. They may never interact with the injured party, never see the specific decision, and never appear in the immediate "who-could-have-prevented-this?" narrative that traditional tort analysis often privileges. Nonetheless, these upstream actors are precisely the ones who can observe patterns across thousands or millions of decisions, compare alternative designs, and internalize the long-run costs of false positives, false negatives, and biased error distributions.

The preliminary claim, then, is that while in simple, non-algorithmic accidents the CCA and the best cost–benefit decision-maker frequently converge, AI pushes them apart. Focusing exclusively on the apparent CCA at the point of harm risks misallocating responsibility to local human operators and under-incentivizing those who actually govern algorithmic risk at scale. As the next sections show, the actor who looks like the cheapest cost avoider in a single incident is often not the actor who is best placed to govern systemic algorithmic risk across cases.[17]

III. The Alignment Breaks—Why AI Rewires Tort Law's Liability Map

A. Rewiring Point One: The Actor Stack—A Multi Layered Value Chain

In AI, the "actor" is rarely a single person or firm. What looks like one decision at the point of harm is usually the end product of an actor stack—a multi layered value chain in which distinct entities make distinct risk shaping choices at distinct moments. Tort law, by contrast, is built to tell a cleaner story: identify the relevant actor, specify the duty, evaluate breach against a standard of care, and connect the conduct to the injury. The actor stack disrupts that narrative. Control is layered. Knowledge is layered. So is the capacity to prevent harm. And once those layers are separated, the familiar question—"who should have done more?"—cannot be answered by looking only at the last human in the room.

Start with model developers and vendors. They do not merely "build" a system; they write the system's risk profile into its DNA. They choose architectures, curate training data, tune parameters, and hard-wire defaults—confidence thresholds, escalation rules, retraining cadence—that silently govern downstream behavior. In effect, they preselect which errors will be common, which will be rare, and which will be tolerated. Long before any end-user encounters a concrete case, the developer has already defined the boundaries of what the system will treat as normal, exceptional, and ignorable.

Then comes the platform and cloud layer. Providers offer the infrastructure and tooling that make deployment feasible, and increasingly they supply pre-trained foundation models or large generative systems that downstream actors adopt as building blocks rather than bespoke

[17] Katherine Drabiak, *Leveraging Law and Ethics to Promote Safe and Reliable AI/ML in Healthcare*, 2 FRONT. NUCL. MED. 1, 4–7, 10–14 (2022); Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 5–11; George Maliha et al., *supra* note 1, at 630–639.

products.[18] This layer is often treated as background, but it is not neutral. Platforms can determine what is measurable (logging and monitoring), what is auditable (access to model behavior and change histories), and what can be constrained (rate limits, filters, safety tooling). Those design choices shape not only safety outcomes, but the practical possibility of proving negligence and correcting it. A world in which the relevant evidence is systematically unavailable—or structurally expensive to obtain—is a world in which the standard tort inquiry is tilted before it begins.

Institutional users—hospitals, banks, insurers, ride sharing and logistics platforms—sit at the integration layer, where "operations" becomes governance.[19] They decide which system to procure, where to place it in the workflow, how strongly to frame its outputs, and how much discretion humans will realistically have. A simple configuration choice can rewrite the risk map: a hospital may lower an alert threshold to reduce "noise" route low risk outputs into a slow queue or compress review time under productivity pressures. The model may not change at all, yet the institution's integration can convert a recommendation into an action-forcing directive—or, just as consequentially, an inaction-forcing default. This is the layer where organizational incentives, staffing levels, and compliance pressures translate algorithmic output into lived harm.

Finally, the frontline user appears—the physician, the driver, the loan officer, the HR staffer, the consumer—closest in time and space to the injury. That proximity makes the end-user the easiest defendant to name and the easiest story to tell. But the actor stack makes that convenience misleading. The end-user's "choice" is frequently constrained by upstream interface design, institutional protocol, and the thinness of information available to evaluate reliability. Tort law's impulse to locate responsibility at the last human touchpoint risks confusing visibility with control.

Responsibility is therefore distributed across multiple private and public actors whose decisions are separated in time, space, and expertise.[20] The point is not merely descriptive ("there are many

---

[18] Chen Chen et al., *Trustworthy, Responsible, and Safe AI: A Comprehensive Architectural Framework for AI Safety with Challenges and Mitigations*, ARXIV:2408.12935 1, 2–3, 10 (2025); Philipp Hacker et al., *Regulating ChatGPT and other Large Generative AI Models*, in PROC. ACM CONF. FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 1112, 1112–1116 (2023).

[19] Muhammad Uzair, *supra* note 2, at 12–13 ; George Maliha et al., *supra* note 1, 630–637; Katherine Drabiak, *supra* note 17, at 10–14.

[20] Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *Attributing Responsibility in AI-Induced Incidents: A Computational Reflective Equilibrium Framework for Accountability*, ARXIV , 1–4 (2025); Gabriel Lima & Meeyoung Cha, *Responsible AI and Its Stakeholders*, ARXIV:2004.11434 1, 2–3 (2020); Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, 5–8.

actors"). It is normative and doctrinal: when control and knowledge are layered, tort law must rethink how it assigns duty, assesses breach, and understands causation across a value chain. The actor stack is the first structural reason why AI rewires the liability map—and it sets up the radical information asymmetries that follow.

B. Rewiring Point Two: The Black Box—Radical Information Asymmetries

AI architectures, training pipelines, and performance characteristics are often opaque even to sophisticated institutional users, let alone to frontline professionals and lay users.[21] This opacity is not a mere inconvenience; it is a structural feature of contemporary algorithmic systems that reshapes how (and whether) tort law can do its ordinary work. In classic negligence settings, the law assumes that key actors can *appreciate* the relevant risk, *select* among precautions, and *explain* their choices in a way that courts can evaluate against a standard of reasonable care. The black box disrupts each of those assumptions at once. It disconnects the point of harm from the point of meaningful knowledge, and it converts what looks like a simple "error" into an evidentiary and governance problem: who actually knows enough to calibrate safety, and who can prove what, when something goes wrong?

Developers and platforms typically control documentation, logs, test suites, performance metrics, and outcome data at scale.[22] That control matters because the most legally salient facts in an AI-driven accident—error rates in the relevant subpopulation, known failure modes, model drift, the conditions under which performance degrades, the distribution of false negatives versus false positives, the existence (or absence) of post-deployment monitoring—are rarely visible from the outside. In practice, the information needed to perform a Hand-style "B versus PL" analysis is often *locked upstream*, and it remains locked precisely when litigation begins to ask for it. The result is a predictable distortion: courts can easily scrutinize what the last human actor did in the room, but they may never see the upstream design and governance choices that set the baseline risk in the first place. When the relevant safety knowledge is treated as proprietary (or is simply

[21] Clara Cestonaro et al., *supra* note 1, at 3–4, 9; Adriano Koshiyama et al., *Towards Algorithm Auditing: Managing Legal*, *Ethical and Technological Risks of AI, ML and Associated Algorithms*, 11 R. SOC. OPEN SCI. 1, 4–6, 13–14 (2024); George Maliha et al, *supra* note 1, at 638.
[22] Adriano Koshiyama et al., *Id*, at 3–6.

not recorded in a usable way), the tort system's ordinary fact-finding process becomes structurally tilted toward the visible and away from the causally important.

End-users operate interfaces that may provide limited explanations, coarse risk scores, or blunt recommendations, with minimal ability to interrogate or recalibrate the underlying model.[23] This means that the "human in the loop" is frequently asked—by designers, institutions, and later by plaintiffs – to supply the very thing the interface withholds: a reasoned, context-sensitive assessment of reliability. Yet the user's practical capacity to do so is often thin. A clinician sees a confidence score without the model's calibration data; a driver is told to monitor a system whose failures are rare but catastrophic; a loan officer is given a binary output without access to the model's feature space, training distribution, or threshold logic. The end-user's discretion is therefore not just institutionally constrained (as the actor-stack analysis explains), but informationally hollowed out. What looks like an opportunity to "double-check" is often a demand to validate a system that is intentionally non-interrogable at the point of use.

These asymmetries create what has been described as "black box" liability and responsibility gaps in medical AI and other fields, where no single frontline actor can meaningfully understand, let alone optimize, system-wide risk.[24] The gap is doctrinal as well as practical. Duty and breach become harder to specify when the content of "reasonable care" depends on facts controlled by another actor; causation becomes harder to prove when the mechanism of error is inaccessible; and failure to warn analysis becomes unstable when the most important warnings would require disclosures that vendors resist or institutions never demand. The black box thus amplifies the very divergence this article emphasizes: the apparent CCA at the point of harm may be asked to carry legal responsibility, even though the BDM—and, more precisely, the BARG—is located upstream where the system's risks can actually be observed, quantified, and reduced across cases. In short, radical information asymmetries do not merely complicate tort adjudication; they help explain why liability rules that fixate on the last human decision systematically misallocate incentives in AI ecosystems.

[23] Clara Cestonaro et al., *supra* note 1, at 2–4; Katherine Drabiak, *supra* note 17, at 3–5.
[24] Benjamin H. Lang et al., *Responsibility Gaps and Black Box Healthcare AI: Shared Responsibilization as a Solution*, 2 DIGIT. SOC. 52, 55–56, 59–61 (2023); George Maliha et al, *supra* note 1, 637–639; Clara Cestonaro et al., *Id*, 3–4, 9–10.

Crucially, this is not a static problem. Once the informational baseline is skewed—once the model's behavior, limitations, and change history are not meaningfully transparent to those who rely upon it—the legal system is primed to misread algorithmic harm as a sequence of isolated "bad calls," rather than as a governance failure in a socio-technical system. That is the bridge to the next rewiring point: the moving target. When systems update, retrain, and drift, the black box does not stay in one shape long enough for ordinary liability narratives to stabilize—making information asymmetry not only radical, but continuously renewed.

C. Rewiring Point Three: The Moving Target—Dynamic Systems and Feedback Loops

Unlike static products, many AI systems are not "finished" when they are shipped, installed, or first deployed. Their risk profile is not a stable attribute that can be assessed once and then treated as fixed. Instead, the system's behavior is often a moving target—because the model changes, the environment changes, the data changes, and, crucially, the system itself changes the data environment in which it operates. That dynamism matters for tort law because negligence doctrine is structured around relatively stable objects of evaluation: a design, a warning, a practice, or a professional decision. In AI ecosystems, those anchors drift.

1. Dynamic Adaptation and Environmental Shifts

Models are regularly updated or retrained, sometimes continuously via online learning. A central feature of modern AI deployment is that performance is expected to be maintained through iterative updating. Models are patched, thresholds are recalibrated, retraining data is refreshed, and "improvements" are pushed in cycles that resemble software development more than classic product manufacture.

Even where a model is not literally learning online, it is often subject to "concept drift" and "distribution shift"—the world that generated the training data is not the world in which the model now operates. A diagnostic model trained on one hospital's imaging pipeline may degrade when scanners, protocols, patient demographics, or prevalence rates change; a driving model may fail when weather, signage, road geometry, or fleet composition shifts; a credit model may become brittle as macroeconomic conditions change or consumer behavior adapts. The moving-target problem is therefore not simply that models get updated—it is that "reasonable care" cannot be

evaluated by looking only at a single frozen snapshot of model performance at time t_0, when the harm occurs at time t_1 in a materially different risk environment.

2. Feedback Loops and Self-Creating Systems

AI systems are often subject to feedback loops, where the system's own outputs influence future data (e.g., predictive policing, credit approvals, and hiring pipelines).[25] These feedback loops intensify the moving-target problem because they make the system partially self-creating. When an algorithm's outputs shape what gets observed, recorded, and later treated as "ground truth," the system can lock in its own assumptions—sometimes amplifying error, sometimes amplifying inequality, and often doing both while appearing to improve by internal metrics.

If police deployment decisions concentrate surveillance in particular neighborhoods, the resulting arrest data can "confirm" the model's premise that those neighborhoods are higher risk. If a lender denies applicants predicted to default, the model may never learn whether those applicants would in fact have repaid—creating selective labels and an evidentiary gap that looks like predictive success but is partly a byproduct of the institution's own denial policy. If an employer's screening model narrows the applicant pool, the workforce data that later "validates" the model is a downstream artifact of the model's own gatekeeping. In these settings, the system does not merely predict the world; it helps produce the world it then claims to measure. Tort law's ordinary intuition—evaluate the defendant's conduct against an external reality of risks—becomes harder to apply when the defendant's system is actively reshaping that reality.

3. The Problem of Scale

AI is typically deployed at scale, so that small design choices (e.g., risk thresholds, loss functions) propagate into thousands or millions of decisions. Scale transforms marginal technical parameters into population-level governance decisions. A minor shift in a confidence threshold, an alert suppression rule, a loss function that privileges one kind of error over another, or a UX choice that frames the output as a "recommendation" versus an "instruction" can translate into system-wide changes in false negatives and false positives.

---

[25] Aurora S. Zhang & Anette E. Hosoi, *Structural Interventions and the Dynamics of Inequality*, in PROCEEDINGS OF THE ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1014, 1015–1016 (2024); Madalina Busuioc, *supra* note 7, at 826–827; Solon Barocas & Andrew D. Selbst, *supra* note 3, 682–687.

In classic negligence narratives, the accident is the salient unit: a discrete event calling for a post hoc evaluation of whether the defendant should have been more careful in that moment. In AI systems, by contrast, the salient unit is often the policy encoded in the system—a policy that is executed repeatedly, at speed, and in ways that no single frontline actor can meaningfully audit decision-by-decision. Once the system is scaled, the tort system's preference for individualized stories can systematically miss the more important question: who set the parameters that governed risk across the entire population of decisions?

4. Structural Patterns of Harm

Harms may emerge not as isolated "accidents" but as repeated, structurally generated patterns—for example, systematic under-diagnosis in underrepresented subpopulations, or recurring disparate impact in algorithmic lending and hiring.[26] This is the core doctrinal pressure point. Traditional tort analysis is comfortable with rare malfunctions, one-off mistakes, and localized departures from reasonable care. But dynamic AI systems can generate harms through stable patterns that are produced by an evolving system operating under stable institutional incentives.

In medicine, the harm may not look like "the system broke" but rather "the system is consistently less sensitive for certain patients," especially where training data, calibration choices, or workflow integration make errors more likely for underrepresented groups. In employment and credit, the harm may be a recurring disparate impact that is statistically predictable yet individually deniable—each single denial can be defended as "reasonable reliance" on a score, while the system as a whole operates as a structural sorting mechanism that predictably burdens the same communities. The moving-target character of AI makes this worse: even when an issue is detected, updates can change the system's behavior before courts, regulators, or injured parties can stabilize the factual record.

5. Doctrinal Implications for Tort Law

For tort law, the moving-target and feedback-loop dynamics do not merely "complicate proof." They disrupt how duty, breach, and causation are conceptually organized. A regime built around

[26] Solon Barocas & Andrew D. Selbst, *Id*, 684–687; Savina D. Kim, Stefan Lessmann, Galina Andreeva & Michael Rovatsos, *Fair Models in Credit: Intersectional Discrimination and the Amplification of Inequity*, ARXIV:2308.02680V1, 12–16 (2023); Clara Cestonaro et al., *supra* note 1, at 3, 9.

one-time design choices and static warnings is poorly matched to systems whose safety depends on continuous monitoring, version control, incident response, and retraining governance. When the relevant "precaution" is not a discrete act by a frontline user but an upstream practice—ongoing auditing, drift detection, rollback capability, human-override realism testing, and post-deployment evaluation across subpopulations—the identity of the legally relevant decision-maker shifts.

This is precisely where the divergence between the apparent CCA (Cheapest Cost Avoider) and the BDM (Best Decision Maker) becomes practically consequential: the actor who can most cheaply prevent this instance of harm (often the last human in the loop) is frequently not the actor who can most cheaply and effectively govern the system's evolving risk over time.

In that sense, the moving target is not just a technical feature of modern AI. It is a structural reason why tort law's liability map must be rewired around the actors who can manage dynamics: the entities who control updating, monitoring, deployment constraints, feedback-loop mitigation, and institutional integration. Those are quintessential BARG functions. And once we see AI risk as dynamic and self-reinforcing, the next step follows naturally: tort law must learn to treat algorithmic harm less as a collection of isolated accidents and more as a set of systemically produced patterns—precisely the shift taken up in the next rewiring point.

D. Rewiring Point Four: From Accidents to Patterns—Systemic vs. Individual Risk

Traditional tort paradigms often imagine a discrete accident between an injurer and a victim: a bounded event, a localized lapse, and a post hoc inquiry into whether a specific actor failed to take a reasonable precaution in that moment. Algorithmic systems, by contrast, tend to produce risk as a feature of the system's repeated operation—across many users, contexts, and time periods. In this setting, the legally salient unit is often not "the accident", but the policy encoded in model design, thresholds, workflows, and update practices—policies that can quietly shift the distribution of harms at scale. AI systems therefore generate systemic risk:

1. Structural harms, such as discrimination, exclusion, and loss of privacy or autonomy, often emerge not from a single deviant output but from stable patterns of design and data use—choices about objectives, feature selection, labeling practices, and representativeness that

predictably burden certain groups or contexts even when each individual decision is facially "routine".[27]

2. Causation is often diffused across the socio-technical chain: no single decision looks clearly wrongful in isolation, and no single human actor can be said to have "caused" the harm in the classic but-for sense, yet the aggregate pattern is both foreseeable and socially costly. This is the familiar paradox of pattern-based injustice: each instance is deniable, while the distribution is systematic.[28]

In such environments, effective governance turns on the capacity to detect and respond to patterns—through population-level performance monitoring, auditing for disparate impact, stress-testing and drift detection, and the ability to revise thresholds, interfaces, and deployment constraints. The actors who can see and alter patterns across many cases—those with access to comprehensive data, auditing tools, and meaningful deployment levers—are therefore central to risk reduction in a way that frontline actors rarely can be.[29]

This shift from accidents to patterns pressures tort doctrine on multiple fronts. Standards of care calibrated for one-off mishaps may miss harms that are "reasonable" in any single case but unreasonable in their cumulative incidence or distribution. Similarly, evidence and causation doctrines that privilege individualized narratives may underweight statistical proof of recurring failure modes, feedback-loop amplification, or systematically skewed error rates.

Once risk is understood as systemic, liability design must follow the points of systemic control: where risk is measured, where it can be recalibrated, and where changes propagate across users. This reframes the tort question from assigning blame for an isolated incident to assigning responsibility for governing an evolving risk-generating system.

E. Rewiring Point Five: The Liability Map – Doctrinal Implications for Tort Law

Point Four reframed algorithmic harms as patterned and systemic rather than isolated mishaps. Point Five translates that shift into doctrine: once the relevant "risk" is a moving, feedback-driven

[27] Madalina Busuioc, *supra* note 7, at 826–827; Xukang Wang et al., *supra* note 3, at 3–4; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 677–692.
[28] Aurora S. Zhang & Anette E. Hosoi, *supra* note 25, at 1015–1017; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 1–3.
[29] Adriano Koshiyama et al., *supra* note 21, at 4–6; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *Id*, at 2, 7–9; Madalina Busuioc, *supra* note 7, at 826–827.

system, the liability inquiry must map onto the actors and institutions with genuine governance leverage over that system—design choices, deployment constraints, monitoring, and iterative updates.

Doctrinally, this pushes tort law away from a purely event-based perspective (a single breach causing a single injury) and toward a governance-based perspective. Courts still rely on familiar tools—duty, breach, causation, and reasonableness—but those tools must be applied with attention to upstream design and training decisions, ongoing post-deployment control, and the allocation of informational and organizational capacity to detect risk patterns before they crystallize into harm.

In practical terms, the "liability map" is a mapping between legal responsibility and institutional control. It asks where the system's risk thresholds are set, who decides when to retrain or rollback, who can meaningfully audit outcomes, and who can internalize the costs of harm through pricing, insurance, and enterprise-level precautions. Where those levers sit upstream, doctrines that default to blaming the most visible downstream actor will misallocate incentives.

That is why an exclusively backward-looking question—who could have prevented this accident tomorrow morning?—is radically incomplete. It invites courts to focus on the last human in the chain, even when that person lacks the authority, information, or time to override a system that is calibrated elsewhere.

Put differently, focusing only on the immediate encounter risks treating downstream discretion as control. In AI settings, the key preventive acts often occur earlier—during system design, calibration, documentation, deployment constraints, and post-deployment monitoring—where risk patterns can be detected and corrected.

Tort law must also ask: ***Who is best positioned to govern systemic algorithmic risk—to monitor, understand, and redesign the system in response to emerging harms?***

Rewiring the liability map therefore has concrete doctrinal consequences. First, it reframes duty and breach around governance capacity: the relevant standard of care may include reasonable auditing, robust logging, meaningful human-override pathways, and timely updating or retraining when performance drift is foreseeable. Second, it changes how courts should think about causation. Where harm arises from aggregate system behavior, the evidentiary focus often shifts from pinpointing a single "but-for" decision to establishing that the system's configuration and

oversight regime made the harm materially more likely—and that a feasible governance intervention would have reduced that risk.

Third, it informs the choice among negligence, product liability, and enterprise-liability-style doctrines. Depending on the context, traditional negligence may under-incentivize prevention because the party with the best information is not the party facing the lawsuit. Doctrines that better track informational asymmetry and cost internalization—whether through design-defect analysis, failure-to-warn framed as failure-to-disclose model limits, or institutional liability for deployment policies—can help align legal incentives with the locus of control.

Finally, it clarifies how tort law should interact with public regulation. Compliance may be relevant evidence of reasonableness, but it cannot substitute for governance when regulation lags, is incomplete, or is gamed through box-checking. Precisely because algorithmic systems are moving targets, tort doctrine remains a complementary mechanism for enforcing ongoing risk governance rather than one-time certification.

Against that backdrop, the next section offers concrete examples showing how a surface-level "cheapest cost avoider" (CCA) intuition often points courts toward the wrong defendant, while the true best cost–benefit decision-maker (BDM)—and the actor who can function as an effective BARG—sits upstream.

This is where the divergence between the apparent CCA and the true best cost–benefit decision-maker (BDM) in AI contexts becomes stark—and why the identity of the BARG is typically determined by system-level governance leverage rather than by proximity to the injurer.

IV. The Frontline Fallacy: When CCA Intuition Misfires

A. Medical AI: The "Independent Judgment" Myth

In contemporary medico-legal discourse, clinicians are routinely positioned as the primary bearers of responsibility when AI-assisted diagnosis or treatment contributes to patient harm. Courts, regulators, and professional guidelines alike tend to emphasize the physician's duty to "exercise independent judgment," framing algorithmic tools as mere aids that must never displace human

decision-making.[30] On this view, the physician remains fully accountable for the ultimate clinical decision, even when that decision is heavily shaped by algorithmic output.

From a superficial CCA perspective, this allocation of responsibility appears intuitively plausible. The physician is physically present at the point of care, occupies a well-defined professional role, and fits comfortably within existing malpractice narratives. In theory, the clinician can disregard an algorithmic recommendation, order additional tests, consult a colleague, or rely on clinical intuition. Courts can easily reconstruct this moment of choice, and negligence doctrine is well equipped to ask whether a reasonable physician should have done more in that encounter.

Yet this intuition collapses once the analysis shifts from the single clinical encounter to the structure of algorithmic decision-making in medicine. The most consequential cost—benefit trade-offs that shape diagnostic risk are rarely made at the bedside. Instead, they are embedded upstream in the design, calibration, and deployment of medical AI systems. Choices about sensitivity versus specificity, acceptable false-negative rates, confidence thresholds, and escalation protocols are typically determined long before a physician encounters a particular patient.[31] These choices define not only the system's error profile, but also the kinds of clinical vigilance that will appear "reasonable" or "excessive" ex post.

Developers and vendors play a central role in this process. Through decisions about model architecture, training data, loss functions, and calibration strategies, they effectively encode normative judgments about which errors matter most and which risks are tolerable. A diagnostic system optimized to minimize false positives will, by design, increase false negatives; one calibrated to reduce unnecessary biopsies may systematically under-detect malignancies in certain subpopulations. These are not incidental technical details. They are policy choices with predictable clinical consequences, replicated across thousands of cases.[32]

Healthcare institutions add another critical layer of risk governance. Hospitals and health systems decide which AI tools to procure, how to configure them, and how tightly to integrate them into

---

[30] Maroudas, Vasileios P., *Fault-Based Liability for Medical Malpractice in the Age of Artificial Intelligence: A Comparative Analysis of German and Greek Medical Liability Law in View of the Challenges Posed by AI Systems*, 57 REV. EUR. & COMP. L. 135, 145–151 (2024); Aagaard Lise, *Artificial Intelligence Decision Support Systems and Liability for Medical Injuries*, 9 J. RES. PHARM. PRACT.125, 126–127 (2020); George Maliha et al., *supra* note 1, at 632–633; Clara Cestonaro et al., *supra* note 1, at 4–5.

[31] Aagaard Lise, *Id*, at 125–126; George Maliha et al., *Id*, at 632–639; Clara Cestonaro et al., *Id*, at 2–4.

[32] Aagaard Lise., *Id*, at 125–127; George Maliha et al., *Id*, at 632–633, 638–641; Clara Cestonaro et al., *Id*, at 3–4, 9.

clinical workflows. Interface design, alert fatigue mitigation, default settings, and documentation practices all shape how clinicians actually experience and rely upon algorithmic output.[33] A system presented as a "recommendation" may, in practice, function as an action-forcing directive when time pressure, staffing constraints, and institutional protocols converge. Conversely, a nominally advisory tool can become an inaction-forcing default when low-risk outputs are routed away from meaningful review. None of these governance choices are controlled by the individual physician facing a patient.

Empirical and doctrinal analyses of medical AI liability increasingly recognize this mismatch. Systematic reviews of AI-related malpractice risks show that existing legal frameworks tend to overload physicians with responsibility while under-deterring developers and institutions, despite the fact that systemic risk is overwhelmingly determined upstream.[34] The physician is asked to supply "independent judgment" precisely where meaningful independence is structurally undermined by information asymmetries, opaque model behavior, and institutional reliance on algorithmic outputs.

In this context, the physician may look like the CCA at the level of a single adverse outcome, but she is not the actor best positioned to perform the relevant cost—benefit analysis about algorithmic risk. She cannot recalibrate the model, retrain it on more representative data, audit its performance across populations, or modify its integration into clinical workflows. Those capacities lie with developers, vendors, and healthcare institutions—the actors who can reduce risk not just for one patient, but for all future patients simultaneously.

Medical AI therefore exemplifies the core claim of this Article: in algorithmic environments, the apparent CCA at the point of harm often diverges sharply from the true best cost—benefit decision-maker. Treating the clinician as the primary locus of liability rests on the "independent judgment" myth—the assumption that frontline professionals retain meaningful control over risks that are in fact governed elsewhere. Once that myth is exposed, responsibility must be reoriented toward the actors who actually set the baseline level of algorithmic risk and who can most

[33] Deimantė Rimkutė, *AI and Liability in Medicine: The Case of Assistive-Diagnostic AI*, 16 BALTIC JOURNAL OF LAW & POLITICS 64, 70–71, 79 (2023); Aagaard Lise., *Id*, at 125–126; Katherine Drabiak, *supra* note 17, at 10–13.
[34] Deimantė Rimkutė, *Id*, at 65–67; George Maliha et al., *supra* note 1, at 629–634; Clara Cestonaro et al., *supra* note 1, at 1, 4, 9–10.

efficiently alter it. In medical AI, those actors are the system's developers and the institutions that deploy and govern it.[35]

The same structural illusion reappears beyond the hospital walls: just as clinicians are told to exercise "independent judgment" over opaque diagnostic systems, drivers of semi-autonomous vehicles are cast as ever-vigilant supervisors—formally responsible, yet practically deprived of meaningful control over risks that are engineered elsewhere.

B. Autonomous and Semi-Autonomous Vehicles: The "Ready-to-Intervene" Myth

In the context of autonomous and semi-autonomous vehicles, contemporary liability frameworks often rely—explicitly or implicitly—on what may be called the "ready-to-intervene" assumption. Under this model, responsibility is anchored in the figure of the human driver who is formally designated as a supervisor: the system may control steering, acceleration, braking, and navigation, but the human operator is expected to remain vigilant, attentive, and capable of taking over instantaneously when the system encounters an unexpected scenario.

At first glance, this allocation of responsibility appears compatible with classical Cheapest Cost Avoider logic. The driver is physically present, directly connected to the vehicle's controls, and—at least in theory—capable of preventing harm by braking, steering, or disengaging automation. Courts and commentators therefore often treat the driver as the natural locus of duty: the last human in the loop, and thus the last chance to avert catastrophe.

Yet a growing body of empirical research in human automation interaction fundamentally undermines this premise. Continuous, high-quality monitoring of a highly reliable automated system is not merely difficult; it is cognitively and psychologically unrealistic.[36] Decades of research on vigilance, automation complacency, and skill degradation show that humans are systematically ill-suited to act as passive supervisors of systems that fail rarely, unpredictably, and at machine speed. The more reliable the system appears during ordinary operation, the more

[35] George Maliha et al., *Id*, at 630–640; Katherine Drabiak, *supra* note 17, at 3–6, 9–14; Deimantė Rimkutė, *Id*, at 65–71.
[36] Alice Guerra, Francesco Parisi & Daniel Pi, *Liability for Robots I: Legal Challenges*, 18 J. INST. ECON. 331, 332–333 (2022); A. Feder Cooper & Karen Levy, *Fast or Accurate? Governing Conflicting Goals in Highly Autonomous Vehicles*, 20 COLO. TECH. L.J. 221, 227–235 (2022); Muhammad Uzair, *supra* note 2, at 5–7.

human attention decays; the rarer the intervention opportunities, the slower and less effective the human response becomes when intervention is suddenly demanded.

The result is a structural mismatch between legal expectation and human capability. The driver is formally tasked with monitoring, but the system is designed in a way that predictably erodes the very vigilance that monitoring requires. Reaction times in takeover scenarios routinely exceed the temporal window in which avoidance is physically possible, especially when failures involve perception errors, misclassification of objects, or complex traffic dynamics that unfold faster than human situational awareness can recover. In such settings, the "ready-to-intervene" driver is less a genuine safeguard than a legal fiction.

By contrast, manufacturers and software providers occupy a radically different position in the risk architecture of autonomous driving systems. They determine how control is shared between human and machine, how and when control is transferred back to the driver, and how clearly takeover requests are communicated.[37] They design the sensor fusion logic, object-classification thresholds, and decision policies that govern how the vehicle prioritizes speed, comfort, and safety.[38] These actors also control the system's learning and updating processes, enabling them to respond to incidents by deploying software updates across entire fleets, redesigning user interfaces, recalibrating thresholds, or modifying default behaviors.[39]

Crucially, these upstream actors are capable of learning from accidents in a way that individual drivers cannot. A single driver experiences one crash; a manufacturer observes patterns across thousands or millions of miles. A driver cannot redesign the perception stack or adjust the system's tolerance for uncertainty; a manufacturer can. From a cost–benefit perspective, the manufacturer is therefore vastly better positioned to evaluate the trade-offs between false positives and false negatives, between earlier alerts and driver overload, and between aggressive automation and conservative fallback strategies.

---

[37] Xuan Di, Xu Chen & Eric Talley, *Liability Design for Autonomous Vehicles and Human-Driven Vehicles: A Hierarchical Game-Theoretic Approach*, 118 TRANSPORTATION RESEARCH PART C: EMERGING TECHNOLOGIES 1, 3–6 (2020); Muhammad Uzair, *Id*, at 5–7, 14–15.
[38] A. Feder Cooper & Karen Levy, *supra* note 36, at 232–235; Xuan Di, Xu Chen & Eric Talley, *Id*, at 1–6.
[39] Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J.L. & TECH. 171, 198–201 (2015); Xu Chen & Xuan Di, *supra* note 2, at 418–420.

Legal scholarship has increasingly recognized that traditional tort doctrines struggle to account for this asymmetry. Attempts to anchor liability in driver negligence rest on an implicit assumption that the driver is meaningfully capable of preventing the harm at reasonable cost. But when supervision itself is the weak link—when the system is engineered in a way that predictably defeats sustained human vigilance—the driver is not the cheapest cost avoider in any meaningful sense. Nor is the driver the best positioned decision-maker about system-level risk. That role belongs to the entities that design, train, deploy, and update the automated driving system in the first place.[40]

From the perspective advanced in this article, semi-autonomous driving thus exemplifies a broader pattern: the apparent CCA at the point of harm diverges sharply from the actor who actually governs risk at scale. Treating the driver as the primary bearer of liability not only misallocates responsibility, but also distorts incentives. It encourages manufacturers to externalize systemic design risk onto individual users, while providing weak legal pressure to redesign interfaces, takeover protocols, and automation boundaries in ways that reflect real human limitations.

Once this misalignment is acknowledged, the normative implication follows naturally. The best cost—benefit decision-maker in semi-autonomous vehicle systems is not the human monitor who is structurally set up to fail, but the manufacturer or software provider who can redesign the system to reduce the probability and severity of harm across an entire fleet. In BARG terms, the manufacturer is the actor best positioned to gather information about failures, evaluate trade-offs among alternative designs, and implement changes that propagate system-wide. Assigning liability accordingly does not absolve drivers of all responsibility, but it rejects the myth that human supervision can function as a robust fail-safe in environments engineered for machine control.

The collapse of the "ready-to-intervene" narrative in semi-autonomous driving thus exposes a broader structural pattern: across AI-mediated domains, tort law repeatedly assigns responsibility to human actors who appear to retain formal control, while the true governance of risk—through design choices, thresholds, and system architecture—resides elsewhere, a dynamic that becomes even more pronounced in algorithmic decision-making systems governing credit, insurance, and employment.

[40] Shi Rui, supra note 2, ta 159–162; Alice Guerra, Francesco Parisi & Daniel Pi, *supra* note **Error! Reference source not found.**, at 332–335; Muhammad Uzair, *supra* note 2; Xuan Di, Xu Chen & Eric Talley, *supra* note 37, at 2–6.

C. Algorithmic Credit, Insurance, and Employment: The "Neutral Score" Myth

In domains such as credit scoring, insurance underwriting, and algorithmic hiring, machine-learning systems are routinely framed as instruments of objectivity. By translating complex personal histories into numerical scores, these systems promise to replace subjective human judgment with neutral, data-driven assessment. Yet a substantial interdisciplinary literature has demonstrated that algorithmic scoring systems can encode, reproduce, and even amplify existing social hierarchies and structural inequalities, even when protected attributes are formally excluded from the model.[41]

From a surface-level tort perspective, the frontline clerk, loan officer, or HR professional applying the algorithmic output may appear to be the Cheapest Cost Avoider. In principle, this actor could question the score, request additional information, or deviate from the automated recommendation. This intuition fits comfortably within traditional negligence narratives: the human decision-maker is visible, proximate to the harm, and seemingly capable of exercising discretion at low cost.

In practice, however, this discretion is often more illusory than real. Institutional policies frequently instruct employees not to deviate from algorithmic outputs in the name of consistency, auditability, and the reduction of "human bias".[42] Deviations from automated recommendations may trigger internal scrutiny, disciplinary action, or accusations of arbitrariness, while adherence to the score is treated as compliance with objective procedure. The result is a familiar inversion: the human operator bears formal responsibility for the decision, but lacks meaningful authority to alter the risk calculus embedded in the system.

Meanwhile, the core cost–benefit decisions that shape discriminatory risk are made upstream, by institutions and vendors that design, train, and deploy the scoring systems. These actors select training datasets, define objective functions, and choose how to trade off accuracy, profit, and fairness across populations.[43] They decide whether, and how, to impose fairness constraints; which

---

[41] Lauri Kai, *supra* note 3, at 3–4, 12–14; Nicholas Schmidt & Bryce Stephens, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination*, arXiv 130, 130, 134–138 (2019); Xukang Wang et al., *supra* note 3, at 1–5; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 673–675, 691–692.
[42] Natalie Sheard, *supra* note 3, at 627–629, 633–635; Xukang Wang et al., *Id*, at 2, 5–7.
[43] Neil Menghani, Edward McFowland III & Daniel B. Neill, *Insufficiently Justified Disparate Impact: A New Criterion for Subgroup Fairness*, ARXIV:2306.11181 1, 1–2, 7 (2023); Greta Coraglia et al., *Evaluating AI Fairness in*

metrics count as acceptable performance; and what levels of disparate impact are tolerated as the price of efficiency.[44] Crucially, they also control access to portfolio-level and population-level data—the only vantage point from which systemic disparate impact can be detected, measured, and addressed.[45]

Extensive work in antidiscrimination law and algorithmic fairness has shown that persistent inequality in automated credit, insurance, and employment decisions is rarely the product of isolated frontline choices.[46] Rather, it emerges from design and deployment decisions that structure how risk is defined, measured, and distributed across groups. Treating the individual clerk or HR officer as the Cheapest Cost Avoider in this setting trivializes structural injustice by collapsing system-level governance failures into individualized moments of application.

From a cost–benefit perspective, the actor best positioned to govern algorithmic discrimination is the institution or vendor that sets the system's goals, fairness parameters, data governance practices, and override policies.[47] These actors can recalibrate thresholds, retrain models on more representative data, audit outcomes across populations, and redesign workflows to mitigate disparate impact at scale. By contrast, the frontline employee can at most alter the outcome of a single case, often at personal or professional risk, and without access to the information necessary to assess systemic effects.

The "neutral score" myth thus mirrors the broader pattern identified throughout this Article. The apparent CCA at the point of decision is not the actor who governs algorithmic risk in any meaningful sense. Liability regimes that fixate on downstream human application misdirect incentives, shielding those who define and control the system's distributive consequences while exposing individual employees to responsibility for harms they neither designed nor can effectively prevent.

Credit Scoring with the BRIO Tool, ARXIV:2406.03292 1, 2–3, 7, 15–16 (2024); Lauri Kai, *supra* note 3, at 2–4, 10–14, 20–28.

[44] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger & Suresh Venkatasubramanian, *Certifying and Removing Disparate Impact*, ARXIV:1412.3756 1, 2–3 (2015); Xukang Wang et al., *supra* note 3, at 2–4; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 688–692.

[45] Xukang Wang et al., *Id*, at 7–8; Solon Barocas & Andrew D. Selbst, *Id*, at 711–714.

[46] Lauri Kai, *supra* note 3, at 2–7; Natalie Sheard, *supra* note 3, at 625–627; Solon Barocas & Andrew D. Selbst, *Id*, at 673–675 ; Xukang Wang et al., *Id*, at 3–4.

[47] Solon Barocas & Andrew D. Selbst, *Id*, at 677–692; Xukang Wang et al., *Id*, at 1, 3–7; Lauri Kai, *Id*, at 12–14, 26–28.

Algorithmic scoring in credit, insurance, and employment thus exposes a familiar liability illusion: the law assigns responsibility to the last human decision-maker while ignoring where algorithmic risk is actually designed, calibrated, and governed. Downstream discretion is treated as control, even when it is procedurally discouraged and informationally empty. This misalignment is not confined to scoring systems. It resurfaces even more starkly in debates over generative AI, where users are portrayed as autonomous risk creators based on prompts and publication choices, while the architecture that shapes and constrains harmful outputs is built and controlled upstream. The next subsection examines this displacement of responsibility through the lens of generative AI and the "user-control" myth.

D. Generative AI and Content Harms: The "User-Control" Myth

Generative AI systems can produce a wide range of content-related harms, including defamation, copyright infringement, privacy violations, incitement, and other legally cognizable injuries. Much of the emerging legal and policy discourse frames these harms through the lens of user behavior. On this account, the end user who crafts the prompt or republishes the output is treated as the primary locus of responsibility: the user could have chosen different prompts, exercised greater restraint, or refrained from dissemination altogether.[48] From a traditional tort perspective, this framing is intuitively attractive. The user appears to be the CCA, exercising direct control at the moment closest to the harm.

Yet this intuition rests on a deeply misleading understanding of how generative AI systems actually generate, constrain, and propagate risk. A growing body of technical and governance-oriented scholarship demonstrates that content risk in generative AI is centrally shaped by platform-level design and deployment choices, including the structure of guardrails, filtering mechanisms, logging and monitoring practices, access tiers, and safety fine-tuning regimes.[49] At the same time, research on adversarial prompting and so-called "jailbreaks" shows that even well-

[48] Richard J. Tong et al., *A First-Principles Based Risk Assessment Framework and the IEEE P3396 Standard*, ARXIV:2504.00091 1, 2–6 (2025); Laura Weidinger et al., *Sociotechnical Safety Evaluation of Generative AI Systems*, ARXIV:2310.11986 1, 6–12 (2023); Susan Hao et al., *supra* note 4, at 2–4.
[49] Laura Weidinger et al., *Id*, at 6–11; Susan Hao et al., *Id*, at 1–5.

intentioned user conduct cannot reliably neutralize content risks when system-level constraints are brittle, incomplete, or strategically bypassable.[50]

These limitations are not accidental. Generative AI models are designed to be flexible, general-purpose, and responsive to a wide variety of inputs. As a result, platform providers face persistent trade-offs between expressive capacity, usability, and safety. The management of these trade-offs occurs upstream, through decisions about training data, reinforcement learning objectives, refusal behaviors, and post-deployment oversight. Empirical work on generative AI safety therefore emphasizes the centrality of systematic risk assessments, red-teaming, incident reporting, and iterative patching at the model and platform level.[51] These practices, when present, shape the baseline level of harm for all users simultaneously; when absent or under-resourced, they leave downstream actors exposed to risks they cannot meaningfully diagnose or mitigate.

By contrast, the user's preventive capacity is narrow and episodic. A user may avoid causing harm in a particular interaction, but lacks the ability to recalibrate default behaviors, redesign safety layers, or deploy updates that reduce risk across the system as a whole. In tort terms, the user's control is real but case-specific, while the platform's control is systemic. Treating the former as the primary focus of liability therefore misdirects incentives. It encourages after-the-fact blame of individual prompting or publication choices while undercutting investment in "safety by design" measures that only platform-level actors can implement.[52]

Once this structural asymmetry is made explicit, the identity of the relevant BDM becomes clearer. The most consequential cost—benefit decisions in generative AI concern how much harmful content is tolerable, which categories trigger refusals, how aggressively models generalize from training data, and how quickly vulnerabilities are detected and patched. These are not decisions that can be made meaningfully by end users. They require access to population-level data, technical expertise, and the capacity to implement changes that propagate across millions of interactions. In most generative AI ecosystems, these capacities reside with model developers and

[50] Banerjee, Somnath et al., *How (Un)Ethical Are Instruction-Centric Responses of LLMs? Unveiling the Vulnerabilities of Safety Guardrails to Harmful Queries*, ARXIV 193, 193–199 (2024); Federico Bianchi & James Zou, *Large Language Models Are Vulnerable to Bait-and-Switch Attacks for Generating Harmful Content*, ARXIV 1, 1–2 (2024).
[51] Richard J. Tong et al., *supra* note 48, at 1–5; Chen Chen et al., *supra* note 18, at 6–14.
[52] Ssan Hao et al., *supra* note 4, at 2–3; Laura Weidinger et al., *supra* note 48, at 7–12 ; Richard J. Tong et al., *Id*, at 2–6.

platform providers, sometimes in conjunction with large institutional deployers that integrate generative systems into consumer-facing products or services.[53]

From the perspective advanced in this Article, the "user-control" narrative therefore exemplifies a recurring liability illusion. Downstream discretion is treated as if it were genuine governance over risk, while the upstream actors who design, calibrate, and continuously update the system remain comparatively insulated from tort-based accountability. The result is a familiar misalignment: liability signals are sent to actors who cannot efficiently reduce future harm, while those who can redesign the system to alter the distribution of risk face weaker legal pressure to do so.

This misalignment does not require absolving users of all responsibility. Users may still bear duties related to intentional misuse, republication, or reliance in particular contexts. But focusing tort liability primarily on user behavior misconceives where meaningful prevention occurs in generative AI systems. The actor best positioned to evaluate cost–benefit trade-offs and to implement system level safeguards is typically the platform or model provider—the actor who functions, in practice, as the Best Algorithmic Risk Governor (BARG) for generative content risks.

Seen in this light, generative AI does not present an isolated anomaly in tort law, but rather a particularly clear manifestation of a broader structural error. Once again, liability intuition gravitates toward the last human actor in the chain, mistaking formal discretion for genuine control over risk. What appears as user choice at the surface is in fact tightly channeled by upstream architectures, defaults, and governance decisions that shape harmful outcomes across users and contexts. This displacement of responsibility is not unique to generative content systems. It recurs across AI-mediated domains whenever human involvement is preserved largely as a legal placeholder rather than as a realistic locus of risk governance. The next subsection distills this recurring error into a general insight—the "last human" pattern—and explains why it poses a foundational challenge to tort law's traditional liability framework.

E. Interim Insight: The "Last Human" Pattern

[53] Philipp Hacker et al., *supra* note 18, at 1115–1117; Laura Weidinger et al., *Id*, at 7–11; Richard J. Tong et al., *Id*, at 3–5.

The preceding sections examined distinct doctrinal arenas—medical AI, semi-autonomous vehicles, algorithmic scoring in credit and employment, and generative AI content systems. Although these domains differ technologically and institutionally, they expose a common structural feature of AI-mediated harm.

Tort intuition repeatedly converges on the same figure: the last human in the chain.

In each setting, the actor temporally and spatially closest to the injury—the physician reading the scan, the driver supervising the vehicle, the loan officer applying the score, the user crafting the prompt—appears, at first glance, to be the CCA. This actor occupies the final decision node, fits comfortably within existing negligence narratives, and offers courts a familiar anchor for duty, breach, and causation analysis. The legal story is straightforward: *a human saw the output, could have done more, and did not.*

But once the analysis shifts from the isolated incident to the architecture of algorithmic risk production, this intuition begins to fail.

Across AI ecosystems, the most consequential cost—benefit trade-offs are rarely made at the point of application. They are embedded upstream in decisions about model architecture, training data, objective functions, confidence thresholds, escalation rules, interface defaults, workflow integration, and post-deployment monitoring and updating. These choices determine the system's baseline error profile—which errors are frequent, which are rare, and which are structurally obscured. They also determine how much room for meaningful human intervention actually exists downstream. The frontline actor's "discretion" is therefore often constrained by design, channeled by institutional protocol, and informationally hollowed out by opacity.

This produces a systematic divergence that recurs across domains:

The apparent CCA is often the last manual link in the chain.
The true best cost–benefit decision-maker—the actor positioned to govern algorithmic risk systemically—is typically located upstream, where architectures, datasets, thresholds, and governance processes are shaped.[54]

[54] Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 13–15; Muhammad Uzair, *supra* note 2, at 12–15; Laura Weidinger et al., *supra* note 48, at 7–9, 22–25; George Maliha et al, *supra* note 1, at 630–634, 637–641; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 677–692, 715–722.

This is the "Last Human" pattern: tort law's liability intuition gravitates toward the most visible human actor, even when that actor lacks meaningful control over the system-level parameters that structure risk across cases. Visibility is mistaken for governance. Proximity to harm is conflated with capacity to reduce future harm.

Doctrinally, this pattern reveals a structural mismatch between classical CCA reasoning and AI-mediated harm. In many traditional accident settings, the CCA and the Calabresi–Hirschoff BDM coincide. AI systems fracture that alignment. The actor who could, in principle, avert *this* instance of harm at the last moment is often not the actor who can most efficiently redesign the system to reduce the class of harms from which this instance emerged.

The consequence is not merely analytical. If courts continue to map liability reflexively onto the last human in the loop, they risk over-deterring actors with limited systemic leverage while under-deterring those who actually shape the distribution of algorithmic risk. Tort law would then function less as an incentive mechanism for safer design and more as a mechanism of scapegoating.

This interim insight therefore necessitates a new organizing concept—one that tracks governance capacity over algorithmic risk, rather than mere proximity to injury. The next section introduces that concept: the BARG—the actor best positioned to gather information, perform the relevant cost–benefit analysis, and implement system-level changes that alter the baseline level of risk for many users at once.

V. Re-Centering Tort Law: The Actor Who Actually Governs Algorithmic Risk

A. Identifying the Real Risk Governor

Once AI-related harm is understood as the product of layered architectures, radical information asymmetries, dynamic updating, and pattern-based risk, the central tort inquiry must be reformulated at a more structural level. The relevant question is no longer merely who could have prevented *this* injury at the last moment, but rather who governs the system that makes injuries of this type systematically more or less likely across cases. The shift is from episodic control to systemic governance, from the moment of harm to the architecture of risk production. The BARG concept provides the analytic framework for performing that shift within tort doctrine.

The BARG is the actor (or set of actors) in an AI ecosystem best positioned to gather and interpret information about algorithmic risks and harm patterns across users and contexts; to evaluate cost—benefit trade-offs among competing design, deployment, and governance options; and to implement system-level changes—technical or organizational—that modify the baseline level of algorithmic risk for many users at once. Each component of this definition is critical. Information-gathering matters because algorithmic risk is statistical and distributed rather than episodic. Cost–benefit evaluation matters because AI safety decisions are rarely binary; they involve continuous trade-offs among accuracy, robustness, fairness, and operational constraints. System-level implementation capacity matters because only interventions that propagate across users can meaningfully alter the overall distribution of harm rather than merely shifting outcomes in isolated instances.

The BARG framework therefore treats algorithmic risk as a governance problem embedded in socio-technical systems rather than as a series of disconnected human mistakes. It recognizes that meaningful control over AI-related harm depends on actors who operate at the level of model design, data governance, deployment architecture, monitoring, and updating practices—domains where the structure of risk is defined before any particular user encounters the system. By centering these actors, the BARG concept translates tort law's concern with efficient accident prevention into the language of algorithmic system governance, preserving the economic logic of liability while adapting it to environments characterized by scale, opacity, and continuous change.[55]

B. Why BARG Is More Than BDM Applied to AI

At this point, a fair objection must be confronted directly. If the BARG framework builds on Calabresi and Hirschoff's best decision-maker analysis, is it doing anything more than giving the BDM concept a new label for the age of artificial intelligence? The answer is yes—but the claim requires precision. BARG does not purport to replace the BDM framework, nor does it reject the economic logic from which BDM emerged. Rather, BARG operationalizes that logic under conditions that the classical framework did not have to confront in systematic form: layered

[55] Gabriel Lima & Meeyoung Cha, *supra* note 20, at 1–3; Chen Chen et al., *supra* note 18, at 1–14, 67–70; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 1–4, 7–9; Madalina Busuioc, *supra* note 7, at 828–834.

algorithmic production, radical information asymmetries, continuous updating, feedback loops, and pattern-based harm at scale.

The BDM test asks the right foundational question: who is best positioned to make the relevant cost–benefit decision and to act upon it? In simple or moderately complex accident settings, that inquiry may be sufficient. The court can ask which party had superior information, superior capacity to compare accident costs and avoidance costs, and superior ability to adjust behavior accordingly. In such settings, the identity of the BDM can often be inferred from ordinary indicia of control: who designed the product, who operated the activity, who selected the precaution, or who could have altered the relevant conduct at reasonable cost. But AI systems alter the structure of the inquiry. They do not merely add technical complexity to familiar tort problems; they change where risk is generated, where it can be observed, and where it can be reduced.

BARG is therefore not a competing theory of accident-cost allocation, but a more specific governance-oriented instantiation of BDM for algorithmic systems. Its analytic contribution lies in shifting the inquiry from decisional capacity in the abstract to risk-governance capacity in a particular socio-technical ecosystem. In AI-mediated environments, the relevant cost–benefit decision is rarely a single discrete choice made at the point of harm. It is usually embedded in a series of upstream and ongoing decisions: what data to use, how to train the model, which loss function to privilege, how to calibrate thresholds, how to design the interface, how to structure human oversight, how to monitor performance after deployment, and when to update, retrain, restrict, or withdraw the system. These are not merely technical implementation details. They are the institutional locations at which the accident-cost calculus is actually performed.

This is why BARG adds something that BDM alone leaves underdeveloped. BDM identifies the type of actor tort law should care about: the actor capable of making and acting upon the relevant cost–benefit judgment. BARG identifies the kinds of capacities that matter when that judgment concerns algorithmic risk: access to training and performance data, control over model architecture, ability to observe outcomes across a population of cases, capacity to detect drift or disparate error patterns, authority to change deployment conditions, and ability to propagate safety improvements across users. Put differently, BARG translates BDM’s general economic question into operational markers suited to algorithmic environments. It asks not only who is best placed to decide, but who is best placed to govern the system through which risk is repeatedly produced.

The distinction also matters because BARG can generate different liability outcomes than a court might reach under an undifferentiated BDM analysis. A general BDM inquiry may still be pulled toward the actor who appears to exercise professional or operational judgment in the individual case: the physician who accepted an AI-generated diagnosis, the driver who failed to resume control, the loan officer who relied on a proprietary score, or the content moderator who applied a platform's automated classification. Each of these actors appears to make a "decision." Each is proximate to the injury. Each may be described, at least formally, as having the last opportunity to avoid harm. But BARG directs attention away from formal decisional moments and toward the actor that controls the risk architecture that made those moments legally and practically meaningful. In many AI cases, that actor will be upstream: the developer, platform provider, or institutional deployer that controls system design, monitoring, updating, and integration.

This does not mean that BARG always selects the same defendant, or that it mechanically assigns liability to the most technologically sophisticated actor. The point is functional, not categorical. A model developer may be the BARG where the relevant risk arises from training data, architecture, calibration, or known model limitations. An institutional deployer may be the BARG where the risk arises from workflow integration, staff incentives, override policies, alert thresholds, or failure to monitor local performance. A platform provider may be the BARG where it controls logging, guardrails, access rules, safety tooling, or system-wide updates. In some cases, algorithmic risk will be jointly governed, and more than one actor may perform BARG-like functions. But even in those cases, the framework clarifies the relevant inquiry: liability should track functional governance over the risk, not merely proximity to the injury or formal participation in the final decision.

BARG also functions as a doctrinal tool rather than merely a descriptive label. It can guide courts in applying familiar tort concepts—duty, breach, causation, and the choice between negligence, product liability, and institutional liability—without requiring courts to invent an entirely new cause of action. At the duty stage, BARG helps identify which actors have sufficient control over algorithmic risk to justify legally cognizable obligations of care. At the breach stage, it focuses the standard of care on governance practices such as validation, documentation, monitoring, updating, warnings, escalation pathways, and realistic human-override design. At the causation stage, it helps courts see that the legally relevant cause may lie not in a single downstream act, but in a system configuration that materially increased a class of foreseeable harms. And at the remedy and

deterrence stages, it channels liability toward the actors capable of changing the system for future cases, rather than merely punishing the last human link in the chain.

In this sense, BARG is best understood as a bridge between economic tort theory and algorithmic governance. It preserves the central insight of CCA and BDM analysis—that liability should be assigned in a way that induces the actor best positioned to reduce accident costs to do so. But it recognizes that, in AI ecosystems, accident costs are often shaped by governance decisions that are distributed, technical, dynamic, and invisible at the point of injury. The legal challenge is therefore not simply to find the cheapest cost avoider or even the best abstract decision-maker. It is to identify the actor with the institutional capacity to observe, evaluate, and alter the algorithmic system that generates risk across cases.

That is the conceptual work BARG performs. It does not claim that BDM was wrong; it claims that BDM requires specification when the "decision" at issue is no longer a discrete human choice but an ongoing governance process embedded in software, data, organizational routines, and platform architecture. BARG is that specification. It turns the BDM insight into a usable liability framework for AI by asking where algorithmic risk is actually governed—and by insisting that tort law's incentives should be directed there.

C. Functional Markers of Algorithmic Control

To operationalize the BARG framework, courts require criteria that move beyond formal labels and toward functional indicators of algorithmic control. The central question is not who is nominally responsible for an AI system, but which actor (or actors) possesses the practical capacity to shape the system's risk profile across cases. The following functional markers identify where meaningful governance over algorithmic risk actually resides.

To make this inquiry more concrete, courts can treat BARG identification as a functional, multi-factor test. The factors below are not mechanically dispositive; rather, they indicate where algorithmic risk is actually designed, observed, updated, and internalized. The greater an actor's control over the high-weight factors, the stronger the case for treating that actor as the relevant BARG.

Table 1. A Multi-Factor Test for Identifying the Best Algorithmic Risk Governor

Factor	Relative Weight	Doctrinal Significance	Typical Indication of BARG Status
Control over model architecture and system design	High	Identifies the actor that defines the system's baseline risk profile before any downstream use occurs.	The actor selects or controls model architecture, training pipelines, parameters, thresholds, interfaces, or core safety features.
Control over training data and data governance	High	Shows who shapes the informational foundation from which algorithmic errors, biases, and blind spots emerge.	The actor selects, curates, labels, validates, cleans, excludes, or updates the data on which the system depends.
Access to outcome data and population-level observability	High	Determines who can detect recurring error patterns, disparate impacts, drift, or systematic underperformance across cases.	The actor possesses logs, performance metrics, incident reports, post-deployment outcomes, feedback data, or auditing capacity across users or populations.
Ability to update, retrain, recalibrate, suspend, or withdraw the system	High	Captures who can respond to liability signals by reducing future risk at scale rather than merely correcting isolated outcomes.	The actor can deploy patches, change thresholds, retrain the model, alter guardrails, rollback versions, restrict use, or remove the system from deployment.
Control over deployment conditions and workflow integration	High/Medium	Identifies who determines how algorithmic output is translated into real-world decisions and whether human oversight is meaningful or merely formal.	The actor controls procurement, configuration, escalation rules, override policies, alert settings, user training, staffing incentives, or institutional reliance on the system.
Capacity to monitor, audit, and document ongoing performance	High/Medium	Links BARG status to the ability to maintain reasonable post-deployment	The actor conducts or controls validation, auditing, documentation, red-teaming, drift

		governance over a dynamic system.	detection, incident response, or periodic review.
Ability to spread costs and internalize residual risk	Medium	Reflects traditional tort-law concerns with efficient cost allocation, insurance, pricing, and enterprise-level risk management.	The actor can insure, price risk into the service, distribute losses across users or customers, or invest in system-wide precautions.
Contractual or practical authority over other actors in the AI value chain	Medium	Shows whether the actor can impose safety obligations on developers, vendors, deployers, or users even without direct technical control.	The actor can require documentation, audits, safety standards, data access, compliance reports, indemnification, or contractual safeguards.
Proximity to the immediate harm	Low	Helps explain why the last human actor should not automatically be treated as the BARG merely because she is closest to the injury.	The actor is present at the point of harm but lacks meaningful control over design, data, monitoring, updating, or system-level risk reduction.
Formal human discretion at the point of use	Low	Distinguishes nominal decision-making from genuine governance capacity.	The actor can accept, reject, or question an output in a single case, but cannot alter the system's risk architecture or observe patterns across cases.

This table should be read functionally rather than formally. No single factor is necessary in every case, and no factor is always sufficient on its own. The inquiry is cumulative: an actor that controls design, data, observability, updating, and deployment conditions will ordinarily be a strong BARG candidate, even if it is distant from the immediate injury. Conversely, an actor who is proximate to the harm but lacks access to system-level information, updating authority, or meaningful control over deployment should not be treated as the BARG merely because she is the last human in the loop. The table therefore translates the BARG concept into a doctrinally usable

test: responsibility should track functional governance over algorithmic risk, not formal proximity to an individual accident.

1. Design control

The most salient marker of algorithmic control is authority over system design. Actors who determine model architecture, training pipelines, parameter tuning, confidence thresholds, user interfaces, and integration into organizational workflows effectively define the system's baseline error distribution and its interaction with human users.[56] These design choices encode *ex ante* judgments about acceptable trade-offs between false positives and false negatives, sensitivity and specificity, automation and discretion. Because such decisions are made upstream and propagate across all downstream uses, design control is a strong indicator of BARG status.

2. Data access and observability

Algorithmic risk is statistical and pattern-based rather than episodic. Accordingly, the ability to observe system behavior at scale is a second critical marker. Actors with access to logs, performance metrics, post-deployment outcome data, and feedback across populations are uniquely positioned to detect systematic failure modes, bias, drift, or feedback-loop amplification.[57]By contrast, actors limited to isolated encounters with the system lack the informational basis required for meaningful cost—benefit analysis of algorithmic precautions. Control over observability therefore maps closely onto control over risk governance.

3. Systemic leverage

A third marker concerns the capacity to implement changes that propagate broadly rather than locally. Actors who can deploy updates, patches, configuration changes, policy modifications, or infrastructural constraints can alter the system's behavior for many or all users simultaneously.[58] This capacity distinguishes governance from discretion: the former reshapes the risk environment itself, while the latter merely affects outcomes in individual cases. Systemic leverage is thus central to identifying who can actually respond to liability signals by reducing future harm.

4. Economic capacity and incentives.

Finally, tort law has long emphasized the importance of assigning liability to actors who can

---

[56] Muhammad Uzair, *supra* note 2, at 5–6, 12–13; Weisz et al., *supra* note 4, at 5–9; George Maliha et al, *supra* note 1, at 633–634.

[57] Savina D. Kim, Stefan Lessmann, Galina Andreeva & Michael Rovatsos, *supra* note 26, at 10–15; Adriano Koshiyama et al., *supra* note 21, at 4–6; Madalina Busuioc, *supra* note 7, at 826, 829–834; Xukang Wang et al., *supra* note 3, at 6–7.

[58] Aagaard Lise, *supra* note 30, at 125–126; Laura Weidinger et al., *supra* note 48. At 7–9, 19–20, 22–23, 27–28; Philipp Hacker et al., *supra* note 18, at 1115–1116, 1118–1120; Adriano Koshiyama et al., *Id*, at 4–7.

internalize the costs of precaution and residual harm. In algorithmic environments, this consideration takes on renewed importance. Actors with the financial capacity to invest in safer design, monitoring, and governance—and to spread residual risk through pricing or insurance—are better positioned to respond efficiently to liability incentives without undermining socially valuable innovation.[59] Economic capacity therefore functions as both an efficiency filter and a realism constraint on BARG identification.

Taken together, these functional markers underscore that BARG status is not tied to formal proximity to the harm or to nominal human involvement at the point of decision. It tracks control over the architecture, data, and update mechanisms that generate risk across cases. Importantly, the BARG is not necessarily a single entity. In many AI ecosystems, a primary BARG (such as a model developer or platform provider) will coexist with secondary BARGs (such as large institutional deployers) whose integration and governance choices materially shape system-level risk.

The BARG is not necessarily a single entity; in many cases, a *primary BARG* (e.g. the model developer) will coexist with *secondary BARGs* (e.g. a large institutional deployer or platform provider).

These functional markers serve a deliberately pragmatic role. They translate the abstract insight that algorithmic risk is governed upstream into criteria that courts can actually apply when allocating liability. Rather than asking who touched the harm or who formally retained discretion at the point of decision, the markers redirect attention to where algorithmic risk is designed, observed, and recalibrated over time. In doing so, they provide a bridge between the economic logic of the best cost–benefit decision-maker and concrete doctrinal analysis.

The sections that follow build on this framework to show how traditional tort doctrines—negligence, product liability, and institutional responsibility—can be reinterpreted once liability is anchored in functional control over algorithmic systems rather than episodic human involvement. Read through the lens of these markers, familiar doctrines no longer ask whether the last human actor behaved reasonably in isolation, but whether the relevant risk governor exercised reasonable

[59] Emanuela Carbonara, Alice Guerra & Francesco Parisi, *supra* note 9, at 173–176, 190–191; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 8–11; Roberto Pardolesi & Bruno Tassone, *supra* note 5, 11–12, 29–31.

care in designing, monitoring, and updating a system whose errors predictably propagate across cases.

D. From Cheapest Cost Avoider to Risk Governor

The conceptual move from the CCA to the BARG does not abandon the economic logic of tort law, but rather exposes its limits in contemporary technological settings. Classical CCA analysis was designed for environments in which accident prevention is relatively localized and where the actor best positioned to avert harm in a given instance is also the actor who can efficiently internalize the costs of precaution. In such contexts, identifying the CCA provides a workable proxy for allocating responsibility in a way that minimizes the social costs of accidents.

Yet even within Calabresi's own framework, this proxy was never meant to be exhaustive. Where harm prevention depends not on simple, observable precautions but on complex judgments about design, information, and system-wide trade-offs, the identity of the cheapest cost avoider becomes increasingly indeterminate. This concern motivated Calabresi and Hirschoff's refinement of the analysis through the BDM concept, which shifts the focus from physical prevention to decisional capacity. The relevant inquiry becomes who is best situated to perform the cost—benefit analysis between accident costs and accident-avoidance costs and to act upon that analysis in a meaningful way. Seen in this light, the BDM framework already anticipates settings in which responsibility should attach not to the last actor in the causal chain, but to the actor who governs the conditions under which risk is produced and managed. The BARG concept builds directly on this insight, extending it to algorithmic environments in which risk is systemic, probabilistic, and generated through layered socio-technical systems rather than isolated human acts.[60]

Algorithmic systems make this extension necessary. In AI-mediated environments, the actor who appears to be the CCA in any individual incident—often a frontline human user—is frequently not the actor who governs risk in a meaningful sense. While such users may retain nominal discretion at the point of application, they typically lack control over the design choices, training data, thresholds, deployment architecture, and updating practices that determine the system's baseline error profile. By contrast, upstream actors—such as developers, platform operators, and large institutional deployers—are positioned to observe algorithmic behavior across cases, to evaluate

[60] *See* generally John C.P. Goldberg, *supra* note 5; Roberto Pardolesi & Bruno Tassone, *supra* note 5, at 11–12.

trade-offs among alternative configurations, and to implement changes that propagate across many decisions simultaneously. In these settings, the capacity to reduce harm lies less in last-moment intervention and more in the governance of the system that repeatedly generates risk. The divergence between the apparent CCA and the actor who actually governs algorithmic risk is therefore not accidental, but structural.[61]

Understood in this way, the Best Algorithmic Risk Governor is best seen as an algorithmic specification of the BDM rather than as a departure from established tort theory. It captures the intuition that, where risk is produced by complex and scalable systems, responsibility should follow governance capacity rather than proximity to harm. The shift from CCA to BARG thus reframes tort law's efficiency inquiry without altering its core commitment: allocating liability so as to incentivize those actors who are best positioned to reduce the social costs of harm.

This conceptual shift from accident prevention to risk governance sets the stage for examining how algorithmic systems systematically decouple apparent control from actual control—and why that decoupling repeatedly misleads tort law's traditional liability intuitions.

E. Limits and Objections: Over-Deterrence, Operational Knowledge, and Shared Governance

The BARG framework is not without limits. Indeed, if it were read too aggressively, it could reproduce a familiar problem in tort law: a concept designed to improve deterrence might become a vehicle for excessive liability. Three objections therefore require direct attention before the doctrinal implications are developed. First, would shifting liability toward developers, platforms, and institutional deployers over-deter socially valuable AI innovation? Second, does the framework understate the operational knowledge possessed by hospitals, banks, employers, and other institutional users—knowledge that may be unavailable to model developers? Third, if algorithmic risk is often jointly produced by developers, deployers, and users, why should tort law search for a single BARG rather than apportion responsibility among multiple actors?

The over-deterrence objection is serious, but it rests on a misunderstanding of what BARG requires. The framework does not impose strict, automatic, or enterprise-wide liability for every

[61] Laura Weidinger et al., *supra* note 48, at 7–12; George Maliha et al, *supra* note 1, at 630–635; Muhammad Uzair, *supra* note 2, at 1–8; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 674–692.

harm involving an AI system. Nor does it treat technological sophistication as a sufficient basis for responsibility. BARG identifies the actor best positioned to govern the relevant risk: the actor with access to the information, design levers, monitoring capacity, updating authority, and institutional ability to reduce future harms across cases. Liability should attach to failures of reasonable algorithmic governance, not to the mere fact that an AI system was used or that an AI-related harm occurred.

This distinction matters because innovation can be chilled not only by excessive liability, but also by poorly targeted liability. A regime that places primary responsibility on frontline users—physicians, drivers, loan officers, or employees—may discourage adoption in high-value settings while doing little to improve system safety. Those actors often cannot redesign the model, audit population-level error patterns, recalibrate thresholds, or update the system. By contrast, liability directed toward genuine BARGs creates incentives for safer innovation: better documentation, stronger validation, more realistic human-override design, continuous monitoring, incident response, and timely updating. The point is not to punish AI development, but to distinguish responsible governance from ungoverned deployment.

Over-deterrence concerns can also be addressed through doctrinal calibration. Courts need not treat BARG status as conclusive of breach. An actor may be the relevant risk governor and still demonstrate reasonable care. Evidence of meaningful pre-deployment testing, post-deployment monitoring, documentation, auditing, version control, incident response, and good-faith recalibration should matter in determining breach. Similarly, carefully designed regulatory safe harbors may be appropriate where an actor can show not only formal compliance, but sustained governance practices capable of detecting and correcting emerging risks. In this sense, the BARG framework is compatible with innovation-protective doctrines, so long as protection follows demonstrated governance performance rather than mere certification or formal labeling.

The second objection is that information asymmetry cuts both ways. Developers and platform providers often control model architecture, training data, documentation, logs, and update mechanisms. But hospitals, banks, insurers, employers, and other institutional deployers may possess operational knowledge that developers lack. A hospital may know how a diagnostic system is actually integrated into emergency-room workflows; a bank may know how loan officers are instructed to rely on a score; an employer may know how an automated screening tool

interacts with productivity pressures, applicant pools, or internal compliance incentives. In these contexts, the developer may understand the model, but the deployer understands the environment in which the model becomes legally consequential.

This objection does not defeat BARG; it refines it. BARG is a functional inquiry, not a categorical preference for upstream developers. Where the relevant risk arises primarily from model design, training data, calibration, or known technical limitations, the developer or platform provider will often be the primary BARG. Where the risk arises from procurement choices, workflow integration, override policies, alert thresholds, staffing pressures, user training, or failure to monitor local performance, the institutional deployer may be the primary or at least a secondary BARG. The framework therefore accommodates the fact that different actors possess different forms of knowledge. The legally relevant question is not who knows everything, but who controls the particular knowledge and intervention points necessary to reduce the relevant risk.

This is especially important in high-stakes institutional settings. A developer may provide a model with disclosed limitations, but a hospital may deploy it in a setting for which it was not validated, suppress alerts to reduce noise, or instruct clinicians to treat low-risk outputs as presumptively reliable. A credit-scoring vendor may design a model, but a bank may determine the threshold at which human review is bypassed or discouraged. An employer may purchase an automated screening tool but decide how heavily to weight its output, whether to audit disparate impact, and whether employees are permitted to override the recommendation. In such cases, the institutional deployer is not a passive consumer of technology. It is an active governor of algorithmic risk.

The third objection concerns shared governance. Many AI harms are not produced by one actor alone. They emerge from the interaction of model developers, platform providers, institutional deployers, and sometimes users. A model may be poorly documented by the developer, inadequately configured by the deployer, and uncritically followed by the end-user. If BARG were understood to require one exclusive defendant, it would oversimplify the very socio-technical complexity the framework is designed to capture.

The better reading is different. BARG need not always be singular. In some cases, there will be one dominant risk governor; in others, there will be primary and secondary BARGs; and in still others, responsibility should be apportioned among several actors according to their respective governance functions. A developer may be responsible for architecture, training data, calibration,

and warnings. A platform may be responsible for logging, guardrails, access controls, and system-wide updates. An institutional deployer may be responsible for procurement, local validation, workflow integration, override policy, monitoring, and escalation practices. A frontline user may remain responsible where she possesses meaningful information and realistic discretion in the particular case. But proximity alone should not transform the user into the principal risk governor.

Shared governance therefore does not undermine the BARG framework. It explains why the framework must remain functional and comparative. The task is not to search metaphysically for "the" actor behind an AI harm, but to map the functions through which risk was designed, observed, deployed, and controlled. Tort law already has tools for dealing with multiple responsible actors, including comparative fault, contribution, indemnity, and joint or several liability where appropriate. BARG does not eliminate those tools. It helps courts decide how to use them by identifying which actors exercised which forms of governance over the risk-generating system.

These objections ultimately sharpen the theory rather than weaken it. BARG is not a command to impose maximal liability on upstream actors. It is a method for aligning responsibility with practical governance capacity. It guards against over-deterrence by tying liability to unreasonable governance failures rather than to AI use as such. It accounts for two-sided information asymmetries by recognizing that institutional deployers may sometimes be primary risk governors. And it accommodates shared governance by allowing courts to identify primary and secondary BARGs and apportion responsibility according to functional control. Properly understood, the framework does not simplify algorithmic ecosystems into a single defendant; it gives tort law a structured way to see where meaningful control over algorithmic risk actually resides.

F. Stop Asking Who Touched the Harm—Ask Who Governs the System

At the point where algorithmic harm materializes, tort law's instinctive move is still to ask a familiar and deceptively simple question: *who touched the harm?* Who made the last decision, who relied on the output, who could have intervened at the final moment before injury occurred? This instinct reflects a deep structural feature of negligence doctrine, which has long been organized around discrete encounters, identifiable human actors, and localized failures of care.

Yet in AI-mediated environments, this question is increasingly orthogonal to the problem tort law purports to solve. The decisive determinants of harm are rarely located at the moment of application. They are embedded upstream—in system architecture, training data, model calibration, deployment design, and post-deployment governance. Focusing liability analysis on the last human touchpoint therefore risks mistaking proximity for control, and visibility for governance.

Once harms are produced by socio-technical systems that operate at scale, update dynamically, and distribute risk across thousands or millions of decisions, the analytically prior question must shift. The relevant inquiry is not *who touched the harm*, but *who governs the system that makes harms of this type systematically more or less likely*. Tort law, if it is to remain an instrument of efficient accident prevention, must reorient its liability lens toward actors with genuine system-level control.

This shift follows directly from the logic of the best cost–benefit decision-maker. In algorithmic environments, meaningful cost–benefit analysis does not occur at the point of use. It occurs where actors can observe aggregate performance, detect patterns of failure, and compare alternative designs and governance strategies. Those actors—developers, platform operators, and large institutional deployers—are uniquely positioned to gather and interpret information about algorithmic risks across populations, contexts, and time.[62] They can observe error distributions, model drift, feedback effects, and disparate impact patterns that remain invisible to frontline users operating case by case.

Crucially, these upstream actors also make the baseline risk decisions in the first place. Choices about training data composition, loss functions, confidence thresholds, escalation rules, and workflow integration determine which errors will be common, which will be rare, and which will be structurally obscured. These design and deployment choices encode normative judgments about acceptable risk long before any particular harm occurs.[63] By the time a physician, driver, clerk, or user encounters an algorithmic output, the system's error profile has already been largely fixed.

---

[62] Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 2–3. 6–7; Adriano Koshiyama et al., *supra* note 21, at 2–5, 15–17, 26–29; Madalina Busuioc, *supra* note 7, at 828–829, 831–834 .

[63] George Maliha et al, *supra* note 1, at 632–634, 638–641; Muhammad Uzair, *supra* note 2, at 5–8, 10–11, 14–16, 24; Lauri Kai, *supra* note 3, at 11–14, 20–22, 30–33.

Finally, tort law's efficiency rationale has always been forward-looking. Liability is justified not merely as retrospective blame, but as a mechanism for inducing future risk reduction. That mechanism operates only if liability is placed on actors who can respond to legal signals by redesigning systems rather than merely altering isolated behavior. Actors with the capacity to update models, recalibrate thresholds, modify interfaces, and deploy fixes across an entire user base are precisely those who can respond most efficiently to liability pressure by reducing future harms at scale.[64]

Seen in this light, the familiar focus on the last human in the loop is not simply incomplete—it is systematically misleading. The actor who appears to be the cheapest cost avoider in a single incident is often not the actor who governs algorithmic risk across incidents. Tort law's traditional fixation on the moment of harm therefore risks over-deterring downstream users while under-deterring the upstream entities that actually shape the system's risk architecture.

The central claim of this section is thus straightforward: in AI contexts, tort law must pivot from an event-based inquiry to a governance-based inquiry. The proper focal point of liability is the actor—or constellation of actors—who controls the informational, architectural, and organizational levers through which algorithmic risk is produced and managed. Asking who governs the system, rather than who touched the harm, is not a normative luxury. It is a doctrinal necessity if tort law is to align responsibility with real capacity for risk reduction in algorithmic societies.

VI. From Accidents to Architecture: Tort Law in the Age of Algorithmic Risk

A. Negligence Revisited: Reasonable Care in Algorithmic Systems

Negligence doctrine has long revolved around a deceptively simple inquiry: did the defendant fail to exercise reasonable care under the circumstances? In economic terms, this inquiry is often glossed through a Hand-style cost—benefit analysis, asking whether the burden of additional precautions would have been lower than the expected reduction in accident costs. In traditional,

[64] Emanuela Carbonara, Alice Guerra & Francesco Parisi, *supra* note 9, at 174–176, 191–194; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 8–13; Roberto Pardolesi & Bruno Tassone, *supra* note 5, at 11–12, 15–16, 25, 29, 31–32.

non-algorithmic settings, this framework typically directs courts toward the actor closest to the injury—the person who acted, failed to act, or could have intervened at the decisive moment.

In algorithmic systems, however, this intuitive mapping between reasonable care and proximity to harm systematically breaks down. The content of "reasonable care" cannot be specified by focusing solely on the last human decision-maker, because the most consequential cost–benefit decisions that shape algorithmic risk are made elsewhere: upstream, *ex ante*, and at scale. As a result, negligence analysis in AI contexts must be reinterpreted through the lens of the BARG—the actor best positioned to evaluate and govern algorithmic risk across cases rather than merely react to it in isolated incidents.

For **developers and vendors**, reasonable care encompasses a suite of system-level governance practices rather than isolated technical competence. It includes robust training and validation procedures; careful selection and documentation of training data; bias, robustness, and stress testing across relevant subpopulations; version control and change tracking; and meaningful mechanisms for post-deployment monitoring, incident response, and recall. These practices are not ancillary to negligence analysis—they define the baseline risk profile of the system itself. Choices about model architecture, loss functions, thresholds, and retraining cadence embed normative judgments about which errors are acceptable and which harms are tolerable, and they predictably shape outcomes across thousands or millions of downstream decisions.[65]

For **institutional deployers**—such as hospitals, banks, insurers, platforms, and large employers—reasonable care lies not in blind reliance on vendor assurances, but in active governance of algorithmic integration. This includes due diligence in system selection; scrutiny of documented limitations and known failure modes; ongoing auditing of performance in real-world conditions; clear and realistic human-override policies; and training for frontline staff that reflects actual system behavior rather than formal disclaimers. Crucially, institutional choices about workflow integration, alert thresholds, default settings, and escalation procedures can transform an ostensibly advisory tool into an action-forcing or inaction-forcing mechanism. Negligence analysis

[65] Katherine Drabiak, *supra* note 17, at 4–8, 11–14; Miriam Buiten, Alexandre de Streel & Martin Peitz, *Id*, at 2–8, 13, 15–16; Adriano Koshiyama et al., *supra* note 21, at 2–7, 15–17, 28; George Maliha et al, *supra* note 1, at 632–634, 638–641.

must therefore attend to how institutional configuration decisions materially shape risk, even when the underlying model remains unchanged.[66]

By contrast, for **frontline users**—physicians, drivers, loan officers, moderators, or individual consumers—reasonable care should be defined more narrowly. It may include appropriate reliance on AI outputs given the information reasonably available, adherence to institutional protocols, and communication with affected individuals. But it should not be stretched to encompass system-level governance responsibilities that users neither control nor meaningfully understand. Treating frontline actors as if they bear responsibility for model calibration, data representativeness, or post-deployment monitoring mistakes formal discretion for genuine control and converts negligence doctrine into a vehicle for scapegoating rather than efficient risk reduction.[67]

Reframed in this way, negligence doctrine no longer asks simply whether the last human actor behaved reasonably in isolation. Instead, it asks whether each relevant actor exercised reasonable care commensurate with its governance capacity over algorithmic risk. This reinterpretation aligns negligence analysis with the Calabresi—Hirschoff insight that liability should follow the best cost–benefit decision-maker, rather than reflexively attaching to the most visible human at the point of harm. In algorithmic systems, reasonable care is thus inseparable from system design, deployment architecture, and ongoing risk governance—and negligence law must evolve accordingly.

B. Product and Strict Liability After Software: When Algorithms Are the Defect

BARG is not meant to displace modern product liability. Rather, it explains when and why product-liability reasoning should be directed toward upstream algorithmic risk governors. In many AI cases, existing doctrines of design defect, failure to warn, enterprise liability, and strict liability already provide the doctrinal vocabulary for shifting responsibility toward actors that design, market, deploy, or profit from scalable risk-generating systems. The difficulty is not that product liability lacks relevance. The difficulty is that, in algorithmic ecosystems, courts must identify which actor actually performs the product-like governance function: defining the system's

[66] Deimantė Rimkutė, *supra* note 33, at 66–71; Clara Cestonaro et al., *supra* note 1, at 3–5, 8–10; Madalina Busuioc, *supra* note 7, at 828–834; Xukang Wang et al., *supra* note 3, 3–7, 9–10.
[67] Natalie Sheard, *supra* note 3, at 625–630, 633–638; Clara Cestonaro et al., *Id*, at 3–4, 6–10; George Maliha et al, *supra* note 1, at 632–634, 638–640.

baseline risk profile, controlling the data and architecture, monitoring performance, issuing warnings, updating the system, and spreading residual losses.

Put differently, product liability supplies the doctrinal form; BARG supplies the sorting principle. Product doctrine asks whether a product was defectively designed, inadequately warned against, or placed into the stream of commerce under conditions that justify enterprise responsibility. BARG helps courts determine where that inquiry should attach when the "product" is not a static object but a dynamic software-and-data system distributed across developers, platforms, vendors, and institutional deployers. The point is therefore not to replace product liability with a new label, but to prevent product-liability analysis from becoming trapped by formal categories or by the visible downstream user. When algorithmic design choices generate systemic risk, product-liability reasoning should follow the actor with functional control over those choices.

Debates over whether AI systems should be characterized as "products" or "services" have intensified, particularly in light of emerging European initiatives and AI-specific liability proposals. Yet from a BARG-oriented perspective, this formal categorization question is less important than a functional one: where is systemic design risk located, and which actor is best positioned to internalize and govern it?[68]

When an AI system operates as a standardized, scalable artifact—such as a medical diagnostic device, an autonomous driving control system, or a prepackaged credit-scoring model—the analogy to classic product liability becomes analytically powerful. In such contexts, strict or product liability directed at the BARG can serve its traditional efficiency function: internalizing systemic design risk in the hands of the actor who defines the system's baseline error profile and who can spread residual losses across users through pricing and insurance. Assigning liability at that level does not merely compensate victims; it incentivizes upstream redesign, recalibration, and safer deployment across all future cases.[69]

---

[68] J.K.C. Kingston, *Artificial Intelligence and Legal Liability*, in *Artificial Intelligence and Law*, 5–7 (Springer 2016); Sundaraparipurnan N. & Mark Potkewitz, *A Risk-Based Approach to Assessing Liability Risk for AI-Driven Harms Considering EU Liability Directive*, 3–8 (2024); also Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 4–6, 12–13; Aagaard Lise, *supra* note 30, at 125–126.

[69] Sundaraparipurnan N. & Mark Potkewitz, *Id*, at 4–8, 10–14; J.K.C. Kingston, *Id*, at 5–7, 8–11; Aagaard Lise, *Id*, at 125–126.

Within this framework, traditional product doctrines—especially design-defect and failure-to-warn analysis—require reinterpretation rather than abandonment. In algorithmic environments, "design" encompasses not only physical components, but model architecture, training data composition, feature selection, objective functions, risk thresholds, and robustness testing practices. A poorly calibrated confidence threshold, systematically unrepresentative training data, or the absence of stress testing across relevant subpopulations may constitute the functional equivalent of a design defect. Similarly, the failure to disclose known limitations, degradation risks, or population-specific performance gaps can be reframed as a failure-to-warn in a data-centric context. These doctrinal tools are already available; what changes is the object of scrutiny.[70]

The dynamic character of many AI systems complicates but does not undermine the product-liability analogy. Unlike static manufactured goods, algorithmic systems are frequently updated, retrained, and recalibrated after deployment. Strict liability in this setting may therefore need to be coupled with ongoing duties of monitoring, updating, and communicating material changes in system behavior. Where risk evolves over time—through model drift, feedback loops, or distribution shifts—the relevant "defect" may lie not in the original release alone, but in inadequate post-deployment governance. Product liability, properly adapted, can accommodate this dynamism by recognizing continuing obligations tied to control over updates and system performance.[71]

In many practical settings, the BARG will be identifiable as a manufacturer, software provider, or integrated platform that exercises design authority, data control, and systemic leverage over deployment conditions. Existing product-liability frameworks therefore offer a natural starting point for allocating responsibility—provided they are recalibrated to address software- and data-centric harms rather than purely physical malfunctions. The key insight is that when algorithms themselves embed the risk-generating design choices, the algorithm can be the defect in a legally meaningful sense, and liability should follow the actor who governs that design.[72]

---

[70] K.C. Kingston, *Id*, at 5–11; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 4–8, 15–16; Aagaard Lise, *Id*, at 125–126.
[71] Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 5–6, 10–14; Katherine Drabiak, *supra* note 17, at 8–9, 12–14; George Maliha et al, *supra* note 1, at 638–641.
[72] Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 4–8, 15–16; Aagaard Lise, *supra* note 30, at 125–126; J.K.C. Kingston, *supra* note 68, at 5–7, 10–12.

C. Who Is the Employer When the Algorithm Decides? Institutional and Vicarious Liability Reframed

As algorithmic systems increasingly organize and structure everyday professional activity, the traditional premise underlying employment-based liability—that employees function as autonomous decision-makers operating under human supervision—becomes progressively unstable. Across a growing range of domains, frontline actors no longer originate the relevant judgment but instead implement outputs generated elsewhere: by models designed, calibrated, and continuously updated at institutional or upstream levels. Physicians rely upon diagnostic scoring systems, loan officers operationalize automated credit determinations, and platform workers execute routing or pricing directives embedded in software architectures. In such environments, human actors frequently operate less as independent evaluators and more as executors of algorithmic instructions.[73]

This transformation exposes a structural tension within classical doctrines of institutional and vicarious liability. Respondeat superior presumes a hierarchical chain of supervision in which responsibility flows upward from identifiable employee negligence. Yet algorithmic governance often redistributes functional authority away from human supervisors and toward socio-technical systems embedded within institutional workflows. When decision thresholds, escalation pathways, and performance expectations are encoded directly into interfaces or automated feedback mechanisms, the meaningful locus of supervision may no longer correspond to formal managerial hierarchy.

Courts should therefore reconsider the tendency to conceptualize institutions merely as derivative bearers of liability for employee misconduct. Hospitals, digital platforms, insurers, and large employers increasingly operate as primary BARGs, exercising system-level authority over how algorithmic risk is generated, distributed, and constrained. Their responsibility should not depend exclusively upon identifying downstream negligence by individual staff members, but rather upon evaluating institutional governance choices concerning procurement, configuration, monitoring, and integration of algorithmic systems into operational decision-making.

[73] Madalina Busuioc, *supra* note 7, at 828–830, 832–833; Natalie Sheard, *supra* note 3, at 618, 622–628, 633–635.

In certain contexts, the algorithm itself functionally assumes characteristics traditionally associated with supervision. Decision architectures may constrain discretion through confidence thresholds, automated escalation rules, or performance metrics that discipline deviation from algorithmic output. A clinician instructed not to override diagnostic confidence scores, or a platform worker evaluated through automated productivity analytics, operates within a chain of control that is algorithmic rather than purely hierarchical. These developments raise a doctrinal question insufficiently addressed by existing vicarious liability frameworks: whether responsibility should continue to follow formal employment relationships alone, or instead track the locus of algorithmic control that meaningfully structures conduct.[74]

Reframing institutional liability around the BARG restores coherence to tort doctrine in algorithmic environments. When an institution elects to embed and rely upon AI systems as integral components of its workflow, it assumes governance authority over a risk-generating socio-technical environment. That authority encompasses decisions regarding vendor selection, calibration choices, override discretion, auditing practices, and post-deployment monitoring—decisions that shape risk distributions across thousands of interactions rather than a single encounter. Liability aligned with those governance capacities therefore directs legal incentives toward actors capable of implementing precaution at scale.

From a cost–benefit perspective, institutional actors uniquely possess the informational and organizational capacity necessary to govern algorithmic risk. They control access to aggregated performance data, maintain contractual leverage over developers and vendors, and possess the operational ability to suspend deployment, recalibrate thresholds, or redesign workflows in response to emerging harms. Unlike individual employees, they can transform lessons drawn from isolated incidents into system-wide preventive redesign affecting entire populations of users. Recognizing institutions as primary BARGs thus does not merely expand liability exposure; it aligns tort responsibility with the actors capable of reducing systemic risk.

Accordingly, when institutions choose to operationalize AI systems within decision-making processes, tort law should impose obligations commensurate with that system-level control. Liability grounded in BARG principles reflects not a doctrinal rupture but a principled extension

[74] Pratik Shukla, *Vicarious Liability or Liability for the Acts of Others in Tort: A Comparative Perspective*, 5 INT'L J. for Multidisciplinary Rsch. 1, 4–6 (2023); Madalina Busuioc, *Id*, at 825. 828, 832–833.

of institutional responsibility into algorithmically mediated environments, ensuring that responsibility follows governance authority rather than formal proximity to the moment of harm.[75]

D. Tort Law and AI Regulation: Complementarity, Not Displacement

If algorithmic harm is best understood as a problem of ongoing governance rather than episodic misconduct, then the question confronting tort law is not whether regulation replaces liability, but how the two systems should interact once responsibility has been relocated upstream along the liability map. The rapid emergence of dedicated AI regulatory regimes—most prominently the European Union's AI Act, the proposed AI Liability Directive, and an expanding body of sector-specific governance guidance—reflects a parallel institutional recognition that algorithmic risk must be managed ex ante through structured obligations imposed across the AI value chain, particularly for systems classified as "high risk".[76] In this respect, contemporary regulation increasingly targets the same institutional actors identified by the BARG framework: developers, large deployers, and platform operators whose design, calibration, and monitoring decisions shape risk distributions long before individual injuries materialize.

The expansion of regulatory governance, however, does not render tort law redundant. Rather, it clarifies tort law's distinctive institutional role. Regulatory regimes establish baseline safety through standardized mechanisms—conformity assessments, documentation duties, transparency requirements, and monitoring protocols—designed to prevent foreseeable harms before deployment. Tort law operates differently. It evaluates institutional competence under conditions of uncertainty and hindsight, responding to failure modes that regulation cannot fully anticipate, to deployment environments that evolve beyond certification assumptions, and to organizational incentives that convert nominal safeguards into ineffective practice.

From a BARG perspective, treating regulation as a substitute for tort liability would therefore risk reproducing the very distortions identified throughout this Article. Regulatory compliance may inform the content of reasonable care, but it cannot automatically negate it. Algorithmic systems are dynamic socio-technical arrangements rather than static products. Performance drift, feedback

[75] Aagaard Lise, *supra* note 30, at 125–126; Natalie Sheard, *supra* note 3, at 620–622, 633–635; George Maliha et al, *supra* note 1, at 630–634, 637–639.
[76] Philipp Hacker et al., *supra* note 18, at 1118–1120; Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 4–6, 10–14 ; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 2–6, 17–18.

loops, changing data environments, and workflow pressures may generate foreseeable risks even where formal regulatory benchmarks have been satisfied. If compliance were treated as a categorical shield, minimum regulatory standards would risk becoming ceilings on responsibility rather than floors for continuous governance.

At the same time, the institutional orientation of contemporary AI regulation provides an opportunity to calibrate tort incentives more precisely. Where actors functioning as Best Algorithmic Risk Governors demonstrate sustained governance capacity—through transparent risk assessments, meaningful auditing, continuous monitoring, and credible incident-reporting mechanisms—carefully structured safe harbors may be normatively justified.[77] Such protections can mitigate excessive deterrence that might otherwise discourage socially beneficial innovation or incentivize withdrawal from high-stakes applications where experimentation is socially valuable. Crucially, however, safe harbors should attach not to formal compliance alone, but to demonstrable institutional practices capable of detecting and correcting emerging risks across populations and over time. In this sense, protection follows governance performance rather than certification status.

The converse implication follows with equal force. Regulatory violations by BARGs should weigh heavily in negligence and strict-liability analysis, not merely as evidence of noncompliance but as markers of governance failure by actors uniquely positioned to internalize and operationalize regulatory knowledge.[78] Obligations directed toward developers, platforms, and institutional deployers function as operational signals about how algorithmic risk must be monitored, documented, and mitigated. When actors possessing system-level visibility and intervention capacity disregard those signals, the resulting lapse reflects not an isolated mistake but a breakdown in institutional competence at the precise point where tort law expects optimization decisions to occur.

Aligning tort doctrine and regulatory governance around BARGs therefore mitigates two symmetrical dangers that increasingly characterize AI ecosystems. On one side lies the over-deterrence of frontline users—physicians, drivers, clerks, and other end-users—who appear

[77] Chen Chen et al., *supra* note 18, at 13–14, 52–53; Adriano Koshiyama et al., *supra* note 21, at 2–7, 27–29; Richard J. Tong et al., *supra* note 48, at 4–6, 8.
[78] Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 4–6, 10–14; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, 2–5, 8–11.

proximate to harm yet lack meaningful informational or organizational control over system risk. On the other lies the under-deterrence of upstream decision-makers whose design choices, deployment constraints, and update practices structure harm across populations and across time. Complementarity between tort and regulation offers a path between these extremes. Properly aligned, regulation establishes the vocabulary of risk governance, while tort law supplies the adaptive pressure that ensures those obligations remain operational rather than symbolic. Accountability thus follows governance leverage rather than mere proximity to injury, allowing tort law to function not as a rival to AI regulation but as its dynamic partner in shaping responsible algorithmic systems.

E. Insuring the Algorithmic Society: Risk Distribution as Governance Architecture

Insurance has long occupied a central place in Calabresi's taxonomy of accident costs, particularly in the domain of secondary cost spreading. Yet in AI-driven environments, insurance performs a function that exceeds classical loss distribution. It becomes a governance multiplier—an institutional mechanism that translates tort liability into continuous oversight of socio-technical systems.

Insurance markets are already reacting to AI-related risks in healthcare, autonomous vehicles, financial scoring systems, and cyber infrastructures.[79] What distinguishes algorithmic risk, however, is not simply technological novelty but its structural character. Harm is produced not episodically but through scalable architectures of design, deployment, and iterative updating. Under these conditions, insuring individual users is analytically incomplete. The economically salient question becomes: who governs the system that generates risk across cases?

The BARG framework supplies the answer. When underwriting aligns with Best Algorithmic Risk Governors, insurance follows governance leverage rather than formal role. This alignment enables three interlocking functions.

First, pricing becomes informational governance. Premium differentiation tied to auditing capacity, drift detection protocols, incident reporting systems, and population-level monitoring transforms actuarial calculation into an incentive for institutional competence. Actors capable of

[79] Jack Boeglin, *supra* note 39, at 176, 193–194, 197–198, 200–202; Katherine Drabiak, *supra* note 17 at, 1–2, 8–11, 13–14; George Maliha et al, *supra* note 1; Sundaraparipurnan N. & Mark Potkewitz, *Id*, at 1–2, 10–14.

observing and recalibrating systemic risk face financial signals that reward sustained governance rather than minimal compliance.

Second, AI-specific liability products can embed structured governance incentives directly into coverage design. Policies directed toward developers, platforms, and large institutional deployers may condition favorable terms on demonstrable oversight practices—independent algorithm audits, documentation of threshold calibration, fairness testing across subpopulations, and transparent retraining procedures.[80] In this way, insurance operates as a private enforcement layer reinforcing tort doctrine's orientation toward upstream risk control.

Third, insurers increasingly function as cross-institutional observers of failure patterns. By aggregating claims experience across sectors, insurers may detect recurring error modes—bias amplification, model drift, inadequate override design—that remain locally invisible.[81] This aggregation capacity positions insurers as systemic risk intermediaries rather than passive indemnifiers.

Routing premiums and loss exposure through BARGs therefore does more than distribute loss. It embeds tort law's cost-internalization logic within a feedback loop of monitoring, pricing, and redesign. Where liability identifies the proper governor of risk, insurance sustains that identification over time. In dynamic algorithmic systems—where harm evolves through updates, feedback loops, and scaling effects—this sustained governance function is indispensable.

Properly structured, insurance thus serves as a bridge between episodic adjudication and continuous institutional adaptation. It amplifies tort law's preventive logic by linking financial exposure to demonstrable governance capacity. In the algorithmic society, insurance is not merely a mechanism of compensation; it is an architecture of accountability.

VII. Conclusion: Rewiring Tort Law for Algorithmic Governance

A. The BARG Turn: Restating the Core Move

Artificial intelligence exposes a growing tension at the center of modern tort theory. Continued reliance on the Cheapest Cost Avoider paradigm as the primary organizing principle of liability

[80] Adriano Koshiyama et al., *supra* note 21, at 3–7, 22–26; Richard J. Tong et al., *supra* note 48, at 2–6, 8.
[81] Susan Hao et al., *supra* note 4, at 3–8; Adriano Koshiyama et al., *Id*, at 4–7, 5–17, 22, 28.

risks directing judicial attention toward actors who are visible at the moment of harm but structurally incapable of governing the risks that produced it. The CCA framework emerged in relatively bounded accident settings in which prevention capacity and decision-making authority frequently converged in a single actor. Algorithmic environments disrupt that alignment. Harm increasingly arises from layered socio-technical systems in which design, training, deployment, monitoring, and institutional integration are temporally and organizationally dispersed. In such settings, focusing liability analysis on the last human in the loop encourages courts to fixate on proximity rather than governance, overlooking the complex, systemic, and data-driven structures through which algorithmic risk is actually produced and can be most efficiently reduced.[82]

The persistence of this downstream focus remains doctrinally understandable. Physicians, drivers, clerks, and other frontline users fit comfortably within familiar negligence narratives: they occupy identifiable roles, exercise apparent discretion, and stand closest to the injury. Yet visibility does not reliably track control. In algorithmic ecosystems, the most consequential safety decisions are embedded upstream—in architectural design choices, data curation practices, calibration thresholds, interface constraints, and institutional deployment policies that silently shape how risk is distributed across populations and over time.

To remain economically rational and normatively credible under these conditions, tort law must adjust its analytic lens. The relevant inquiry is no longer merely who could have prevented a particular accident at the lowest immediate cost, but which actor is best positioned to gather information about algorithmic risk, evaluate competing precautionary strategies, and implement interventions capable of reducing harm across cases. This move does not abandon the economic logic underlying the CCA or the best decision-maker framework. Rather, it extends that logic to environments characterized by opacity, scale, and continuous updating.

The Best Algorithmic Risk Governor captures this shift. The BARG is the actor best positioned to observe systemic risk patterns, internalize long-term accident costs, and recalibrate socio-technical systems in response to emerging harms. Directing liability toward such actors allows tort law to preserve its deterrence and loss-allocation functions while avoiding the systematic over-attribution

[82] Clara Cestonaro et al., *supra* note 1, at 9–10; Muhammad Uzair, *supra* note 2, at 1–3, 12–15; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 671–672, 677–692; Laura Weidinger et al., *supra* note 48, at 6–12, 17–20; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 5–11; George Maliha et al, *supra* note 1, at 630–640.

of responsibility to downstream human actors whose apparent discretion masks structural constraint—and whose proximity to harm too often substitutes for genuine governance authority.

B. Liability as Governance Infrastructure: The Normative Payoff

Reorienting tort liability around the BARG yields an important normative payoff. Once liability is aligned with the actors who actually govern algorithmic risk, tort law ceases to function merely as a retrospective device for allocating losses and instead begins to operate as governance infrastructure for AI systems.

First, locating liability at the level of BARGs strengthens incentives for safe AI design and deployment. When responsibility attaches to the actors who control model architecture, training data, deployment parameters, and institutional governance structures, legal incentives are directed toward the points where systemic risk is created and can most efficiently be reduced. Developers, platforms, and large institutional deployers are thus induced to internalize the long-term costs of algorithmic failures and to invest in robustness, fairness, transparency, and meaningful post-deployment oversight. Liability in this configuration does not merely respond to harm; it structures the incentive environment within which algorithmic systems are designed and governed.[83]

Second, the BARG framework prevents the systematic over-attribution of responsibility to frontline human actors. Physicians, drivers, clerks, moderators, and similar operators often appear—at the level of a single incident—to be the cheapest cost avoiders. Yet their apparent control is frequently superficial. They neither design the model nor determine its training data, calibration thresholds, or institutional deployment constraints, and they typically lack the information necessary to evaluate the system's reliability across contexts. Treating these actors as the primary locus of liability therefore misdirects legal incentives while producing little systemic improvement. A BARG-oriented approach instead aligns liability with those actors who possess the informational and organizational capacity to govern algorithmic risk across cases.[84]

---

[83] Miriam Buiten, Alexandre de Streel & Martin Peitz, *Id*, at 8–11; George Maliha et al, *Id*, at 630–637; Adriano Koshiyama et al., *supra* note 21, at 2–6, 10–16, 21–28; Madalina Busuioc, *supra* note 7, at 829–834.
[84] Muhammad Uzair, *supra* note 2, at 1–8, 12–17; Natalie Sheard, *supra* note 3, at 621–635; Clara Cestonaro et al., *supra* note 1, at 2–4, 9–10; George Maliha et al, *Id*, at 630–635, 638–639.

Third, focusing liability on BARGs improves the legal system's ability to address harms that are structural rather than episodic. Algorithmic discrimination illustrates this dynamic clearly. Disparate outcomes in credit scoring, hiring, insurance underwriting, and other algorithmic domains rarely arise from isolated deviations in individual decision-making; they emerge from patterns embedded in data selection, model objectives, and deployment practices. Effective mitigation therefore depends on population-level monitoring, auditing, and recalibration—functions typically performed by developers, platforms, and institutional deployers rather than by individual decision-makers at the point of application. Locating liability at the level of algorithmic governance thus enables tort law to respond to systemic harms while encouraging institutions to detect and correct discriminatory patterns before they crystallize into widespread injury.[85]

The broader implication is that artificial intelligence need not be treated as an external disruption to tort doctrine. Properly structured, tort liability can shape the trajectory of AI integration itself. By directing legal incentives toward actors capable of governing algorithmic systems, tort law helps steer technological development toward configurations that are safer, more accountable, and more consistent with public values.

C. The Next Map: Evidence, Institutions, and Multi-BARG Futures

Calabresi's original insight—that accident law should focus on the actor best positioned to reduce the costs of harm—was never meant to be a static formula. It was a method for adapting legal responsibility to changing technological and institutional conditions. The BARG framework continues that project. By redirecting attention toward the Best Algorithmic Risk Governor, it offers a way to translate the logic of the cheapest cost avoider into the architecture of contemporary algorithmic systems. Yet identifying the relevant BARG is only the starting point. The next challenge is institutional and doctrinal: adapting procedural mechanisms, evidentiary rules, and responsibility doctrines so that tort law can operate effectively in technologically complex environments. Several directions for further inquiry follow from this shift.

One such direction concerns the problem of mass harms. Many AI-related injuries do not arise as isolated accidents but as systemic effects experienced simultaneously by large populations.

---

[85] Savina D. Kim, Stefan Lessmann, Galina Andreeva & Michael Rovatsos, *supra* note 26, at 2–4, 11–13; Xukang Wang et al., *supra* note 3, at 1–7, 9–10; Solon Barocas & Andrew D. Selbst, *supra* note 3, 671–672, 677–692, 717–719; Madalina Busuioc, *supra* note 7, at 825–826, 831–834.

Algorithmic systems used in domains such as credit scoring, employment screening, pricing, and content moderation can produce patterns of harm that are individually modest yet collectively substantial. In such circumstances, individualized litigation risks obscuring the structural nature of the harm and diffusing accountability. BARG-based liability therefore raises an important institutional question: how should responsibility be operationalized where algorithmic harms are widely distributed? Mechanisms such as class actions, collective redress procedures, and forms of public enforcement may become essential complements to the BARG framework, enabling courts and regulators to identify and discipline the relevant risk governors at scale rather than through fragmented individual claims.[86]

A second direction concerns the role of the state. Governments increasingly deploy AI systems in areas including policing, welfare administration, taxation, and immigration control. These deployments raise distinctive questions about how the BARG concept should operate in the public sector. In some cases, the relevant BARG may be the public authority that decides to adopt and structure the system; in others, it may lie with private actors responsible for design, training data, or operational architecture. Applying BARG analysis in this domain inevitably intersects with doctrines of public law, including sovereign immunity, administrative accountability, and constitutional limits on governmental power. Clarifying when public bodies themselves should be treated as BARGs—and how responsibility should be distributed between public institutions and their technological partners—remains a central challenge for the emerging law of algorithmic governance.[87]

A third direction concerns proof and causation. The epistemic structure of many AI systems complicates the operation of traditional evidentiary doctrines. Black-box models, probabilistic decision processes, and complex technological supply chains may make it difficult for injured parties to establish how a particular algorithmic output produced a legally cognizable harm. Without doctrinal adaptation, this informational asymmetry risks undermining deterrence by insulating the relevant BARG from effective liability. Courts may therefore need to reconsider evidentiary frameworks in ways that align informational access with responsibility. Instruments

[86] Aurora S. Zhang & Anette E. Hosoi, *supra* note 25, at 1014–1018, 1020–1023; Xukang Wang et al., *Id*, at 1–2, 4–8, 10; Solon Barocas & Andrew D. Selbst, *Id*, at 673–675, 684–687, 691–693, 701–714; Madalina Busuioc, *Id*, at 825–827, 823–833 .

[87] Annette Zimmermann & Chad Lee-Stronach, *Proceed with Caution*, 52 CAN. J. PHIL. 6, 6–9, 19–22 (2022); Madalina Busuioc, *Id,* at 826–830, 832–834.

such as reversed burdens of proof, evidentiary presumptions, and duties of disclosure imposed on actors controlling the relevant technological infrastructure may help restore that alignment. Integrating the BARG framework with such evidentiary innovations represents an important frontier for both scholarship and doctrinal development.[88]

Finally, many AI systems operate within layered technological ecosystems involving multiple actors distributed across global supply chains. Model developers, platform operators, downstream deployers, and regulators may each exercise partial control over the creation and management of algorithmic risk. In such environments, responsibility may not reside in a single actor but in a constellation of overlapping BARGs. The challenge, therefore, is not merely to identify the actor best positioned to govern risk, but to develop doctrines capable of allocating responsibility among multiple such actors operating at different stages of the technological pipeline. Crafting coherent principles for these multi-BARG environments—particularly in cross-border contexts—will be essential as AI systems become increasingly integrated into global infrastructures of production and governance.[89]

These lines of inquiry ultimately return tort theory to the broader ambition that animated Calabresi's work. The goal was never to freeze accident law within a single doctrinal rule, but to equip courts and lawmakers with a conceptual vocabulary capable of evolving alongside technological change. In the age of artificial intelligence, that vocabulary must now include the Best Algorithmic Risk Governor.

If courts accept this invitation, tort law will do more than assign liability after algorithmic harms occur. It will help structure the incentives that shape how AI systems are designed, deployed, and governed in the first place. By directing responsibility toward the actors best positioned to anticipate and control algorithmic risks, the BARG framework can help build a legal and

---

[88] David Fernández Llorca, Vicky Charisi, Ronan Hamon, Ignacio Sánchez & Emilia Gómez, *Liability Regimes in the Age of AI: A Use-Case Driven Analysis of the Burden of Proof*, 76 J. ARTIFICIAL INTELLIGENCE RSCH. 613, 616–617, 620–622, 630–631 (2023); Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 1–4, 7–9; Clara Cestonaro et al., *supra* note 1, at 3–4, 9–10; Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 1, 3. 5–6, 9, 12–13.

[89] Gabriel Lima & Meeyoung Cha, *supra* note 20, at 2–3; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *Id*, at 1–3, 7–8; Philipp Hacker et al., *supra* note 18, at 1115–1117, 1120; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 5–6, 11–13.

economic environment in which the development of AI is continuously recalibrated in light of human safety, fairness, and dignity.

In the algorithmic age, the central question of accident law remains the one Calabresi posed half a century ago: who is best positioned to govern risk? The BARG framework offers a contemporary answer.

CORPORATE LIABILITY IN THE AGE OF ARTIFICIAL INTELLIGENCE

Fujiao Xie, Ph.D.*

I. INTRODUCTION

Artificial intelligence ("AI") represents one of the most significant transformations in the structure of corporate decision-making since the emergence of the modern business corporation. Algorithmic systems increasingly determine pricing strategies, allocate investment capital, conduct securities trading, evaluate employee performance, manage supply-chain logistics, and shape consumer engagement across global markets. Unlike earlier technological innovations, contemporary machine-learning systems possess adaptive capabilities that enable them to modify operational behavior in response to evolving data environments. This capacity for autonomous adjustment complicates foundational assumptions embedded in corporate liability doctrine—particularly the premise that legally relevant corporate conduct can be attributed to identifiable human decision-makers exercising conscious judgment.

Traditional corporate law evolved within an industrial economy characterized by hierarchical governance structures and deterministic technologies. Liability doctrines such as respondeat superior, fiduciary duty, and negligence presupposed that harmful outcomes could be traced to discrete managerial decisions or organizational failures.[1] AI systems disrupt this framework by introducing probabilistic reasoning processes, opaque algorithmic architectures, and distributed responsibility across engineers, executives, data scientists, and automated systems themselves.[2] As corporate reliance on AI deepens, courts will increasingly confront disputes in which causal chains linking governance decisions to operational harm are technologically mediated and analytically complex.

This Article advances a new theoretical and doctrinal model for addressing corporate liability in the algorithmic era. It argues that while existing fault-based doctrines remain necessary, they are insufficient to address systemic risks generated by high-impact AI deployment. Instead, courts and regulators should adopt an enterprise risk internalization framework grounded in a rebuttable presumption of corporate liability for algorithmic harms. Under this model, corporations deploying autonomous or semi-autonomous decision systems must demonstrate implementation of structured governance safeguards to avoid liability exposure.

By integrating fiduciary oversight obligations with enterprise liability principles derived from tort and securities law, the proposed framework aligns accountability with institutional capacity to manage technological risk. The Article proceeds by examining conceptual tensions in attribution doctrine, analyzing evolving fiduciary oversight jurisprudence, evaluating liability exposure

* Associate Professor of Accounting, Murray Koppelman School of Business at Brooklyn College of the City University of New York

across doctrinal domains, and situating emerging governance models within comparative regulatory developments.

II. THE LIMITS OF TRADITIONAL ATTRIBUTION DOCTRINE

Corporate liability depends upon doctrinal mechanisms that connect organizational conduct to legal responsibility. Agency law permits attribution of employee misconduct when actions occur within the scope of employment.[3] Negligence principles similarly impose liability where corporations fail to implement reasonable safeguards against foreseeable harm.[4]

AI systems complicate these frameworks by mediating decision authority through complex socio-technical architectures. Algorithmic harms may result from interactions among training datasets, model design choices, environmental inputs, and continuous learning processes.[5] In such contexts, identifying a single culpable actor may be both practically difficult and conceptually misleading.

Scholarly responses have diverged along three principal lines. First, some commentators argue that AI should be treated as a sophisticated instrumentality analogous to industrial machinery.[6] Second, others propose recognizing algorithmic systems as functional agents capable of generating independent liability implications.[7] Third, an emerging body of scholarship emphasizes enterprise risk allocation, contending that corporations deploying AI for profit maximization should bear responsibility for resulting externalities regardless of fault.[8]

Courts confronting AI-related disputes must therefore balance doctrinal continuity with the need for adaptive legal frameworks capable of addressing technologically mediated risk.

III. FIDUCIARY DUTY AND THE EVOLUTION OF TECHNOLOGICAL OVERSIGHT

A. Caremark and the Rise of Mission-Critical Risk Governance

Delaware courts have increasingly emphasized that directors must implement monitoring systems capable of identifying mission-critical risks. In *In re Caremark International Inc. Derivative Litigation*, the Court of Chancery recognized that sustained failure to establish reporting mechanisms could constitute a breach of fiduciary duty.[9] Subsequent decisions have reinforced this principle. In *Marchand v. Barnhill*, the Delaware Supreme Court held that boards must maintain oversight structures tailored to core operational risks.[10]

AI governance may soon emerge as a paradigmatic mission-critical risk domain. Consider a corporation deploying autonomous logistics software that prioritizes efficiency metrics over safety constraints. Following a series of accidents, shareholders bring derivative litigation alleging that directors ignored internal warnings regarding algorithmic bias and system instability. Courts evaluating such claims may examine whether boards established technology risk committees, algorithmic audit frameworks, and escalation protocols for technological anomalies.

B. Algorithmic Reliance and the Business Judgment Rule

The business judgment rule protects directors who act in good faith with reasonable diligence.[11] However, blind reliance on algorithmic outputs without meaningful oversight could undermine this protection. Judicial expectations regarding technological literacy are therefore likely to evolve as AI adoption becomes widespread.

IV. LIABILITY EXPOSURE IN THE ALGORITHMIC CORPORATION

A. Tort and Product Liability

Negligence doctrine remains a central pathway for imposing liability on corporations deploying AI technologies. Firms may face claims alleging inadequate training-data validation, failure to monitor model drift, or insufficient cybersecurity safeguards.[12]

Strict product liability raises additional complexities. Courts must determine whether adaptive software constitutes a "product" and how defectiveness should be evaluated when performance evolves through learning processes.[13] Litigation involving automated driver-assistance systems illustrates these tensions.

B. Securities Disclosure

Public corporations must disclose material risks affecting financial performance.[14] As AI becomes central to corporate strategy, inadequate disclosure of algorithmic vulnerabilities may give rise to securities fraud litigation. Enforcement actions involving automated trading controls demonstrate that firms cannot avoid liability by attributing misconduct to autonomous software.[15]

C. Employment Discrimination

Automated hiring tools have generated scrutiny due to potential disparate impact on protected classes.[16] Employers deploying such technologies may face liability if they fail to conduct bias audits or provide meaningful human review mechanisms.

D. Antitrust

Algorithmic pricing systems may facilitate tacit coordination without explicit agreements among competitors.[17] Antitrust authorities increasingly examine whether firms deploying coordination-prone algorithms should bear responsibility for supracompetitive pricing outcomes.

V. ENTERPRISE RISK INTERNALIZATION MODEL

This Article proposes a rebuttable presumption of enterprise liability for harms arising from high-impact AI deployment. Plaintiffs would establish a prima facie case by demonstrating that algorithmic systems caused foreseeable harm within corporate operations.

Corporations could rebut liability by demonstrating implementation of governance safeguards, including:

- board-level technology oversight committees
- independent algorithmic auditing procedures
- documented bias testing protocols
- continuous monitoring and incident reporting systems
- transparent disclosure to regulators and investors

Certain ultra-hazardous AI applications—such as autonomous weapons systems or safety-critical transportation technologies—may warrant strict liability regardless of governance measures.

This model aligns legal accountability with corporate risk-internalization capacity while promoting innovation through governance safe-harbor incentives.

VI. COMPARATIVE REGULATORY TRAJECTORIES

The European Union has adopted a comprehensive risk-based regulatory framework imposing ex ante compliance obligations on high-risk AI systems.[18] The United States has pursued a decentralized enforcement strategy grounded in existing statutory authority exercised by agencies such as the Federal Trade Commission and Securities and Exchange Commission.[19]

The United Kingdom has adopted a principles-based regulatory model emphasizing innovation flexibility. Canada has proposed legislation targeting high-impact automated decision systems. China has implemented algorithm governance rules addressing recommendation systems and data security. These divergent approaches may converge toward hybrid governance-centered liability regimes as multinational corporations seek regulatory harmonization.

VII. CONCLUSION

Artificial intelligence is transforming the institutional foundations of corporate accountability. As algorithmic systems assume greater autonomy, doctrines grounded in human agency will face increasing conceptual strain.

An enterprise risk internalization framework offers a principled mechanism for adapting corporate liability to technological transformation. By aligning responsibility with institutional capacity to

manage systemic risk, courts and regulators can promote innovation while safeguarding investors, consumers, employees, and competitive markets.

The evolution of corporate liability in the AI era will ultimately reflect broader societal choices regarding technological power and economic governance. Structured governance-centered liability regimes represent a necessary step toward ensuring that innovation proceeds alongside meaningful accountability.

ENDNOTES

1. Restatement (Third) of Agency § 7.03 (Am. L. Inst. 2006).
2. Frank Pasquale, *The Black Box Society* (2015).
3. Burlington Indus., Inc. v. Ellerth, 524 U.S. 742 (1998).
4. Restatement (Third) of Torts § 7.
5. Ryan Calo, Artificial Intelligence Policy: A Primer and Roadmap, 51 U.C. DAVIS L. REV. 399 (2017).
6. Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (2015).
7. *Id.*
8. Mark A. Geistfeld, Strict Products Liability 2.0, 95 NOTRE DAME L. REV. 1325 (2020).
9. In re Caremark Int'l Inc. Derivative Litig., 698 A.2d 959 (Del. Ch. 1996).
10. Marchand v. Barnhill, 212 A.3d 805 (Del. 2019).
11. Aronson v. Lewis, 473 A.2d 805 (Del. 1984).
12. Andrew Tutt, An FDA for Algorithms, 69 ADMIN. L. REV. 83 (2017).
13. James A. Henderson & Aaron D. Twerski, *Products Liability* (8th ed. 2019).
14. Basic Inc. v. Levinson, 485 U.S. 224 (1988).
15. Commodity Futures Trading Comm'n v. Coscia, 866 F.3d 782 (7th Cir. 2017).
16. Griggs v. Duke Power Co., 401 U.S. 424 (1971).
17. Ariel Ezrachi & Maurice Stucke, *Virtual Competition* (2016).
18. Proposal for a Regulation Laying Down Harmonized Rules on Artificial Intelligence, COM (2021) 206 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
19. FTC, Algorithmic Fairness Statement (2023) https://www.ftc.gov/system/files/ftc_gov/pdf/2023190_commissioner_bedoya_riteaid_statement.pdf

BEYOND THE HIRED GUN: LAWYER JOKES, NDA ABUSE, AND THE PROMISE OF PURPOSE-DRIVEN LAW

Hershey H. Friedman, Ph.D.
Professor of Business
Department of Management, Marketing, and Entrepreneurship
Koppelman School of Business
Brooklyn College of the City University of New York
Email: x.friedman@att.net

Xianfang Zeng, PhD
Assistant Professor of Marketing
Department of Management, Marketing, and Entrepreneurship
Brooklyn College of the City University of New York
Email: Xianfang.Zeng@brooklyn.cuny.edu

Abstract

Lawyer jokes occupy a unique position in popular culture because they are remarkably violent, strangely normalized, and deeply revealing how society regards a profession it simultaneously fears, needs, and resents. This paper uses such humor as a lens to examine the ethical failures that have eroded public trust in the legal system, with particular attention to how nondisclosure agreements (NDAs) are weaponized to shield serial misconduct. Drawing on Freudian theories of tendentious humor, disparagement humor research, and the growing literature on conscious capitalism, the paper argues that the legal profession is overdue for a values-based transformation. It identifies collaborative law as one promising model for reorienting lawyers from "hired guns" into the healers of conflict envisioned by Chief Justice Burger and practiced by Gandhi. By situating this shift within the broader corporate move from shareholder primacy toward stakeholder accountability, the paper suggests that purpose-driven law is both a practical possibility and a growing demand from clients, employees, and society at large.

Keywords: lawyer jokes, professional ethics, nondisclosure agreements (NDAs), purpose-driven leadership, collaborative law, conscious capitalism, disparagement humor, legal reform

INTRODUCTION

Why are there no lawyer jokes? Because they are all true. This punchline is worth pausing on, not for its humor but for its sociology. Jokes are not random. They encode cultural grievances and social observations that are too uncomfortable for direct expression. Humor serves as a powerful lens through which cultural values, stereotypes, and power dynamics are revealed. Because jokes are more than mere entertainment, they often encode social attitudes and serve as revealing diagnostics of bias and belief systems. Listening to a person's humor can be highly instructive for those seeking to discern whether that individual harbors sexist, racist, or otherwise prejudiced tendencies. As Helmreich (2004) demonstrated, humor and anecdotes provide valuable insights into stereotypes and the subtle ways they are reinforced or challenged in everyday life. Similarly, Davies (2011) argues that jokes frequently tap into deeply embedded assumptions about race, gender, and class. This offers insight into what audiences are willing to laugh at and what they find objectionable. Humor can serve as a socially acceptable outlet for expressing biased attitudes. Such "disparagement humor" allows individuals to communicate prejudice under the guise of play, which can effectively normalize discriminatory worldviews. Beyond revealing individual attitudes, these jokes also reflect broader cultural narratives (Ford & Ferguson, 2004).

The psychological study of aggressive humor dates back to Sigmund Freud, who categorized jokes according to their underlying intent. He distinguished between innocent, non-tendentious wordplay and tendentious humor, which serves a specific psychological function (Freud, 1960, pp. 90-116). Within the tendentious category, Freud identified three primary types: obscene or exposing jokes, hostile or aggressive jokes, and cynical or blasphemous jokes (p. 115). This framework highlights that humor is often far from neutral, frequently acting as a vehicle for complex social and personal motives. Freud argued that hostile jokes function as a psychological tool to express aggression indirectly. By using wit as a mask, individuals can avoid direct conflict while still channeling suppressed impulses in a socially acceptable format. In this view, aggressive humor provides a calculated release for built-up psychological tension. It allows for the expression of hostility through a disguised medium, effectively transforming potentially destructive energy into a sophisticated form of social interaction.

Friedman and Friedman (2019) find that humor can be a meaningful tool for advancing social justice. This perspective emphasizes that everyone deserves fair treatment and equal opportunities, regardless of their background, ethnicity, religion, circumstances, or appearance. Through this lens, creators of social justice comics view humor as a transformative tool for addressing systemic imbalances. They aim to leverage wit to rectify social inequities while advocating for the empathetic treatment of every individual. This perspective positions comedic narratives as a deliberate means of social correction. By highlighting institutional flaws through satire or storytelling, these artists seek to promote a just and compassionate society for everyone. However, beyond its potential for social commentary or positive change, humor also frequently serves as a vehicle for expressing collective anxieties and criticisms. It is within this latter context, specifically through the prevalent genre of disparagement humor targeting lawyers, that this paper explores the profound implications for professional ethics and public perception.

This paper brings these strands together. First, it deems lawyer humor as a cultural barometer of public trust in law. Second, the paper examines how nondisclosure agreements (NDAs), arbitration clauses, and contract design have been weaponized as “machinery of silence” in ways that challenge traditional conceptions of lawyers’ ethical duties. Third, we suggest purpose-

driven law informed by developments in purpose with profit corporate forms and purpose-driven leadership. Finally, this paper considers collaborative and healing-oriented practices in law as emerging models of a profession that sees itself as a steward of values rather than merely a technician of rules.

LAUGHING AT THE BAR: WHAT LAWYER JOKES REVEAL ABOUT PUBLIC TRUST

Marc Galanter's classic study of lawyer jokes situates them as a "rich and time-honored genre" that tracks changing public attitudes toward the legal profession (Galanter, 2005). Jokes about specific professions, such as law, accounting, politics, or academia, reveal collective perceptions of those roles. Examining the content and circulation of such humor reveals how society constructs hierarchies of respectability and moral worth. Thus, studying the jokes told about various professions provides a unique perspective on public attitudes and the symbolic boundaries of social groups. Jokes about other professionals mock certain aspects of the profession. Thus, accounting jokes play into the stereotype that accountants are dull, insipid, and unoriginal.

Note the following two famous jokes:

> A man takes a balloon ride at a local country fair. A fierce wind suddenly kicks up, causing the balloon to violently leave the fair and carry its occupant out into the countryside. The man has no idea where he is, so he goes down to five meters above ground and asks a passing wanderer, "Excuse me, sir, can you tell me where I am?"
> The passer-by says, "You are in a downed red balloon, five meters above ground."
> The balloon's unhappy resident replied, "You must be an economist.
> "How could you possibly know that?" asked the passer-by.
> "Because your answer is technically correct but absolutely useless, and I am still lost."
> "Then you must be in management," said the passer-by.
> "That's right! How did you know?"
> "You have a great view from up there, and yet you don't know where you are, and you don't know where you're going. The fact is you're in the exact same position you were in before we met, but now your problem is somehow my fault!"
>
> A businessman, a doctor, and an engineer were all sentenced to die in the electric chair. The warden explained the rules, stating that the law allows only one execution attempt. If the machine fails to work, the prisoner is set free. The businessman was strapped in first. When asked for a final request, he shook his head. The executioner flipped the switch, but nothing happened. True to the law, the guards released him and let him walk away. Next, the doctor took his seat. He declined a final request and waited. Again, the switch was thrown, the machine remained silent, and the doctor was allowed to go free. Finally, the engineer was strapped into the chair.

> The warden asked if he had any last words. The engineer looked up at the ceiling, squinted at the wiring, and pointed a finger. He said, "Do you see the wire up there? You should probably fix that loose ground wire near the transformer if you want it to work properly."

Lawyer jokes occupy a peculiar place in popular culture, distinguished by their extreme and often violent humor. A striking number of them imagine scenarios of mass harm or even extermination directed at lawyers, reflecting a deep-seated cultural ambivalence toward the profession. This intensity may help explain why Lynch and Friedman (2013) suggest that lawyer jokes could serve as an effective pedagogical tool in teaching business ethics. Few, if any, other professions are the subject of so many jokes in which the punchline often involves their collective destruction, revealing both society's fascination with and resentment toward those who practice law. A significant portion of lawyer jokes relies on the trope of professional misconduct and a lack of integrity. Jokes about “600 lawyers at the bottom of the sea” are not just playful hostility; they condense anxieties about the “ubiquitous place” of law in modern life and the “legalization” of everyday conflicts (Galanter, 2005).

> What do you call 600 dead lawyers at the bottom of the ocean?
> Answer: A good start! [Note that this joke does not work with any profession other than lawyers. Try it with nurses or engineers.
>
> The trouble with the legal profession is that 99% of its members give the rest a bad name.
>
> A woman and her little girl were visiting the grave of the little girl's grandfather. On their way back to the car, the little girl asked, "Mommy, do they ever bury two people in the same grave?"
> "Of course not, dear," replied the mother. "Why would you think that?"
> "The headstone back there said... 'Here lies a lawyer and an honest man.’"
>
> Why do you bury lawyers 10 ft deep instead of 6?
> Because deep down, they really are good people.
>
> A lawyer dies and arrives at the Pearly Gates. St. Peter says, "We've been expecting you for quite some time." The lawyer is shocked and replies, "But I'm only 42 years old!" St. Peter consults his records and says, "That's impossible. According to your billable hours, you're at least 98!"
>
> What's the difference between a lawyer and a vulture? Vultures wait until you're dead to rip your heart out.
>
> "It is so dark in here. Why are the windows covered?" the lawyer muttered as he regained consciousness. "Purely for your peace of mind," the nurse whispered. "There’s a huge fire next door, and we didn't want you to assume you’d died and started your afterlife.

What do you throw to a drowning lawyer?
His partners.

During a high-stakes lawsuit, a young lawyer proposed sending a gift to the judge to influence the outcome. His mentor shut him down immediately, insisting the judge's integrity would lead to an automatic loss for anyone who tried to bribe him. After they won the case, the mentor praised the young lawyer for his restraint. The lawyer admitted, "I actually sent the gift anyway. I just signed my opponent's name to the card."

How many lawyers does it take to stop a speeding truck? Never enough.

Why did God create snakes before he made lawyers?
He needed the practice.

A woman is at the post office when she notices a man at the next counter meticulously sticking "Love" stamps onto hundreds of bright pink, heart-covered envelopes. After he finishes the stamps, he pulls out a bottle of expensive Chanel perfume and starts misting every single one. Unable to contain her curiosity, she walks over and asks, "I couldn't help but notice... that is the most romantic gesture I've ever seen! Are you some kind of modern-day Cyrano de Bergerac?" "Not exactly," the man says, without looking up. "I'm mailing out a thousand of these to married men all over the neighborhood. Inside, they all say, 'I love you—don't tell a soul.'" The woman is horrified. "That's heartless! Why on earth would you do that?" The man finally looks up and grins. "I'm a divorce lawyer. Business has been a little slow lately."

A man walked into a bar with his alligator and asked the bartender,
"Do you serve lawyers here?"
"Sure do," replied the bartender. "Good," said the man. "Give me a beer, and I'll have a lawyer for my alligator."

What's the difference between a dead lawyer on the road and a dead skunk on the road? There are skid marks in front of the skunk.

What do you call a lawyer up to his neck in cement? Not enough cement.

What do you call 22 skydiving lawyers?
Answer: Skeet.

Why are scientists now using lawyers instead of rats for their experiments?
There is no longer an endless supply of rats.
Researchers don't get emotionally attached to lawyers.
Ethics committees are more likely to approve inflicting pain on lawyers.

> There are simply some things even a rat won't do.
> One problem, however, remains: you can't extrapolate the test results to humans.

Violent lawyer jokes provide a revealing window into the moral ambivalence surrounding the legal profession and can therefore serve as a powerful entry point for teaching ethics. Building on Lynch and Friedman's (2013) insight that many lawyer jokes convey the idea that "the only good lawyer is a dead one," such humor often imagines mass death, extermination, or other forms of extreme violence directed at lawyers as an acceptable or even satisfying outcome. In contrast to jokes about most other professions, a substantial subset of lawyer jokes treats the wholesale elimination of lawyers as an intelligible punchline, which highlights the intensity of public anger and mistrust toward the bar.

Galanter (2005) finds that as societies rely more heavily on formal legal processes, ambivalence about lawyers grows: citizens value rights and legal protections yet fear opportunistic litigation, high fees, and procedural gamesmanship. From a pedagogical perspective, these jokes can be analyzed not for their shock value but as cultural artifacts that encode stereotypes about greed, dishonesty, and the perceived corruption of legal institutions, making them a vivid catalyst for discussion about professional responsibility and the social consequences of unethical conduct. When students examine why such jokes "work" and why they seem acceptable to many audiences, they are pushed to confront how humor normalizes hostility, how stereotypes about lawyers are constructed and maintained, and how ethical reform might alter the narrative that makes violent fantasies about lawyers seem amusing in the first place.

FROM PRIVACY TOOL TO WEAPON: NDAS AND THE MACHINERY OF SILENCE

NDAs were originally justified as tools to protect trade secrets and legitimate privacy interests, maintaining the confidentiality of sensitive negotiations, and preventing the disclosure of proprietary information by employees who have been entrusted with it. In practice, they have become central instruments in suppressing disclosure of harassment, discrimination, and other workplace misconduct (Altman, 2022; Grossi, 2025). NDAs play a role in the failure of individualized employment and equality law to combat abuse, normalizing silence and reinforcing workplace hierarchies (Barmes, 2023). Employers, “mediated by lawyers,” impose NDAs to protect serial wrongdoers and organizational reputations, while claimants—facing shame and career risk—often feel they must accept secrecy to obtain any remedy (Barmes, 2023). The use of NDAs in cases of executive sexual misconduct is inherently immoral, functioning as a mechanism of corporate complicity. These legal contracts are effectively weaponized to shield serial predators from accountability and public scrutiny. While often framed as neutral tools for privacy, research indicates that these agreements primarily serve to facilitate a systemic cover-up.

Research indicates that while these agreements are often presented as neutral tools for privacy, they frequently serve to silence survivors and prevent them from warning others, thereby enabling harmful patterns of behavior to persist unchecked within corporate hierarchies. By legally mandating silence, NDAs inflict "secondary traumatization" and "spirit injury" on

survivors, who are forced to witness the continued victimization of colleagues while being contractually prohibited from intervening or disclosing the perpetrator's history. In some cases, lawyers might experience "vicarious trauma," feeling responsible for a harm they were legally prevented from stopping (Chan et al., 2024; Williams, Faber, & Zare, 2025).

Some prominent examples of powerful individuals who weaponized the legal system and NDAs to shield their misconduct include Harvey Weinstein and Roger Ailes. Harvey Weinstein used highly restrictive confidentiality agreements and NDAs in at least eight settlements with women who alleged sexual harassment or assault, enabling him to continue his abuse over many years. Roger Ailes, the former chairman of Fox News, similarly used NDAs in harassment settlements to prevent numerous victims from speaking publicly about their experiences, thereby protecting his position and reputation (Ence, 2019; Hill, 2017).

Here, the 'hired gun' role is starkly illustrated, as lawyers are paid to construct legal shields that enable wrongdoing. The ethical conflict is acute: in serving the client's immediate interests, the lawyer facilitates a greater harm, directly contributing to the erosion of justice and public trust that the profession is meant to uphold.

BEYOND NDAS: BINDING ARBITRATION, CONTRACT SLANTING, AND THE LAWYER'S ETHICAL DILEMMA

Beyond the misuse of NDAs, this systemic erosion of fairness is entrenched in the 'hired gun' ethos of modern contract drafting. The ubiquity of mandatory binding arbitration clauses in employment and consumer contracts serves as a case in point. By burying these clauses in boilerplate text, drafters strip individuals of their day in court, funneling them instead into a private system characterized by corporate bias and diminished transparency (Stone & Colvin, 2015). The creation of such one-sided instruments raises a fundamental ethical question: at what point does a lawyer’s duty to their client become a systematic assault on the rights of the public? While lawyers are obligated to be zealous advocates, this duty is frequently used to justify drafting contracts that are intentionally slanted. The result is a "moral hazard" in which the lawyer acts as a strategic architect of legal shields designed to insulate clients from accountability (Zacks, 2017).

Many scholars argue that a lawyer’s responsibility should encompass the public good instead of being restricted to a client’s immediate demands (Duhl, 2010). By embedding legally dubious clauses meant to intimidate or provide one-sided advantages, lawyers deviate from their fundamental purpose as mediators and healers of conflict (Kunz, 2006). They become agents of institutional power, ultimately reinforcing the public cynicism that fuels the disparaging jokes with which this paper began. From a contract law perspective, scholars question both the enforceability and legitimacy of such instruments. An analysis of the NDA signed by Zelda Perkins in the Weinstein case applies standard doctrines—capacity, duress, public policy—to argue that many high-pressure settlement NDAs, particularly when paired with arbitration and unusually broad confidentiality, are legally vulnerable and normatively suspect (Grossi, 2025). The broader literature on “hushing contracts” and “other people’s contracts” challenges the assumption that private agreements are morally self-justifying where they impose significant negative externalities on third parties and the public (Altman, 2022).

The challenge is conceptual as much as regulatory: how should lawyers understand their role when the “legal” use of contracts conflicts with justice, dignity, or public safety? This question

points toward a need to re-embed values and purpose within the self-conception of the profession itself.

A PROFESSION IN NEED OF A SOUL: THE CASE FOR VALUES-BASED REFORM

Modern business paradigms are shifting to emphasize the necessity of ethical integrity in corporate strategy. Research indicates that 92% of millennials favor purchasing from ethical companies (Shewan, 2020; Aflac Corporate Social Responsibility study, n.d.), while nearly half view traditional capitalism skeptically due to its association with greed (Edwards, 2019). This shift is equally prevalent in the labor market. Younger workers are increasingly motivated by mission and values rather than pay alone (Robison, 2019). Modern employees increasingly demand that their work serve as a path to personal fulfillment (Batuchina et al., 2025). Data shows that 86% of Gen Z and 89% of millennial professionals consider organizational purpose vital for their well-being, making it a non-negotiable priority for talent acquisition (Deloitte, 2024). Indeed, a strong sense of purpose has become a baseline requirement for the modern workforce, with 70% of employees stating they would refuse to work for a company that lacks it (Stobierski, 2021). This cultural shift translates into direct action during the hiring process, as approximately four in ten professionals have already rejected specific job offers that conflicted with their personal values. This prioritization of integrity over financial gain is so pronounced that many professionals would even accept a 15% salary reduction to work for a purpose-driven organization (Meister, 2012).

Several themes of research argue that contemporary legal practice suffers from a value-neutral conception of professionalism. A dominant approach emphasizes strong client advocacy within the limits of the law, while treating personal morality and broader social values as irrelevant or even inappropriate considerations (Fitzpatrick & Queenan, 2020). Work on professional identity formation in legal education highlights an alternative: lawyers who have integrated moral commitments—such as deep responsibility to others, integrity, and public service—tend to be more effective and resilient (Fitzpatrick & Queenan, 2020; Bilionis, 2019). Leadership-oriented courses that invite students to explore their "story of self," emotional intelligence, and ethical decision making help bridge the gap between personal and professional values, reinforcing ideals of fairness, honesty, and trust as central to legal practice (Fitzpatrick & Queenan, 2020).

Similarly, value-oriented pedagogical experiments demonstrate that teaching law through value balancing—using case studies, simulations, and value-based analysis of regulatory texts—enhances students' ability to weigh competing interests and cultivates an "ethical culture" among future lawyers (Zavhorodnia et al., 2019). A Giving Voice to Values (GVV) framework, adapted from business ethics, encourages lawyers to move beyond abstract debates over right and wrong to practical strategies—reframing, building coalitions, bridging value gaps—for implementing ethical intentions in organizational contexts (Plump, 2021). These initiatives suggest that the profession's "soul" cannot be restored solely through external regulation of NDAs or arbitration clauses. Instead, law requires internal transformation of identity and purpose: lawyers should see themselves not as neutral service providers in a moral vacuum but as participants in institutions whose legitimacy depends on aligning legal technique with substantive justice and human flourishing.

PURPOSE-DRIVEN LEADERSHIP AND WHAT LAW CAN LEARN FROM IT

Leadership studies provide a rich vocabulary for thinking about purpose in organizations. While some practice-oriented approaches caution against over-romanticizing purpose, they still treat democratic, emancipatory, and collaborative processes as central to "leaderful practice," where direction emerges from shared values and dialogue rather than top-down commands (Raelin, 2021). Leadership-as-practice perspectives emphasize that ethical commitments are co-constructed through social interaction and reflexive engagement with others, not simply asserted as individual virtues (Raelin, 2021). Corporate law has begun to institutionalize these insights. Emerging profit with purpose or "société à mission" forms in France and Europe embed explicit social or environmental purposes in corporate constitutions, alongside duties of vigilance and accountability (Segrestin & Levillain, 2023; Ventura, 2023). By making purpose a legal element of the corporate form, these reforms aim to resolve the long-standing ambiguity between shareholder primacy and broader corporate interests (Segrestin & Levillain, 2023). Purpose-driven corporate law illustrates how legal frameworks can both constrain and empower management to pursue goals beyond short term profit, while preserving entrepreneurial freedom (Segrestin & Levillain, 2023; Ventura, 2023).

Purpose-driven leaders connect daily tasks to a mission that transcends financial objectives. These leaders do not merely manage; they articulate a compelling vision that helps employees understand how individual contributions create a positive impact. By forging an organizational identity rooted in purpose, leaders instill a robust sense of belonging that boosts morale and is a critical factor in retaining top talent. This sense of meaning is not confined to specific industries or nonprofits (Friedman & Pham, 2023). As Duncan (2025) notes, meaningful work is not defined by job titles but can be cultivated in any role through strong leadership. By encouraging this profound sense of purpose, leadership enhances employee engagement and resilience during difficult times. This style of management promotes authentic, values-based decision-making and an empathetic understanding of what motivates each team member. Consequently, purpose becomes the ultimate talent attractor, securing candidates who are passionate about making a difference and forging high-performance teams united by shared aspirations.

The demand for purpose-driven leadership has become critical as employees seek evidence that their work contributes to the broader community and the world. When employees perceive their work as meaningful, they demonstrate higher levels of motivation, creativity, and loyalty. This alignment between personal and organizational values unlocks creative thinking and prompts employees to propose groundbreaking solutions. Extensive research confirms that purpose-centered organizations cultivate higher employee engagement and stronger affective commitment (Allan et al., 2019; Henderson & Van den Steen, 2015). This crucial alignment is essential for sustaining cultures where efforts create genuine positive change (Ribiero, Costa, and Ramos, 2024). The resulting collaboration drives further innovation, allowing these organizations to outperform traditional competitors. This cultural transformation has compelled business leaders to reassess the corporation's fundamental purpose. A landmark signal of this shift occurred in 2019 when nearly 200 CEOs signed a Business Roundtable declaration. This document committed their organizations to serving the interests of all stakeholders, including employees, communities, and the environment, rather than focusing exclusively on shareholders (Gelles & Yaffa-Bellany, 2019).

The Conscious Capitalism movement reflects a similar orientation through its focus on higher purpose and conscious leadership (Mackey & Sisodia, 2013). Proponents of this movement

contend that capitalism can possess a soul. This philosophy is formally articulated in the "Four Tenets of Conscious Capitalism" as outlined by the organization (Conscious Capitalism, 2025; Wickam, 2022). At its core, this worldview holds that businesses should be animated by a meaningful purpose that transcends profit-making. Within this framework, financial returns are treated as a byproduct of doing good rather than the primary objective. This approach calls for creating value across the full spectrum of stakeholders. It is guided by leaders who can reconcile competing interests while deliberately shaping organizational cultures that reflect the company's deeper values and mission. The business case for this strategy is compelling, as companies aligned with these principles have reportedly outpaced S&P 500 firms by a factor of 14 (Lewis, 2020).

Investors are also applying pressure through Environmental, Social, and Governance (ESG) criteria. The Big Four accounting firms are currently developing ESG standards for annual reporting, while organizations like Gallup are building tools to embed these metrics into daily business practice (Clifton, 2021). Furthermore, Ethisphere's annual list of the world's most ethical companies (Ethisphere, 2025) and JUST Capital's rankings of responsible firms (JUST Capital, 2025) reflect a growing public appetite for accountability. The fundamental shift in corporate ethos from greed to contentment and ethical stewardship. As of early 2026, there are over 10,500 Certified B Corporations worldwide (B Lab, 2026). Clearly, balancing profit with social and environmental responsibility is achievable.

The legal field will likely feel the ripple effects of this corporate sea change.

IS PURPOSE-DRIVEN LAW POSSIBLE?

Chief Justice Warren E. Burger frequently advocated for a streamlined judicial system, urging the legal community to find alternatives to traditional jury trials to resolve disputes more efficiently. His vision for reform extended beyond the courtroom to include improvements to the prison system and the elevation of standards in legal education. Most notably, Burger challenged attorneys to transcend the role of a "hired gun" and instead embrace their potential as "healers of conflict." (Burger, 1984).

Unfortunately, the modern legal landscape often prioritizes financial gain over ethical service and integrity. This shift in values is underscored by a staggering statistic: the United States is home to more than 650,000 lawyers. This figure represents two-thirds of the world's legal advocates and shows that the U.S. has the highest number of lawyers per capita globally. Rather than a point of pride, this saturation suggests an overly litigious society in which the pursuit of justice is often overshadowed by a bloated, expensive legal industry (Burger, 1984).

There is a movement to see lawyers as healers, as individuals who know how to listen and can solve problems. Vijayendran quotes Gandhi to demonstrate that lawyers can have a soul.

> I had learnt the true practice of law. I had learnt to find out the better side of human nature and to enter men's hearts. I realized that the true function of a lawyer was to unite parties riven asunder. The lesson was so indelibly burnt into me that a large part of my time during the twenty years of my practice as a lawyer was occupied in bringing about private compromises of hundreds of cases. I lost nothing thereby—not even money, certainly not my soul (Gandhi quoted in Vijayendran, 2017).

Whether purpose-driven law is possible in practice depends on institutional and cultural change, but some areas of practice already reflect a more healing-oriented approach to lawyering. For example, in family law and restorative justice contexts, lawyers are recast not as adversaries but as facilitators of dialogue and repair (Webb & Ousky, 2006). Values-based and GVV-informed training equips lawyers with practical tools to resist unethical uses of legal instruments. For instance, when confronted with demands for sweeping NDAs or forced arbitration, a purpose-driven lawyer can reframe issues for clients, highlight reputational and ethical risks, and propose alternative resolutions that respect both legitimate confidentiality and victims' autonomy (Plump, 2021; Zavhorodnia et al., 2019). The question of whether purpose-driven law is achievable in practice is not merely theoretical. It has been implemented in the domain of the collaborative law movement.

COLLABORATIVE LAW: THE HEALER MODEL IN PRACTICE

Collaborative law may initially sound like an oxymoron, yet it has developed into a well-established and increasingly influential process in contemporary legal practice. It is built around a structural commitment to non-adversarial problem-solving: clients and lawyers sign a participation agreement committing all parties to honest disclosure, good-faith negotiation, and the pursuit of mutually acceptable solutions (Webb & Ousky, 2006). In a typical collaborative matter, both lawyers and clients commit in advance that counsel will be engaged solely to facilitate negotiation and settlement. In addition, if the process breaks down and the dispute proceeds to litigation, the parties must retain new attorneys to represent them in court (Hoffman, 2007). This structural feature creates a strong incentive for all participants to invest in problem-solving rather than posturing, since escalation to trial automatically triggers the cost and disruption of changing counsel.

Beyond its procedural design, collaborative law reflects a broader shift from adversarial advocacy toward interest-based negotiation, transparency, and joint decision-making. It often incorporates interdisciplinary teams, including financial professionals and mental health practitioners, to address the legal, emotional, and practical dimensions of conflict in an integrated way. In family and business disputes alike, the model aspires to preserve ongoing relationships, protect privacy, and empower clients to craft durable, customized solutions that a court might be ill-equipped to design. In this sense, collaborative law not only reconfigures the lawyer's role from gladiator to trusted counselor but also serves as a counter-narrative to the popular image of lawyers as combatants whose success depends on the destruction of the other side (Hoffman, 2007).

CONCLUSION

Lawyer jokes are more than mere entertainment; they are cultural diagnostics that expose a profession perceived to have drifted from justice toward self-interest. When a punchline about mass lawyer fatalities reliably draws laughter, it signals something deeper than simple irreverence. Instead, it reflects widespread public disillusionment with a system that too often protects the powerful at the expense of the vulnerable. The abuse of NDAs to silence survivors of sexual misconduct represents perhaps the starkest modern example of law weaponized against

the very people it should protect. Yet developments in purpose-driven corporate law, leadership studies, and legal education offer a counter trajectory: embedding explicit purposes and values into organizational forms, training lawyers as ethical leaders, and cultivating professional identities grounded in responsibility, integrity, and public service (Segrestin & Levillain, 2023; Ventura, 2023; Plump, 2021; Raelin, 2021; Zavhorodnia et al., 2019; Fitzpatrick & Queenan, 2020; Bilionis, 2018). This profession contains the seeds of its own redemption. Chief Justice Burger's vision of lawyers as "healers of conflict," Gandhi's practice of uniting parties rather than dividing them, and the emergence of collaborative law all point toward a different model grounded in integrity, empathy, and genuine service. This is not naïve idealism. This shift aligns with a documented transformation already underway in the broader business world, where purpose-driven organizations outperform purely profit-driven competitors and where younger professionals increasingly refuse to separate their work from their values.

A purpose-driven law would not eliminate conflict or ambiguity, nor would it render NDAs or arbitration inherently illegitimate. Instead, it would demand that legal tools be evaluated against the ends they serve and the human impacts they produce, especially on the most vulnerable. Collaborative and healing-oriented practices demonstrate that such a reorientation is practically possible. The deeper question is whether the profession will embrace its role not just as an operator of rules but as a steward of the values that make law worthy of the public’s trust. The legal profession would do well to heed this evolution. As clients and talent migrate toward organizations they trust, law firms that embrace ethical purpose as an operating principle rather than a marketing posture will be better positioned to thrive. By doing so, they can attract top talent and restore the public legitimacy that the profession has long been hemorrhaging. The lawyer as healer is not an anachronism; it may, in fact, be the most viable future the profession has.

References

Aflac (n.d.). Corporate social responsibility. https://chronicle-assets.s3.amazonaws.com/7/items/biz/pdf/AflacCorporateSocialResponsibility.pdf

Allan, B. A., Batz-Barbarich, C., Sterling, H. M., & Tay, L. (2019). Outcomes of meaningful work: A meta-analysis. *Journal of Management Studies*, 56(3), 500-528. https://doi.org/10.1111/joms.12406

Altman, S. (2022). Selling silence: the morality of sexual harassment NDAs. *Journal of Applied Philosophy, 39*(4), 698-720.

Barmes, L. (2023). Silencing at work: Sexual harassment, workplace misconduct and NDAs. *Industrial Law Journal, 52*(1), 68-106.

B Lab. (2026). *B Corp global directory*. https://www.bcorporation.net/en-us/

Batuchina, A., Iždonaitė-Medžiūnienė, I., & Lecaj, R. (2025). Multidimensional scale of meaningful work: construction and validation. *Frontiers in Psychology, 16*. https://doi.org/10.3389/fpsyg.2025.1578825

Bilionis, L. D. (2019). Law School Leadership and Leadership Development for Developing Lawyers. *Santa Clara Law Review, 58*, 601.

Burger, W. E. (1984, February 14). Burger on lawyers. https://www.csmonitor.com/1984/0214/021436.html?utm

Chan, A. L., Nichols, B. J., Crossley, A. D., & Daub, A. (2024). Assessing the impact of nondisclosure agreements and forced arbitration clauses on survivors of workplace sexual harassment and discrimination. https://gender.stanford.edu/media/6696/download?inline

Clifton, J. (2021, April 28). Does capitalism need a soul transplant? https://www.gallup.com/workplace/347156/capitalism-need-soul-transplant.aspx?

Conscious Capitalism (2025). Conscious capitalism philosophy: Learn about the four tenets of conscious capitalism. https://www.consciouscapitalism.org/philosophy

Davies, C. (2011). *Jokes and targets.* Indiana University Press.

Deloitte (2024). 2024 Gen Z and Millennial Survey: Living and working with purpose in a transforming world. https://www.deloitte.com/global/en/issues/work/content/genz-millennialsurvey.html

Duhl, G. M. (2010). The ethics of contract drafting. *Lewis & Clark Law Review*, 14(3), 989-1033.

Duncan, R. D. (2025, April 10). Meaningful work shows the power of purpose. https://www.forbes.com/sites/rodgerdeanduncan/2025/04/10/meaningful-work-shows-the-power-of-purpose/

Edwards, L. (2019, April 24). These are the most telling failures of socialism. https://www.heritage.org/progressivism/commentary/these-are-the-most-telling-failures-socialism

Ence, J. (2019). "I like you when you are silent": The future of NDAs and sexual-harassment claims. *Journal of Dispute Resolution*, *2019* (2), https://scholarship.law.missouri.edu/jdr/vol2019/iss2/10/

Ethisphere (2025). The 2025 world's most ethical companies® honorees list. https://worldsmostethicalcompanies.com/honorees/

Fitzpatrick, M. W., & Queenan, R. (2020). Professional identity formation, leadership and exploration of self. *UMKC Law Review, 89*, 539.

Ford, T. E., & Ferguson, M. A. (2004). Social consequences of disparagement humor: A prejudiced norm theory. *Personality and Social Psychology Review*, 8(1), 79-94. https://doi.org/10.1207/S15327957PSPR0801_4

Freud, S. (1960/1905). *Jokes and their relation to the unconscious.* (James Strachey, translator). W. W. Norton.

Friedman, H. H., & Friedman, L. W. (2019, May 30). The pen is mightier than the sword: Humor in the service of social justice (May 30, 2019). *Review of Contemporary Philosophy*, 19, 26-42. Also available at SSRN: https://ssrn.com/abstract=3396640 or http://dx.doi.org/10.2139/ssrn.3396640

Friedman, H. H., & Pham, N. C. (2023). Self-centered vs. humanity-centered: The most critical continuum for choosing today's leadership." *Journal of Values-Based Leadership,* Summer/Fall 2023, 1-24. https://scholar.valpo.edu/cgi/viewcontent.cgi?article=1455&context=jvbl

Galanter, M. (2005). *Lowering the Bar: Lawyer Jokes and Legal Culture*. University of Wisconsin Press. https://doi.org/10.2307/jj.36032641

Gelles, D., & Yaffe-Bellany, D. (2019). Shareholder value is no longer everything, top CEOs say. https://www.nytimes.com/2019/08/19/business/business-roundtable-ceos-corporations.html

Grossi, R. (2025). NDAs: Legally unenforceable or just unethical? A contract law perspective. *Alternative Law Journal, 50*(4), 284-290.

Helmreich, W. (2004). *The things they say behind your back*. Transaction Publishers.

Henderson, R., & Van den Steen, E. (2015). Why do firms have "purpose"? The firm's role as a carrier of identity and reputation. *American Economic Review, 105*(5), 326–330. https://doi.org/10.1257/aer.p20151072

Hill, A. (2017). Nondisclosure agreements: Sexual harassment and the contract of silence. *Gender Policy Report.* https://genderpolicyreport.umn.edu/nondisclosure-agreements-sexual-harassment-and-the-contract-of-silence/

Hoffman, D. (2007). A healing approach to the law. https://www.csmonitor.com/2007/1009/p09s01-coop.html

JUST Capital. (2025). 2025 rankings: America's most just companies. https://justcapital.com/rankings/

Kunz, C. L. (2006). The ethics of invalid and iffy contracts clauses. *Loyola of Los Angeles Law Review*, 40, 487-512. https://digitalcommons.lmu.edu/llr/vol40/iss1/13/

Lewis, M. (2020). What is conscious capitalism—definition and social responsibility in business. https://www.moneycrashers.com/conscious-capitalism-definition-social-responsibility-business/

Lynch, J. A., & Friedman, H. H. (2013, July). Using lawyer jokes to teach business ethics: A course module. *SSRN*. http://ssrn.com/abstract=2302910.

Mackey, J., & Sisodia, R. (2013). *Conscious Capitalism*. Harvard University Press.

Meister, J. (2012, June 7). Corporate social responsibility: A lever for employee attraction & engagement. http://www.forbes.com/sites/jeannemeister/2012/06/07/corporate-social-responsibility-a-lever-for-employee-attraction-engagement/

Plump, C. (2021). Giving Voice to Values in the Legal Industry. In *Giving Voice to Values* (pp. 123-143). Routledge.

Raelin, J. (2021). Leadership-as-practice: Antecedent to leaderful purpose. *Journal of Change Management, 21*(4), 385-390.

Ribeiro, M.F., Costa, C.G., & Ramos, F.R. (2024). Exploring purpose-driven leadership: Theoretical foundations, mechanisms, and impacts in organizational context. *Administrative Sciences, 14*(7), July, 1-33.

Robison, J. (2019, March 22). What millennials want is good for your business. https://www.gallup.com/workplace/248009/millennials-good-business.aspx

Segrestin, B., & Levillain, K. (2023). Profit-with-purpose corporations: Why purpose needs law and why it matters for management. *European Management Review, 20*(4), 733-740.

Shewan, D. (2020, February 25). Ethical marketing: 5 examples of companies with a conscience. https://www.wordstream.com/blog/ws/2017/09/20/ethical-marketing.

Stone, K. V. W., & Colvin, A. J. S. (2015). The arbitration epidemic: Mandatory arbitration deprives workers and consumers of their rights. https://www.epi.org/publication/the-arbitration-epidemic/

Ventura, L. (2023). New trends in legal frameworks for purpose-driven companies—The European way (s). *European Management Review, 20*(4), 725-732.

Vijayendran, G. (2017). The lawyer as healer. *Law Gazette.* https://lawgazette.com.sg/news/presidents-message/pm-november/

Webb, S. G., & Ousky, R. D. (2006). *The collaborative way to divorce: The revolutionary method that results in less stress, lower costs, and happier kids — without going to court.* Hudson Street Press.

Wickam, M. (2022). Conscious capitalism: An emerging economic philosophy for higher purpose in business. *Journal of Biblical Integration in Business.* https://doi.org/10.69492/jbib.v25i1.625

Williams, M. T., Faber, S. C., & Zare, M. (2025). The mental health consequences of nondisclosure agreements on survivors of workplace discriminatory harassment and abuse. *Routledge Open Research*, 4(10). https://doi.org/10.12688/routledgeopenres.19466.1

Zacks, E. A. (2017). The moral hazard of contract drafting. *Florida State University Law Review*, 42(4), 991-1034. https://ir.law.fsu.edu/lr/vol42/iss4/3

Zavhorodnia, V. M., Slavko, A. S., Degtyarev, S. I., & Polyakova, L. G. (2019). Implementing a Value-Oriented Approach to Training Law Students. *European Journal of Contemporary Education, 8*(3), 677-691.