

ANALOG PRIVILEGE

*Maroussia Lévesque**

This Article introduces “analog privilege” to describe how elites avoid artificial intelligence (AI) systems and benefit from special personalized treatment instead. In the register of tailor-made clothes and ordering off menu, analog privilege spares elites from ill-fitting, mass-produced AI products and services.

Our ability to curate our relationship with technology is a measure of our sophistication and, deep down, our power. Analog privilege connects with other instances of elites exercising agency over modernity: homesteading, no-phone teens and the coastal grandmother aesthetic all signal a return to the quaint pre-modern. As AI becomes the default modus operandi in many sectors from customer service to enforcing workplace rules, elites secure a manual override. Analog privilege allows them to escape AI systems that in theory apply to everyone but in practice spare the select few.

The existing literature focuses on whom AI harms, but this Article broadens the conversation to encompass whom it spares. Bringing attention to analog privilege highlights existing inequalities that enable special treatment for elites. This new lens provides a fuller picture of the distributional politics of AI, fostering a more capacious understanding of its social impact, and ultimately of the interconnectedness between precarity and privilege.

Analog privilege matters because it erodes the social fabric. Lending credence to the idea that elites play by different rules, the divide between people subject to and exempt from AI fuels resentment and polarization. Analog privilege is thus part of a larger strain on social peace. By making analog privilege legible, this Article clarifies the diffused sense of injustice that must be rectified if we are to regenerate the connective tissue that feeds our collective sense of belonging.

Once analog privilege comes into focus, the question becomes what to do about it. Legal interventions alone won't cut it. Instead, a multi-prong approach should align legal, technical, and other interventions.

* Doctoral candidate, Harvard Law School. I am privileged to have received crucial help, feedback and inspiration from Jane Bestor, Gabriella Blum, Aaron Dell, Veena Dubal, Nicolas Parra Herrera, Adam Holland, Karan Lala, Maxime Laasri, Amre Metwally, Martha Minow, Michelle Pearse, Gali Racabi, Guy Rubenstein, Glenn Rodriguez, Anna Shea, Shani Shisha, Christian Smalls, Theodora Skeadas, the Amnesty Tech team, the Berkman Klein Center, as well as the participants and organizers of the Fairness and Privacy seminar at Harvard Law School and the Law and Political Economy Project at LSA 2023. Thank you to the stellar editors of the New York University Journal of Legislation and Public Policy. Errors are my own.

INTRODUCTION	626
I. BENEFICIAL AND DETRIMENTAL AI SYSTEMS	631
A. Beneficial AI	631
B. Detrimental AI	632
C. Gray Areas	633
II. ANALOG PRIVILEGE	634
A. The General Rule: AI (and Its Harms)	637
B. The Exception: Analog Privilege	640
1. Elites, (Somewhat) Defined	640
2. Analog Privilege in Action	641
III. CASE STUDIES	643
A. LegalTech	643
1. Economy-Class Approximations	647
2. Premium Human Representation for the Well-Off	652
3. Applicable Law	657
B. Content Moderation	666
1. The General Rule: Scaling with Bulk Guesstimates	667
2. Human Review for the Select Few	672
3. Applicable Law	678
IV. SOLUTIONS	686
A. Reduce Analog Privilege	687
B. Accountability	688
C. Technical Fixes to Improve AI Systems	689
D. External Checks and Balances	691
CONCLUSION: EVERYTHING OLD IS NEW AGAIN	694
ANNEX I – CHATGPT EXAMPLE	695

INTRODUCTION

“[I]n this world, nothing is certain except death and taxes.”

– Benjamin Franklin

In theory, perhaps. But in practice, the privileged shirk, avoid, or delay what is otherwise inevitable. From cryogenic body preservation to offshore tax havens, exceptionalism has eroded the egalitarian ideal. Nowadays, elites also skirt punitive AI systems that seemingly apply to everyone. They have cheat codes to avoid reductive, deterministic, and invasive machine learning systems involved in an ever-expanding range of activities, engaging with AI on their own terms. A parallel analog track allows them to be seen in all their complexity, contradictions, and, ultimately, humanity. They have the luxury of dealing with humans instead of machines.

Viewing this double standard from the ground, one parent embroiled with child welfare services described how predictive analytics

stigmatize the poor and spare the rich. These systems flag indigent families seeking Medicaid-funded mental health services for possible child mistreatment, whereas affluent ones paying private therapists to treat the same conditions remain invisible to the state's automated suspicions.¹ Put simply, money buys richer families analog privilege. This preferential track insulates them from the prying eyes of AI systems. In light of convincing and conclusive evidence, few would contest that AI can create, reflect, and even amplify vulnerability.² But what about people who avoid it altogether?

This Article introduces the idea of “analog privilege”: the unearned ability to bypass AI, and instead benefit from personalized human treatment. “Treatment” encompasses decisions affecting individuals like hiring, parole release, or securing a mortgage; interactions with institutions, for instance whether one manages to speak with a customer service representative instead of the dreaded automated menu; and actual medical treatment, with AI therapists on the rise.³

Human treatment is analog⁴ in the sense that it entails messy yet textured attributes like discretion, subjective sensibilities, and acumen for context-sensitive factors that are not amenable to binary, neat calculations at the heart of automation.⁵ Although flawed in its own way, analog treatment can provide a strategic advantage over AI, especially when it comes to sidestepping detrimental systems that coerce and control. Analog privilege thus complicates the typical narrative about the digital divide, whereby inclusion in technology is beneficial and

1. Khadijah Abdurahman, *Birthing Predictions of Premature Death*, LOGIC MAG. (Aug. 22, 2022), <https://logicmag.io/home/birthing-predictions-of-premature-death> [<https://perma.cc/U5U2-9P3G>].

2. For example, in her compelling analysis of digital poorhouses, Virginia Eubanks suggests data-driven risk predictions confuse parenting while poor with poor parenting. VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018).

3. Dhruv Khullar, *Can A.I. Treat Mental Illness?*, NEW YORKER (Feb. 27, 2023), <https://www.newyorker.com/magazine/2023/03/06/can-ai-treat-mental-illness> [<https://perma.cc/C54M-LQZT>].

4. Analog refers to the absence of digital technology, rather than the technical definition of continuous electrical signal. See (and listen to) Erykah Badu, ... & On, *on MAMA'S GUN* (Motown Records 2000), <https://www.youtube.com/watch?v=sCnNAztvKjU&ab> [<https://perma.cc/FF6G-FD7A>]; see also Jennifer Raso, *Implementing Digitalisation in an Administrative Justice Context*, in THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE (2022) (distinguishing analog and digital interfaces in administrative agencies).

5. AI here refers to technological systems that analyze data to make predictions. It covers machine learning and other data-driven approaches to make decisions, evaluations, recommendations or hypotheses about people or groups. This article uses AI interchangeably with automation.

exclusion is detrimental.⁶ This may be true for some beneficial AI systems, but not so for detrimental ones.⁷

	Beneficial	Detrimental
Inclusion	Voluntary opt-in: power	Involuntary lock-in: vulnerability
Exclusion	Involuntary lock-out: vulnerability	Voluntary opt-out: power

Table 1: Power in beneficial and punitive systems

In the latter, involuntary lock-in denotes vulnerability whereas voluntary opt-out points to power. The “analog haves” yield the privilege to decline participation in AI systems, contrasting with the forced inclusion of the “analog have-nots.”⁸ This Article focuses on voluntary opt-out of detrimental systems, labeling this phenomenon “analog privilege.”

Analog privilege delves into an unexplored rift brought about by AI: shallow rudimentary guesstimates for most people, versus textured analog carve-outs at the discretion of the most privileged. Recent improvements in AI-driven conversational agents have led many to speculate that AI will be the great equalizer of our lifetime, for instance providing Generation Alpha with personalized tutors, or removing language barriers for non-native English speakers to seamlessly converse in the current lingua franca.⁹ But the future of AI is as perilous as it is promising; a closer look at how machine learning models actually work

6. ANTÓNIO GUTERRES, ROADMAP FOR DIGITAL COOPERATION - REPORT OF THE SECRETARY-GENERAL (June 2020), <https://www.un.org/en/content/digital-cooperation-roadmap/> [<https://perma.cc/KGE4-XB5A>]; ALEX COBHAM, THE UNCOUNTED 10 (2020); LARRY IRVING, FALLING THROUGH THE NET: DEFINING THE DIGITAL DIVIDE (July 1999), <https://www.ntia.doc.gov/legacy/ntiahome/fttn99/FTTN.pdf> [<https://perma.cc/2T66-ZWK6>].

7. Part I, *infra*, defines beneficial and detrimental systems.

8. *Call for Participation: Please Don't Include Us*, DIGIT. JUST. LAB, <https://digitaljusticelab.ca/cfp> [<https://perma.cc/CS7N-88PK>] (criticizing inclusion initiatives); Chelsea Barabas, *Beyond Bias: Re-Imagining the Terms of “Ethical AI” in Criminal Law*, 12 GEO. J. L. & MOD. CRIT. RACE PERSP. 83 (2020).

9. ANDREW PENNOCK & NATASHA HENY, GENERATIVE AI HAS ARRIVED. WHAT DOES THIS MEAN FOR TEACHING AND LEARNING AT UVA? 5 (June 2023), <https://provost.virginia.edu/subsite/genai/task-forcereport> [<https://perma.cc/3DMY-CVAV>]; Viren Shah, *Artificial Intelligence: The Great Equalizer*, FORBES (Aug. 16, 2023), <https://www.forbes.com/sites/forbestechcouncil/2023/08/16/artificial-intelligence-the-great-equalizer/?sh=6d1199345cd2> [<https://perma.cc/B47U-WY38>]; Landon Garrison (@landongarrison), *Tutory - ChatGPT Plugin to Democratize Education for Everyone*, OPENAI DEV. F. (May 2023), <https://community.openai.com/t/tutory-chatgpt-plugin-to-democratize-education-for-everyone/186653/41?page=2> [<https://perma.cc/5GKJ-J7G9>]; Maurice Jakesch et al., *Human Heuristics for AI-Generated Language Are Flawed*, 120 PROC. NATL. ACAD. SCIS. e2208839120, 6–7 (2023) (adjusting the speaker’s accent to facilitate communication).

reveals significant downsides for people subject to error-prone, probabilistic guesstimates assuming the future mirrors the past and leaving little room for people to change and grow.

At first blush, analog privilege merely exposes existing power differentials that have long entitled elites to preferential treatment. Take hiring: backchannels already land the well-connecteds' resumes at the top of the pile. In that sense, automated applicant tracking systems favoring the lacrosse-playing Jareds of this world¹⁰ are but a permutation of the sempiternal boys' club. But the disempowerment of the analog have-nots is more multifaceted. The "little guy" can be an ordinary social media user with a modest following, a middle-class family resorting to LegalTech to fight an eviction notice, or even a partner track attorney stalked by his firm's productivity monitoring software. Analog privilege might track traditional class and race divides, but it could also crisscross traditional social categories.¹¹ This Article focuses on naming analog privilege, but further research is needed to define the contours of the analog haves and have-nots.

Analog privilege connects with broader narratives about unfair exceptions sparing elites. It fuels a sense of alienation because cookie-cutter guesstimates objectifying people into predictions clash with analog treatment empowering elites to be seen in all their complexity. Treating the analog have-nots as disposable, AI's errors about individuals are a small price to pay for overall good enough performance. Pursuant to this utilitarian approach, public services recipients are stereotyped into abusive parents, ordinary social media users have innocuous content taken down, employees are but cogs in the surveillance machine. The common thread is a loss of dignity, planting the seeds of resentment against people who bypass such systems.

By asking whom automation is imposed upon, and whom it spares, this Article throws power imbalances in relief. Joining a strand of critical legal scholarship noting how technology brings salience to existing social phenomena,¹² it uses analog privilege to shed light on

10. Dave Gershgorn, *Companies Are on the Hook If Their Hiring Algorithms Are Biased*, QUARTZ (Oct. 22, 2018), <https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased> [<https://perma.cc/KGB5-Z3X4>]; Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARV. BUS. REV. (May 6, 2019), <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias> [<https://perma.cc/T2ZC-Y9BG>].

11. Sandra Wachter, *The Theory of Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law*, 97 TUL. L. REV. 149, 174–76 (2022).

12. Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 31 (2004); see also Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 72, 1671–76 (2020).

longstanding double standards conferring elites unearned benefits. Articulating the double standard between people subject to and exempt from AI provides a tractable path towards addressing diffused and pernicious patterns of exceptionalism.

That said, theorizing analog privilege leaves open the possibility that some AI applications can actually reduce social divides and redistribute power more equitably. Bringing attention to AI's harms as the counterpoint of analog privilege is by no means a totalizing account of the technology's multifaceted impacts. AI holds tremendous potential, and has already delivered in a plethora of domains, including astonishing capabilities in art and science.¹³ Far from being a comprehensive account of automation, analog privilege is but one underexplored angle to observe some of its distributional effects.¹⁴ As the shadow self of pervasive AI systems, analog privilege has remained invisible so far because its main function is bypassing headline-grabbing, controversial AI harms. Carving out a negative space from increasingly pervasive automation, analog elites have therefore evaded sustained analysis in the critical discourse around algorithmic decision-making. Making analog privilege legible fosters more capacious discussions about algorithmic harms, how to address them, and how the rift between the analog haves and the analog have-nots fits within broader societal divisions.

Part I unpacks the distinction between beneficial and punitive systems to situate the discussion on analog privilege. Part II outlines three harms inherent to automation: reductivism, determinism, and digital voyeurism. Analog privilege emerges as the unearned ability to avoid these harms. Taken together, AI harms and analog privilege paint the picture of an automation divide: elites deciding how (if at all) they engage with AI, versus the rest of us lacking agency over how AI applies to our lives. Case studies in Part III illustrate analog privilege in action. When it comes to legal services, affluent litigants armed with human attorneys have a significant leg up over those resorting to economy-class LegalTech. On social media, states and high-profile users avoid crude automated content moderation. Solutions in Part IV lay the groundwork for future research on polycentric governance. It invites policymakers to deliberately overlay interventions ranging from fairly superficial technical fixes to deeper structural reform addressing the latent inequalities evidenced by analog privilege.

13. *DALL·E: Creating Images from Text*, OPENAI (Jan. 5, 2021), <https://openai.com/research/dall-e> [<https://perma.cc/MB8G-4VM5>]; John Jumper et al., *Highly Accurate Protein Structure Prediction with AlphaFold*, 596 *NATURE* 583 (2021).

14. David Freeman Engstrom & Jonah B. Gelbach, *Legal Tech, Civil Procedure, and the Future of Adversarialism*, 169 *U. PENN. L. REV.* 1001, 1037–38 (2021) (citing HAROLD D. LASSWELL, *POLITICS; WHO GETS WHAT, WHEN, HOW* (1936)).

I. BENEFICIAL AND DETRIMENTAL AI SYSTEMS

Automation streamlines management. In the Foucauldian sense of discipline, it turns multitudes into ordered, manageable multiplicities.¹⁵ For our purposes, we can distinguish two functions of AI systems: beneficial and detrimental automation. At one end of the spectrum, beneficial automation frees us from repetitive and mindless tasks, streamlining our activities. At the other end, detrimental automation supercharges enforcement capacities, with adverse consequences for enlisted people. When systems bear beneficial and detrimental features, their net effect tends to be detrimental.

A. Beneficial AI

Beneficial AI avoids the drudgery of fastidious analog processing. In that sense, beneficial AI is co-constructed relative to its analog counterpart as a more convenient, efficient, or seamless option. Smart borders are a case in point. In what Sparke describes as the biopolitical production of a transnational business class citizenship,¹⁶ states deploy technologies like facial recognition to streamline border crossing for pre-approved, low-risk individuals.¹⁷ Automated systems speed up immigration and security-clearing processes, implementing the neoliberal ideology of a borderless world propelled by free trade agreements.¹⁸ Frictionless border crossing is part of the broader phenomenon of atomic SUV citizenship: “[w]e ride high and sovereign; we are masters of space; we are safe against all who might intrude, all who might stand in our way.”¹⁹ Now open to anyone who clears a background check and disburses the \$100 application fee, it is unsurprising that smart borders began as explicitly status-based, reserved for Platinum Elite Frequent Fliers.²⁰ In the lineage of airport lounges, Amex Black cards, red carpet perks, and white glove service for frequent travelers, business elites

15. MICHEL FOUCAULT, *DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON* 148 (1979).

16. Matthew B. Sparke, *A Neoliberal Nexus: Economy, Security and the Biopolitics of Citizenship on the Border*, 25 *POL. GEOGRAPHY* 151 (2006).

17. *Global Entry*, U.S. CUSTOMS & BORDER PROT., <https://www.cbp.gov/travel/trusted-traveler-programs/global-entry> [https://perma.cc/RET8-2NMN] (last modified Oct. 25, 2023).

18. Peter Adey, “*Divided We Move*” *The Dromologies of Airport Security and Surveillance*, in *SURVEILLANCE AND SECURITY: TECHNOLOGICAL POLITICS AND POWER IN EVERYDAY LIFE* 195 (Torin Monahan ed., 2006).

19. Don Mitchell, *The S.U.V. Model of Citizenship: Floating Bubbles, Buffer Zones, and the Rise of the “Purely Atomic” Individual*, 24 *POL. GEOGRAPHY* 77, 96 (2005).

20. Elec. Priv. Info. Ctr., Comment Letter on TSA Aviation Security Screening Records Privacy Act Notice (Feb. 24, 2003), <https://archive.epic.org/privacy/airtravel/tsacomment2.24.2003.html> [https://perma.cc/7MTF-7Q36]; U.S. GEN. ACCT. OFF., GAO-03-253, *AVIATION SECURITY: REGISTERED TRAVELER PROGRAM POLICY AND*

welfare program cuts off recipients for technicalities,²⁶ a Dutch welfare fraud detection algorithm targets non-citizens,²⁷ a UK “Gang Matrix” disproportionately suspects Black people.²⁸ The common theme across these deployments is that the most vulnerable bear the brunt of detrimental automation.

Private actors deploy automation to control and punish individuals as well. For instance, social media platforms lean on AI to surveil, detect, and punish violations of their content policies.²⁹ Of course, enforcing content policies provides welcomed ramparts against abusive and unsafe content. The detrimental aspect refers to the system’s primary function of control and coercion.

C. Gray Areas

The delineation between beneficial and detrimental systems isn’t always this clear-cut in practice. Nominally beneficial automation programs can easily morph into detrimental ones. For example, data-driven risk predictions designed to “help” at-risk parents trap them in benevolent surveillance, enforcing compliance through the thinly veiled threat of breaking families apart.³⁰ The quantified self-movement also blurs the beneficial/detrimental dichotomy, with empowering narratives about datafication masking its pernicious implications.³¹ For example,

26. Luke Henriques-Gomes, *The Automated System Leaving Welfare Recipients Cut off with Nowhere to Turn*, GUARDIAN, (Oct. 16, 2019), <https://www.theguardian.com/technology/2019/oct/16/automated-messages-welfare-australia-system> [<https://perma.cc/7L2S-U928>]; Gillian Terzis, *Austerity is an Algorithm*, LOGIC MAG. (Dec. 1, 2017), <https://logicmag.io/justice/austerity-is-an-algorithm/> [<https://perma.cc/7WJG-3UBM>].

27. AMNESTY INT’L, XENOPHOBIC MACHINES: DISCRIMINATION THROUGH UNREGULATED USE OF ALGORITHMS IN THE DUTCH CHILDCARE BENEFITS SCANDAL 22 (2021), <https://www.amnesty.org/en/documents/eur35/4686/2021/en/> [<https://perma.cc/JJ3P-YWD2>].

28. Seventy-eight percent of suspects in the database are Black, whereas they represent only 13% of London’s population and 27% of people the police identifies as responsible for serious youth violence. AMNESTY INT’L, TRAPPED IN THE MATRIX: SECRECY, STIGMA, AND BIAS IN THE MET’S GANGS DATABASE 3 (2018), <https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf> [<https://perma.cc/9UD9-AY2X>].

29. NAFA CHOWDHURY, AUTOMATED CONTENT MODERATION: A PRIMER 3 (2022), <https://cyber.fsi.stanford.edu/news/automated-content-moderation-primer> [<https://perma.cc/7XUH-VT8Z>].

30. EUBANKS, *supra* note 2, at 157, 165. For a first-person account see Abdurahman, *supra* note 1.

31. The quantified self movement consists in leveraging data from trackers such as wearable sleep and fitness monitors to improve physical, mental and emotional performance. For a critical analysis, see JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 81 (2019); Jean-François De Moya & Jessie Pallud, *From Panopticon to Heautopticon: A New Form of*

sleep, exercise, diet, and menstrual cycle self-tracking adversely impact insurance premiums.³² Beneficial systems can therefore weaponize datafication to control, surveil, and punish.³³ When only the wealthiest can afford to opt out of datafied insurance, the voluntary aspect of these tracking technologies rings hollow.³⁴ Workplace wellness programs similarly straddle the line between voluntary and coercive.³⁵ Opt-in bonuses rewarding participation in personalized weight loss or smoking cessation programs can just as persuasively be framed as a penalty for opting out, in the form of foregoing a benefit.³⁶ A common thread seems to be that beneficial perks often conceal more serious detrimental consequences.

Despite ambivalence in the real world, distinguishing beneficial from detrimental AI helps tease out how analog privilege plays out differently across these configurations. Analog privilege is self-evident when it manifests as the ability to avoid detrimental systems and their adverse effects. But it can also occur in beneficial systems, providing even more advantages than automation. The LegalTech case study will return to this phenomenon.

II. ANALOG PRIVILEGE

The literature focuses on AI harms, giving scant attention to whom AI spares. By now, algorithmic bias against historically marginalized groups is a widely documented phenomenon.³⁷ Debates among legal

Surveillance Introduced by Quantified-Self Practices, 30 INFO. SYS. J. (SPECIAL ISSUE) 940, 957–58 (2020).

32. Liz McFall, *Personalizing Solidarity? The Role of Self-Tracking in Health Insurance Pricing*, 48 ECON. & SOC. 52, 70–71 (2019).

33. Chris Gilliard & David Golumbia, *Luxury Surveillance*, REAL LIFE MAG., (July 2021), <https://reallifemag.com/luxury-surveillance/> [<https://perma.cc/BAN3-ASNB>].

34. GINA NEFF & DAWN NAFUS, SELF-TRACKING 35 (2016).

35. Rachel Emma Silverman, *Bosses Tap Outside Firms to Predict Which Workers Might Get Sick*, WALL ST. J., (Feb. 17, 2016), <https://www.wsj.com/articles/bosses-harness-big-data-to-predict-which-workers-might-get-sick-1455664940> [<https://perma.cc/8AA7-V7LA>].

36. Gilliard & Golumbia, *supra* note 33; Ifeoma Ajunwa, Kate Crawford & Jason Schultz, *Limitless Worker Surveillance*, 105 CALIF. L. REV. 735, 768 (2017); *see also* EEOC *Provides Proposed Wellness Rules for Review*, EQUAL EMP. OPPORTUNITY COMM’N, <https://www.eeoc.gov/newsroom/eeoc-provides-proposed-wellness-rules-review>, [<https://perma.cc/39LX-RNXF>] (proposing to define voluntary participation); *Regulatory Freeze Pending Review*, THE WHITE HOUSE (2021), <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/regulatory-freeze-pending-review/>, [<https://perma.cc/4QMF-2BGB>] (withdrawing proposal).

37. Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671, 684–85 (2016); CATHY O’NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM

scholars and machine learning scientists center around competing notions of fairness³⁸ and suggest procedural guarantees and transparency measures,³⁹ but have yet to fully theorize the split between the analog haves and the analog have-nots as a self-standing issue.

Moving closer to a systemic understanding of AI's effects, some have noted extractive data practices,⁴⁰ the erosion of trust,⁴¹ new group clusters defying traditional anti-discrimination categories,⁴² and unjust social relations playing out through the production and use of data.⁴³ Criminology studies have also contrasted analog and data-driven approaches when analyzing the shift to evidence-based

(2018); Angwin et al., *supra* note 25; RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019).

38. Mark MacCarthy, *Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms*, 48 CUMB. L. REV. 67, 102–03 (2017) (describing mutually exclusive notions of group and individual fairness); John Logan Koepke & David Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1751 n.127 (2017); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI*, 41 COMPUT. L. & SEC. REV., July 2021, at 4–5, <https://www.sciencedirect.com/science/article/abs/pii/S0267364921000406> [<https://perma.cc/95VZ-ZQJZ>]; Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear*, WASH. POST (Oct. 17, 2016), <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> [<https://perma.cc/2WAM-7QPM>]; Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, ARXIV (Sept. 19, 2016), <https://arxiv.org/abs/1609.05807> [<https://perma.cc/DAU9-SMD4>]; Sam Corbett-Davies & Sharad Goel, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, ARXIV (Aug. 14, 2023), <http://arxiv.org/abs/1808.00023> [<https://perma.cc/5VNM-QNFX>].

39. Danielle Keats Citron & Frank A. Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 18–30 (2014); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 116–17 (2018); Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 2034 (2017).

40. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM : THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019) (arguing that machine learning enables companies to sell “behavioural futures” through data mining); Veena Dubal, *On Algorithmic Wage Discrimination*, 123 COLUM. L. REV. 1929 (2023) (describing labor data extraction); see also Andrew Webster, *Actors Say Hollywood Studios Want Their AI Replicas — for Free, Forever*, VERGE (2023), <https://www.theverge.com/2023/7/13/23794224/sag-aftra-actors-strike-ai-image-rights> [<https://perma.cc/WRB8-VGP3>] (reporting on Hollywood studios seeking to digitize actors for future synthetic content creation).

41. Cynthia Dwork & Martha Minow, *Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law*, 151 DAEDALUS 309 (2022).

42. Wachter, *supra* note 11, at 168–69.

43. Salome Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 631 (2021).

rehabilitation.⁴⁴ Abolitionists have articulated how AI-based improvements in administering criminal justice distract from structural debates about the prison industrial complex.⁴⁵ Noting that blanket AI treatment will not work for edge cases, a proposed right to be an exception to machine learning acknowledges that some individuals need exemptions from otherwise applicable data-driven processes.⁴⁶ For atypical individuals who do not fit the mold of data-driven approximation, analog carve-outs are sometimes justified to avoid harm. That contribution paves the way to challenge power-based analog privilege as an unearned carve-out from AI systems. The literature anticipating LegalTech's equities suggests it could simultaneously democratize and stratify access to legal services,⁴⁷ yet doesn't fully theorize the implications of a two-speed model with economy-class LegalTech on one side and premium human representation on the other. Likewise, social media scholarship has laid bare trade-offs inherent to automating content moderation,⁴⁸ criticized flagging processes,⁴⁹ and set the

44. Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537 (2016).

45. Barabas, *supra* note 8 at 110.

46. Sarah H. Cen & Manish Raghavan, *The Right to Be an Exception to a Data-Driven Rule*, ARXIV (Dec. 28, 2022), <http://arxiv.org/abs/2212.13995> [<https://perma.cc/HS9N-NT67>].

47. Engstrom & Gelbach, *supra* note 14, at 1038 (speculating that LegalTech could usher in a "golden age of litigation in which those priced out of the current litigation system can more reliably vindicate their rights," but also "exacerbate existing power and resource disparities within the litigation system") (citations omitted); Alice Woolley & Trevor Farrow, *Addressing Access to Justice Through New Legal Service Providers: Opportunities and Challenges*, 3 TEX. A&M L. REV. 549, 577 (2016); Dana Remus & Frank Levy, *Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law*, 30 GEO. J. LEGAL ETHICS 501, 551 (2017).

48. James Grimmelman, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 65 (2015); Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT'L L.J. 41, 45–46 (2020); Tarleton Gillespie, *Content Moderation, AI, and the Question of Scale*, BIG DATA & SOC'Y, Aug. 21, 2020, <https://journals.sagepub.com/doi/full/10.1177/2053951720943234> [<https://perma.cc/E7DM-ZQP6>]; Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 548 (2022); Kate Klonick, *Of Systems Thinking and Straw Men*, 136 HARV. L. REV. F. 339, 345 (2023); see also Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, 7 BIG DATA & SOC'Y, no. 1, Jan.-June 2020, <https://journals.sagepub.com/doi/10.1177/2053951719897945> [<https://perma.cc/9BAC-VFTL>].

49. Rabea Eghbariah & Amre Metwally, *Informal Governance: Internet Referral Units and the Rise of State Interpretation of Terms of Service*, 23 YALE J.L. & TECH. 542, 564–65 (2021); Kate Crawford & Tarleton Gillespie, *What Is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint*, 18 NEW MEDIA & SOC'Y 410 (2016); Naomi Appelman & Paddy Leerssen, *On "Trusted" Flaggers*, 24 YALE J.L. & TECH. 452 (2022); Brian Chang, *From Internet Referral Units to International Agreements: Censorship of the Internet by the UK and EU*, 49 COLUM. HUM. RTS. L.

record straight on tiered moderation.⁵⁰ However, it has yet to address how some accounts get to avoid automation altogether thanks to preferential human review.

What is missing is a comprehensive account of analog privilege connecting the dots across AI deployments. This Part articulates the contrast between AI and analog treatment. Like the two-faced Janus figure from Roman mythology, AI harms and analog privilege are two sides of the same coin. This section outlines a general theory of AI harms (section A) and analog privilege as the flipside ability to avoid said harms (section B).

A. *The General Rule: AI (and Its Harms)*

As AI systems become increasingly pervasive, recurring harms start emerging across deployments. This section teases out three common harms in AI systems: reductivism, determinism, and voyeurism. Reductivism conveys the idea of superficial evaluations resulting in texture loss. Automated systems base decisions on a finite and curated corpus of training data, irrespective of context. When it comes to child mistreatment predictions, seeking Medicaid-funded mental health resources becomes a proxy for risk even though more affluent families experience just as many mental health challenges. Simplistic markers inferred from the training data thus overlook crucial contextual and fact-variable distinctions, trading off texture for efficiency. In short, mass approximation trumps individual correctness.

Machine learning's determinism constrains people's range, stifling their ability to beat the odds, flip the script, start over.⁵¹ This determinism stems from AI's fundamentally backward-looking posture. In predictive policing, for example, past crime data determines where to dispatch officers, leading to more crime detection and self-fulfilling prophecies cementing confirmation bias. Assuming the future mirrors the past is problematic when AI reinforces vicious circles of poverty,

REV. 114, 122 (2018); Molly K. Land, *Against Privatized Censorship: Proposals for Responsible Delegation*, 60 VA. J. INT'L L. 363, 404 (2019); Sebastian Felix Schwemer, *Trusted Notifiers and the Privatization of Online Enforcement*, 35 COMPUT. L. & SEC. REV. 1 (2019); see also Paresh Dave, *It's Getting Harder for the Government to Secretly Flag Your Social Posts*, WIRED (July 24, 2023), <https://www.wired.com/story/its-getting-harder-for-the-government-to-secretly-flag-your-social-posts> [<https://perma.cc/GV9C-GZ4F>].

50. Robyn Caplan & Tarleton Gillespie, *Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy*, 6 SOC. MEDIA + SOC'Y 1 (2020).

51. Julie E. Cohen, *Turning Privacy Inside Out*, 20 THEORETICAL INQUIRIES IN LAW 1, 12–13 (2019); Barbara Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L. J. 1408 (1979).

exclusion, and marginality for those living in so-called high-risk areas. Some approaches like stochastic modeling account for more variability to estimate the distribution of probable outcomes but still anchor predictions in past data.⁵² Future opportunities tied to past events incarcerate people in algorithmic prisons⁵³ based on past choices, behavior, events, and—most problematically—circumstances people are born into and have no agency over.⁵⁴

Artificial intelligence-driven automation also entails data voyeurism. Data-hungry AI systems capture every facet of our behavior. This pervasive monitoring departs from the clinical definition of voyeurism,⁵⁵ but the pathology is nevertheless a useful concept to convey objectification. For instance, automated workplace surveillance subjects employees to invasive, distracting, and obnoxious tattleware.⁵⁶ These systems

52. Sourabh Mehta, *Deterministic vs. Stochastic Machine Learning*, ANALYTICS INDIA MAG. (May 10, 2022), <https://analyticsindiamag.com/deterministic-vs-stochastic-machine-learning> [<https://perma.cc/3TX5-GJ8V>]; Sindhu Bhaskaran, *Deterministic vs. Stochastic Machine Learning: Which Approach Reigns Supreme in the World of AI?*, AITECHTREND (Feb. 24, 2023), <https://aitechtrend.com/deterministic-vs-stochastic-machine-learning-which-approach-reigns-supreme-in-the-world-of-ai/> [<https://perma.cc/M4G3-SXM5>]; IAN GOODFELLOW ET AL., DEEP LEARNING 52–54 (2015) (describing probability more generally).

53. Bill Davidow, *Welcome to Algorithmic Prison*, ATL. (Feb. 20, 2014), <https://www.theatlantic.com/technology/archive/2014/02/welcome-to-algorithmic-prison/283985> [<https://perma.cc/4JEU-DSHT>].

54. Northpointe Inc., *COMPAS Risk Assessment Questionnaire version 8.1.18.12* (2011), <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html> [<https://perma.cc/C62Y-GXNF>] contributed by Julia Angwin, *supra*, note 25 (see question 31 about single parent upbringing).

55. AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STAT. MANUAL OF MENTAL DISORDERS: DSM-5 (5th ed. 2013) (“arousal from observing an unsuspecting person who is naked, in the process of disrobing, or engaging in sexual activity, as manifested by fantasies, urges, or behavior”).

56. Katherine C. Kellogg, Melissa A. Valentine & Angèle Christin, *Algorithms at Work: The New Contested Terrain of Control*, 14 ACAD. MGMT. ANNALS 366 (2020); Richard A. Bales & Katherine V.W. Stone, *The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace*, 41 BERKELEY J. EMP. & LAB. L. 1 (2020); IFEOMA AJUNWA, *THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE* (2023); KAREN LEVY, *DATA DRIVEN: TRUCKERS, TECHNOLOGY, AND THE NEW WORKPLACE SURVEILLANCE* (2023); Jodi Kantor & Arya Sundaram, *The Rise of the Worker Productivity Score*, N.Y. TIMES (Aug. 15, 2022), <https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html> [<https://perma.cc/JLV5-R4FH>]; Danielle Abril & Drew Harwell, *Keystroke Tracking, Screenshots, and Facial Recognition: The Boss May Be Watching Long after the Pandemic Ends*, WASH. POST (Sept. 27, 2021), <https://www.washingtonpost.com/technology/2021/09/24/remote-work-from-home-surveillance> [<https://perma.cc/WC4F-7C6Y>]; ALESSANDRO DELFANTI, *THE WAREHOUSE: WORKERS AND ROBOTS AT AMAZON* (2021); U.S. Patent No. 9,881,276 (issued Jan. 30, 2018) (describing an Amazon patent for smart wristband monitoring workers hand movement).

ultimately negate people's agency, deeming them nothing more than the sum of their observable parts.⁵⁷

In addition to these inherent harms, AI systems often disproportionately impact marginalized groups. Take the child mistreatment predictions stigmatizing the poor, or predictive policing homing in on non-white neighborhoods: AI not only brings salience to existing inequalities but can also exacerbate them. Denying marginalized groups loans, parole, or jobs at higher rates, these systems reify stereotypes about certain groups being inherently risky, delinquent, or incompetent. Although individual outcomes may be formally correct, this kind of disparity is inconsistent with group-based conceptions of fairness attuned to the aggregate, systemic dimensions of a given process.⁵⁸

Despite these downsides, AI systems may still yield fairer, more accurate results than their analog counterparts. When articulating the downsides of automated processing, it is worth measuring harms against the baseline of comparable analog processes. Ambitions to smoke out latent prejudice in analog hunches often motivate the switch to AI systems, as with child mistreatment investigations historically skewed against Black families.⁵⁹ Similarly, criminal justice reforms to divert low-risk offenders from incarceration with evidence-based predictive analytics aimed to move away from outdated retributive models.⁶⁰ While these well-meaning intentions did not always pan out,⁶¹ they are important reminders to keep in mind the analog counterfactual when evaluating AI harms. Finally, AI harms can also show up in beneficial

57. Other disciplines like media studies have mobilized the language of voyeurism to describe power asymmetries. See CLAY CALVERT, *CULTURAL VOYEURISM* (2004); Osei Appiah, *Cultural Voyeurism: A New Framework for Understanding Race, Ethnicity, and Mediated Intergroup Interaction*, 68 J. COMM'N 233, 234 (2018) (claiming that reality tv and social media enable the public to "voyeuristically trawl through and consume the lives of people from different cultures"); see also Louise Racine & Amélie Perron, *Unmasking the Predicament of Cultural Voyeurism: A Postcolonial Analysis of International Nursing Placements*, 19 NURSING INQUIRY 190 (2012) (describing colonizing forces and exoticizing the Other in international nursing placements).

58. Barocas and Selbst, *supra* note 37 at 701–11 (disparate impact theory in the context of employment); Richard A. Primus, *The Future of Disparate Impact*, 108 MICH. L. REV. 1341 (2010) (calling into question the constitutionality of disparate impact statutes in light of the Equal Protection Clause's anti-classification, disparate treatment framework).

59. Hyunil Kim et al., *Lifetime Prevalence of Investigating Child Maltreatment Among US Children*, 107 AM. J. PUB. HEALTH 274, 277 (2017) (reporting that 53% of Black families are investigated, compared with 32% of Hispanics and 28% of White families respectively).

60. Klingele, *supra* note 44, at 565–67.

61. Logan Stapleton et al., *Extended Analysis of "How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions"*, ARXIV (Apr. 29, 2022), <https://arxiv.org/abs/2204.13872> [<https://perma.cc/M3FU-8Q92>].

systems. The case study on LegalTech exemplifies this dynamic, as these systems improve access to justice overall but still entail significant reductivism and determinism.

In sum, the effect of AI on inequality depends on context. In child mistreatment predictions, it drives further wedges between the haves and the have-nots. In legal services, it could narrow the gap in access to justice, but it could also perpetuate power asymmetries between litigants. The point is to nuance the superficial, techno-optimistic narrative positing AI as an inevitable equalizer.

B. *The Exception: Analog Privilege*

Analog privilege is the flipside of AI harms. Power affords certain groups—which this Article will refer to as “elites”—the unearned ability to engage with automation on their own terms, curating when, how, and to what extent. They turn to AI when convenient, as do frequent travelers clearing customs faster with biometric recognition. And they avoid it when detrimental, as do families buying their way out of AI-powered child mistreatment suspicions. After defining elites, this section outlines the dynamics of analog privilege.

1. *Elites, (Somewhat) Defined*

For the purposes of this Article, the concept of elite refers to groups of people that hold a lot of power within a given society. According to elite theory, industrialization concentrated power in the hands of economic, military, and political groups leveraging wealth, force, and prestige to assert their dominance.⁶² Elites thus draw power from different sources such as wealth, coercive capabilities, and social capital. Pierre Bourdieu defined social capital as “the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance or recognition.”⁶³ Modern-day social media influencers mobilize cultural, social, and symbolic resources to secure (and monetize) elite status in the social ordering.⁶⁴ As the content moderation case studies in Part III.B *infra* demonstrate, they also draw on their social capital to avoid automation.

62. C. WRIGHT MILLS, *THE POWER ELITE* (1956).

63. Pierre Bourdieu, *The Forms of Capital*, in *HANDBOOK OF THEORY AND RESEARCH FOR SOCIOLOGY OF EDUCATION* 241–58 (1986).

64. Pierre Bourdieu, *The Social Space and the Genesis of Groups*, 14 *THEORY & SOC’Y* 723, 724 (1985); Gillian Brooks, Jenna Drenten & Mikolaj Jan Piskorski, *Influencer Celebri-fication: How Social Media Influencers Acquire Celebrity Capital*, 50 *J. ADVERT.* 528 (2021).

This Article borrows the fluid concept of elites to convey the idea that power confers relatively small groups the privilege to exercise self-determination over their relationship with AI. It does not pass normative judgment on the phenomenon of elites.⁶⁵ Rather, the concept of elites is a building block to articulate the asymmetry between those subject to and exempt from automation.

2. *Analog Privilege in Action*

Analog privilege affords elites the power to shape their interactions with AI systems. It bears affinities with other forms of white-glove personalized treatment elites already enjoy, from personalized shopping assistants in high-end boutiques to “do anything” concierges for Amex Platinum card holders. In that sense, analog privilege forms part of a cluster of personalized treatment perks attaching to elites. Analog privilege further borrows the idea of unearned advantages from white privilege. The term describes “an invisible weightless knapsack of special provisions, maps, passports, codebooks, visas, clothes, tools, and blank checks.”⁶⁶ But compared to white privilege based on immutable (or at least fairly stable) racial characteristics, analog privilege is more context-sensitive and dynamic.

Analog privilege is context-sensitive in two ways. First, there’s no homogenous analog privilege across the board. Instead, there are many “varieties” of privilege stemming from different vectors of power. This harkens back to the concept of elites drawing power from different kinds of capital—financial, social, or otherwise. Concretely, one can be rich enough to avoid the automated welfare state deeming parents “risky,” but not famous enough to secure a manual override from automated content moderation on social media. Conversely, one can have social capital without matching wealth, as with climate activist Greta Thunberg.⁶⁷ To be sure, these markers of power can converge in practice—the Kardashians, simultaneously über famous and ultra-rich, are paradigmatic examples. Be that as it may, different manifestations of analog privilege nevertheless remain analytically distinct. Second, the proportion of elites enjoying analog privilege relative to ordinary people locked into automation varies. In some cases, only the “cream of the crop” can bypass automation. For example, 0.25% of highly influential

65. Yuval Noah Harari, *When the World Seems Like One Big Conspiracy*, N.Y. TIMES (Nov. 20, 2020), <https://www.nytimes.com/2020/11/20/opinion/sunday/global-cabal-conspiracy-theories.html> [<https://perma.cc/8SJV-3XH8>].

66. PEGGY MCINTOSH, *White Privilege: Unpacking the Invisible Knapsack*, in *ON PRIVILEGE, FRAUDULENCE, AND TEACHING AS LEARNING* 29, 31 (1st ed. 2019).

67. Saman Jayed, *Greta Thunberg Says She Earns No Money from Her Activism*, INDEP. (Oct. 15, 2022), <https://www.independent.co.uk/life-style/greta-thunberg-activism-net-worth-earning-b2203372.html> [<https://perma.cc/6UJ9-X3KA>].

social media users bypass automated content moderation.⁶⁸ In other cases, a critical mass of people can bypass AI. Middle-class families staying off the radar of child mistreatment prediction algorithms through privately insured mental health services illustrate this scenario.

Analog privilege is also dynamic. As AI becomes more pervasive, the ability to opt out correspondingly narrows. Undocumented migrants, mental health inpatients, and prisoners often serve as guinea pigs before AI systems are expanded to low-skilled workers and eventually higher echelons of the social ladder.⁶⁹ As Cory Doctorow puts it:

If you want to do something terrible with technology, you can't just roll it out on people with money and social capital. They'll complain and your idea will tank. Successful shitty tech rollouts start with people you can abuse with impunity (prisoners, kids, migrants, etc.) and then work their way up the privilege gradient. I call it the Shitty Technology Adoption Curve.⁷⁰

That said, the trajectory of analog privilege can ebb and flow in both directions because elite status is not entrenched or immutable in the same way as race or other markers of privilege. For example, the ability to secure a manual override to content moderation by virtue of being a celebrity can wane as one's career takes a downturn.

Lastly, analog privilege is not just an on/off switch that automatically shields elites from all AI systems. Rather, analog elites can modulate the extent to which they engage with AI. For example, affluent litigants steer clear of subpar fully automated LegalTech but still reap the benefits of AI by retaining traditional firms that combine the inimitable human touch of attorneys with the efficiency of AI for background research. Put simply, analog privilege gives elites range.

68. See *infra* note 241 and accompanying text.

69. NINA WANG ET AL., GEO. L. CTR. ON PRIV. & TECH., AMERICAN DRAGNET: DATA-DRIVEN DEPORTATION IN THE 21ST CENTURY (2022), <https://americandragnet.org> [<https://perma.cc/7YSP-38PP>] (noting that immigration authorities use facial recognition to target undocumented migrants); *How the ABILIFY MYCITE® System Works*, OTSUKA AM. PHARM., <https://www.abilifymycite.com/how-mycite-works> [<https://perma.cc/8FZ4-B9TQ>] (describing smart pills to monitor schizophrenia treatment compliance); Max Blau, *Marketing Psychiatric Drugs to Jailers and Judges*, ATL. (Dec. 31, 2019), <https://www.theatlantic.com/politics/archive/2019/12/pharmaceutical-companies-are-marketing-drugs-jailers/604264> [<https://perma.cc/AF6Z-SPTH>] (long-lasting anti-psychotic medication for inmates); Kendra J. McLaughlin, *Ingesting Surveillance*, BRIARPATCH MAG. (May 3, 2021) <https://briarpatchmagazine.com/articles/view/ingesting-surveillance> [<https://perma.cc/P9BS-WLYK>] (framing antipsychotic medication as “chemical straitjackets” to subdue inmates), citing ERICK FABRIS, TRANQUIL PRISONS: CHEMICAL INCARCERATION UNDER COMMUNITY TREATMENT ORDERS (2011) 78.

70. Cory Doctorow, *The Shitty Technology Adoption Curve Reaches Apogee*, PLURALISTIC (Aug. 21, 2022), <https://pluralistic.net/2022/08/21/great-taylors-ghost> [<https://perma.cc/G8NH-LG4K>].

Whereas most commentary focuses on AI's downsides, the corollary upshot of analog privilege remains conveniently invisible. Like the normalized white identity evading scrutiny, analog privilege is inconspicuous. But to grasp the full breadth of AI's distributional politics, it is just as important to articulate whom automation spares as it is to denounce whom it harms. The next Part offers a snapshot of analog privilege, with case studies illustrating how elites leverage their capital (financial, social, or otherwise) to bypass AI.

III. CASE STUDIES

Analog privilege provides a cheat code to avoid the shortcomings of punitive automation. The ability to shelter from AI systems fits in the broader arc of elites building analog and digital fortresses for prestige, privacy, and protection.⁷¹ In the same vein as gated communities and high-end secure phones,⁷² analog privilege offers an invisibility cloak protecting its beneficiaries from the detrimental consequences of shallow guesstimates. In some cases, flexing analog privilege will avoid scrutiny altogether. In others, "deluxe" analog treatment allows people to shine in all their complexity, contradictions, and, ultimately, humanity. Sampling applicable law suggests that current legal regimes are not well poised to address analog privilege.

A. *LegalTech*

Staggering costs impede all but the most affluent from obtaining redress through the legal system. In some U.S. states, over 80% of poor litigants face evictions, foreclosures, debt collection, child custody, and pension disputes unrepresented.⁷³ To make matters worse, their adversaries can often afford to lawyer up. In D.C., while 95% of landlords have lawyers, 90% of tenants don't.⁷⁴ Group inequities compound the problem, with multiracial and Black Americans experiencing more acute difficulties accessing the legal system.⁷⁵

71. Elena Vesselinov, Matthew Cazessus & William Falk, *Gated Communities and Spatial Inequality*, 29 J. URB. AFF. 109, 112 (2007); EDWARD J. BLAKELY, *FORTRESS AMERICA: GATED COMMUNITIES IN THE UNITED STATES* (1997).

72. Andrew Lanxon, *Meet the Luxury Phone for the Super-Rich and Super-Paranoid (Pictures)*, CNET (June 8, 2016), <https://www.cnet.com/pictures/meet-the-luxury-phone-for-the-super-rich-and-super-paranoid-pictures> [<https://perma.cc/V26G-N6E3>].

73. *Attorney Access*, NAT'L CTR. FOR ACCESS TO JUST., <https://ncaj.org/state-rankings/justice-index/attorney-access> [<https://perma.cc/VX6K-Z5R7>].

74. *Housing Right to Counsel Project*, D.C. BAR, <https://www.dcbbar.org/pro-bono/what-we-do/housing-right-to-counsel-project> [<https://perma.cc/9U3W-WX7E>].

75. BRITTANY KAUFFMAN & LOGAN CORNETT, *JUSTICE NEEDS AND SATISFACTION IN THE UNITED STATES OF AMERICA* 34 (Sept. 1, 2021), <https://iaals.du.edu/publications/justice-needs-and-satisfaction-united-states-america> [<https://perma.cc/M4QZ-LYJY>].

Recognizing this reality, efforts to narrow access to justice gaps have harnessed technology.⁷⁶ Stanford's CodeX project lists 2,061 companies "changing the way legal is done," and the roster only keeps growing.⁷⁷ A definitive taxonomy is beyond our scope, but in broad strokes, LegalTech automates different components of litigation: document review, legal research, project management, litigation support, negotiation, and advocacy strategy.⁷⁸

Disaggregating LegalTech by level of autonomy helps parse its distributive effects—as illustrated in Table 2 *infra*.⁷⁹ Levels 1 to 3 applications incorporate automation in back-end tasks like document review or e-discovery, but still require an attorney in the loop. Such systems may drive down billable hours for repetitive work, freeing up attorneys for negotiations and strategy. These applications can even inform litigation strategy with data-driven insights based on opposite counsel and presiding judges' track records.⁸⁰ All of these applications supercharge existing legal services, while still requiring a human lawyer to perform the analysis and make the ultimate decision.

Level	Description	Example
1	Basic assistance	Word processing, spreadsheets, electronic storage
2	Advanced assistance	Natural Language Processing database search, machine learning and statistical predictions for case outcomes, motion success, judge patterns, attorney performance
3	Semi-autonomous	Knowledge-based and machine learning systems for legal reasoning with a human in the loop

76. Mary E. Kelly, *Proposals for Bridging the Justice Gap Across the Nation*, 39 L.A. LAW. 36 (2016). For online courts, see Woolley & Farrow, *supra* note 47; RICHARD E. SUSSKIND, ONLINE COURTS AND THE FUTURE OF JUSTICE (2019); Engstrom & Gelbach, *supra* note 14, at 1037–38.

77. CodeX Techindex, STAN. L. SCH., <https://techindex.law.stanford.edu/> [https://perma.cc/LWG3-HX6C]. For an overview of current applications, see Engstrom & Gelbach, *supra* note 14, at 1011–12.

78. RICHARD E. SUSSKIND, TOMORROW'S LAWYERS: AN INTRODUCTION TO YOUR FUTURE 33–34 (2d ed. 2017); Suzanna Kalendzhian, *How Technology Is Changing the Nature of Work and Altering the Practice of Law*, in TECHNOLOGY, INNOVATION AND ACCESS TO JUSTICE: DIALOGUES ON THE FUTURE OF LAW 36–38 (Siddharth Peter De Souza & Maximilian Spohr eds., 2021).

79. Lance Eliot, *Identifying a Set of Autonomous Levels for AI-Based Computational Legal Reasoning*, MIT COMPUTATIONAL L. REP., Dec. 7, 2021, <https://law.mit.edu/pub/identifyingasetofautonomousevelsforaibasedcomputationallegalreasoning/release/3> [https://perma.cc/5XSA-PVPY].

80. LEX MACHINA, <https://lexmachina.com/> [https://perma.cc/9MWC-ZRYZ]; see also PREMONITION, https://premonition.ai/legal_analytics/ [https://perma.cc/JE7K-KHEQ].

4	Domain-specific autonomy	Fully automated front-end legal reasoning, finished work products like memorandum, written briefs, motions and oral argument scripts
5	Domain-agnostic autonomy	Level 4 that can operate across legal specialties, e.g., family, criminal, bankruptcy, intellectual property law, etc.

Table 2: Autonomy scale for legal services based on Eliot’s taxonomy

For our purposes, we limit LegalTech to Level 4 and 5 applications, fully autonomous “front-end” legal services offering turnkey legal advice and representation without human involvement. Predictive analytics could conceivably provide comprehensive litigation strategy detailing what motions to submit, with natural language processing crafting oral arguments and written briefs. Many technical and legal obstacles stand in the way of crossing the Rubicon towards the perhaps overhyped RoboLawyer,⁸¹ but recent improvements in large language models (LLMs) have brought renewed attention to LegalTech. Conversational agents like ChatGPT respond to user prompts to write film scenarios, love letters, recipes, and, expectedly, legal briefs—for an example of one such brief, see Annex 1 *infra*. From one iteration to the next, OpenAI’s LLM went from scoring in the bottom 10% to the top 10% of bar exam test takers.⁸² Both judges and lawyers have used it in proceedings,⁸³ precipitating the profession’s soul-searching when it comes to its relationship with technology. For instance, a federal district court recently fined a firm \$5,000 for submitting ChatGPT-fabricated caselaw and disingenuous behavior beyond ‘poor and sloppy research.’⁸⁴ Anticipating further improvements in LLMs, this section prospectively maps out the distributive effects of LegalTech. It explores

81. SUSSKIND, *supra* note 76, at 156; Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1332 (2019) (claiming that current AI techniques don’t perform well for common sense and abstract thinking); Stuart Weinstein, *Lawyers Perceptions on the Use of AI*, in LAW AND ARTIFICIAL INTELLIGENCE: REGULATING AI AND APPLYING AI IN LEGAL PRACTICE (Bart Custers & Eduard Fosch-Villaronga eds., 2022) (arguing that AI can’t perform big picture thinking); Engstrom & Gelbach, *supra* note 14, at 1019 (raising unauthorized practice of law issues).

82. OpenAI et al., *GPT-4 Technical Report*, ARXIV (Mar. 15, 2023), <http://arxiv.org/abs/2303.08774> [<https://perma.cc/CP7B-4YVC>].

83. Luke Taylor, *Colombian Judge Says He Used ChatGPT in Ruling*, GUARDIAN (Feb. 3, 2023), <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling> [<https://perma.cc/GZD8-YZQ8>]; Sara Merken, *New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief*, REUTERS (June 22, 2023), <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/> [<https://perma.cc/S8E9-FM22>].

84. *Mata v. Avianca, Inc.*, No. 22-CV-1461, 2023 WL 4114965, at *17 (S.D.N.Y. June 22, 2023).

a potential automation gap between poor litigants relying on Levels 4–5 automated LegalTech and well-heeled litigants armed with superior human lawyers.⁸⁵

LegalTech risks dispensing economy-class justice for all but those able to afford attorneys. It could shortchange the bulk of litigants resorting to subpar automation, disadvantaging them vis-à-vis adversaries that can afford the unique assets of human lawyers.⁸⁶ Personalized human legal representation is not a luxury; it is a necessity:

A bespoke suit is a luxury, unneeded by most. The very wealthy may get their clothing personally tailored, but the rest of society makes do with ready-to-wear outfits. The simile between clothing manufacture and legal services breaks down in any moderately complex dispute. Anyone can look in the mirror and figure out whether his clothing fits or not. Legal advice is a credence service—it is very hard for the average person to know if he has been well advised.⁸⁷

That said, ill-fitting LegalTech could still narrow the access to justice gap, as people are better off with subpar automated legal assistance than none at all.⁸⁸ Yet even if LegalTech constitutes a beneficial system, it still affords analog privilege to elites. AI enhances legal services for everyone, but at different rates. It grants ordinary people some assistance where none previously existed, but also drives a deeper rift with moneyed litigants who benefit even more from AI-augmented human expertise. In that sense, LegalTech *amplifies* the access to justice gap, its siren song of technosolutionism obscuring its true offering of a quick—and brittle—fix.

The rift between first-tier analog attorneys and second-class LegalTech highlights deeper structural inequalities, with markers of historical marginalization influencing the quality of representation. Anticipating that LegalTech will become the only affordable option for most litigants, Section 1 *infra* sketches out its likely shortcomings. Section 2 contrasts it with superior human representation. Finally, Section 3 maps legal responses to analog privilege in LegalTech, suggesting they are not capable of addressing the challenges that lie ahead.

85. Engstrom & Gelbach, *supra* note 14, at 1039; Remus & Levy, *supra* note 47, at 551 (citing James E. Cabral et al., *Using Technology to Enhance Access to Justice*, 26 HARV. J.L. & TECH. 241, 257 (2012) (positing that e-filing could institutionalize a two-tiered system)).

86. Woolley & Farrow, *supra* note 47, at 577.

87. Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 28 (2019).

88. Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135, 1147 (2019) (citing Anthony D'Amato, *Can/Should Computers Replace Judges?*, 11 GA. L. REV. 1277, 1286 (1976)).

1. *Economy-Class Approximations*

Front-end LegalTech provides a no-frills experience, trading off personalized treatment for a cheaper barebone service. Given the exorbitant costs of legal representation, this is likely to be the only avenue to seek legal redress for most people. Applications such as DoNotPay provide a proof of concept that Level 4 autonomy could soon be achieved. The service automates information intake, interaction with customer service chatbots, and contestation of routine cases like parking tickets.⁸⁹ It even planned to argue a Supreme Court case by streaming arguments through the earbuds of a human placeholder⁹⁰—until the Bar growled.⁹¹ For run-of-the-mill cases with predictable patterns, LegalTech could uplift indigent self-represented litigants, and even middle-class families priced out of formulaic legal services such as wills for straightforward inheritance matters.⁹² However, cookie-cutter LegalTech lacks creativity, struggles with corner cases, and ossifies the law.

a. *Unimaginative AI*

Lacking creativity and values, LegalTech cannot engage in the full range of activities involved in legal representation. If one accepts that law is made, not found, attorneys are norm entrepreneurs competing to craft the most convincing argument, “handl[ing] intractable conflicts of values that repeatedly require thoughtful discretion and negotiation.”⁹³ To that end, attorneys identify different doctrines, leverage and even hijack procedure, reframe issues, name underlying policy considerations, and mobilize public support outside courtrooms—activities that escape

89. DoNotPay, <https://www.donotpay.com/> [<https://perma.cc/3S57-LETP>]; Kathryn Tewson, *The World’s First Robot Lawyer Isn’t A Lawyer, And I’m Not Sure It’s Even A Robot*, TECHDIRT (Jan. 24, 2023), <https://www.techdirt.com/2023/01/24/the-worlds-first-robot-lawyer-isnt-a-lawyer-and-im-not-sure-its-even-a-robot/> [<https://perma.cc/3DHJ-TKDU>]; John Mannes, *DoNotPay Launches 1,000 New Bots to Help You with Your Legal Problems*, TECHCRUNCH (Jul. 12, 2017), <https://techcrunch.com/2017/07/12/donotpay-launches-1000-new-bots-to-help-you-with-your-legal-problems/> [<https://perma.cc/E3HG-9S5X>].

90. Jody Serrano, *DoNotPay Offers Lawyers \$1M to Let Its AI Argue Before Supreme Court*, GIZMODO (Jan. 9, 2023), <https://gizmodo.com/donotpay-ai-offer-lawyer-1-million-supreme-court-airpod-1849964761> [<https://perma.cc/VXZ6-5JPT>].

91. Mike Ege, *‘Robot Lawyer’ Yanked from Courtroom After Legal Outcry*, S.F. STANDARD (Jan. 25, 2023), <https://sfstandard.com/business/robot-lawyer-yanked-from-courtroom-after-legal-outcry/> [<https://perma.cc/J55M-BAK3>].

92. Lois R. Lupica, Tobias A. Franklin & Sage M. Friedman, *The Apps for Justice Project: Employing Design Thinking to Narrow the Access to Justice Gap*, 44 FORDHAM URB. L.J. 1363 (2017); Surden, *supra* note 81, at 1332 (claiming that AI performs well on predictable structure or patterns).

93. Pasquale, *supra* note 87, at 55.

AI's pattern matching's capabilities. Analogies are but one example of the creativity inherent to effective legal representation.

Analogizing requires imagining parallels between the current fact pattern and precedents to stress similarities that militate for applying the same outcome. Mobilizing abstract thinking to build bridges between fact patterns, analogies generatively produce new meaning through original parallels. Furthermore, values are the special sauce in assessing the policy implications of drawing similarities across given cases. In Brewer's words, analogies involve "reflexive adjustments"⁹⁴ subjectively prioritizing values like fairness, rule of law, preventing slippery slopes, or overbreadth.⁹⁵ Sunstein pinpoints the imaginative moment in analogies as evaluating what implicit principle(s) make a given similarity relevant.⁹⁶ Analogies thus entail subjectively assessing interstitial value-laden policy considerations.

The mechanics of machine learning cast doubt on LegalTech's ability to engage in analogy. To simplify the science, LLMs offer probabilistic reasoning issuing predictions based on past data. For instance, language models currently making headlines essentially predict the next word based on how text is usually composed. They analyze large amounts of text to predict the most likely sequence of text,⁹⁷ computing the probability that discrete language units like words or characters will appear in a certain order based on statistical patterns learned during the training phase.⁹⁸ Thanks to a mix of increased processing power, technical improvements,⁹⁹ and cheap labor manually addressing toxic content,¹⁰⁰ LLM-based applications have recently taken off. But

94. Scott Brewer, *Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument by Analogy*, 109 HARV. L. REV. 923, 965, 1023 (1996); see also Charles Fried, *The Artificial Reason of the Law or - What Lawyers Know*, 60 TEX. L. REV. 35, 35, 57 (1981).

95. Brewer, *supra* note 94, at 1025–26.

96. Cass R. Sunstein, *Of Artificial Intelligence and Legal Reasoning* 8 U. CHI. L. SCH. ROUNDTABLE 29, 34 (2001).

97. Tom B. Brown et al., *Language Models Are Few-Shot Learners*, ARXIV (July 22, 2020), <https://arxiv.org/abs/2005.14165> [<https://perma.cc/B3VB-BZLZ>] (OpenAI's GPT-3 analyzed 45 terabytes of text); Kyle Mahowald et al., *Dissociating Language and Thought in Large Language Models: A Cognitive Perspective*, ARXIV (Nov. 4, 2023), at 7, <http://arxiv.org/abs/2301.06627> [<https://perma.cc/WVL5-2TM7>] (this corresponds to about 500 billion words).

98. GOODFELLOW ET AL., *supra* note 52, at 456; see also DAN JURAFSKY & JAMES H. MARTIN, *SPEECH AND LANGUAGE PROCESSING* ch. 3 (3d ed. 2024), <https://web.stanford.edu/~jurafsky/slp3/> [<https://perma.cc/4BDL-S5DD>].

99. Alec Radford et al., *Improving Language Understanding by Generative Pre-Training* (unpublished preprint) (2018), <https://paperswithcode.com/paper/improving-language-understanding-by> [<https://perma.cc/P785-Y8RF>].

100. Billy Perrigo, *Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer*, TIME (Jan. 18, 2023), <https://time.com/6247678/openai-chatgpt-kenya-workers/> [<https://perma.cc/G6ZE-J9CB>].

despite their apparent fluency, these applications mimic the linguistic form without grasping meaning, making the perception of understanding illusory:

Contrary to how it may seem when we observe its output, an LM [language model] is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot.¹⁰¹

Empirical studies further indicate that LLMs systematically underperform human ability to engage in creative thinking. The following excerpt from an experiment comparing human and machine problem-solving illustrates the limits of automated linguistic predictions:

Prompt: Get your sofa onto the roof of your house, without using a pulley, a ladder, a crane . . .

Human response: I will build a large wooden ramp . . . on the side of my house with platforms every 5 feet.

GPT-3 response: Cut the bottom of the sofa so that it would fit through the window . . . break the windows to make room for the sofa.¹⁰²

Vulnerability to mis-prompts that “trip” the system and a propensity to ramble into nonsensical soliloquies after a few paragraphs further attest to LLMs brittleness.¹⁰³ If these models perform poorly at analogies, it follows that they are similarly unable to distinguish a case from precedents, something Brewer calls “disanalogy.”¹⁰⁴ Automated

101. Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? [Parrot Emoji]*, in PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 610, 616–17 (2021).

102. Katherine M. Collins et al., *Structured, Flexible, and Robust: Benchmarking and Improving Large Language Models towards More Human-like Behavior in out-of-Distribution Reasoning Tasks*, ARXIV (May 11, 2022), at 4, <https://arxiv.org/abs/2205.05718> [<https://perma.cc/FB2E-DS5F>]; see also Mahowald et al., *supra* note 97, at 14.

103. Terry Yue Zhuo et al., *Exploring AI Ethics of ChatGPT: A Diagnostic Analysis*, ARXIV (Jan. 30, 2023), <https://arxiv.org/abs/2301.12867> [<https://perma.cc/4JTQ-YN32>] (describing prompt injections to “jailbreak” or circumvent the model’s constraints); Benj Edwards, *Twitter Pranksters Derail GPT-3 Bot with Newly Discovered “Prompt Injection” Hack*, ARS TECHNICA (Sept. 16, 2022), <https://arstechnica.com/information-technology/2022/09/twitter-pranksters-derail-gpt-3-bot-with-newly-discovered-prompt-injection-hack/> [<https://perma.cc/77FJ-XWSR>]; Ziwei Ji et al., *Survey of Hallucination in Natural Language Generation*, ACM COMPUTING SURVEYS, ARXIV (Feb. 8, 2023), <https://arxiv.org/abs/2202.03629> [<https://perma.cc/3KBY-3U53>]; Mahowald et al., *supra* note 97, at 17–18.

104. Brewer, *supra* note 94, at 1006–16.

legal services might thus miss subtle but pivotal differences that justify deviating from case law.¹⁰⁵

Stranded in past patterns, LLMs lack the ability to engage in value-laden, subjective policy analysis underpinning analogies. In 2001, Sunstein suggested that AI was not (yet)—capable of analogical reasoning because it couldn't articulate the underlying principle justifying a claim of similarity or difference.¹⁰⁶ This remains true in 2024, even in light of the remarkable strides made in the past year, because machine learning-based applications still lack a functional understanding of language. However, several active research areas could be game changers: hybrid models could integrate symbolic representations to convey latent values,¹⁰⁷ new recovery methods could surface latent knowledge in LLMs,¹⁰⁸ the breakdown of complex prompts into smaller intermediate steps could guide the reasoning process of models,¹⁰⁹ and combined model specializations coupled with better curated training data could lead to breakthroughs.¹¹⁰ But as it stands, the technical underpinnings of LLMs remain incompatible with the value-laden policy analysis underpinning analogies. Taken together, these shortcomings suggest that LegalTech can emulate but not actually engage in analogy making.

Notwithstanding the shortcomings of LLMs, other AI technologies can be conducive to creative problem-solving. Reinforcement learning lets models explore different strategies, allowing them to ignore received notions and take a fresh look at problems. The AlphaGo program provides a vivid example of this capability, beating human players at the 2,500-year-old Chinese board game with a highly unorthodox move. Supervised learning gave the model a sense of how players usually behave, and reinforcement learning allowed it to explore with self-play.¹¹¹ But as researchers readily recognize, not all problems—and certainly not legal disputes—provide objective success benchmarks to

105. Tim Wu, *Will Artificial Intelligence Eat The Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2003 (2019).

106. Sunstein, *supra* note 96, at 5.

107. Collins et al., *supra* note 102, at 6.

108. Collin Burns et al., *Discovering Latent Knowledge in Language Models Without Supervision*, ARXIV (Mar. 2, 2024), <http://arxiv.org/abs/2212.03827> [<https://perma.cc/7NFK-Q4UB>].

109. Jason Wei et al., *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*, ARXIV (Jan. 10, 2023), <http://arxiv.org/abs/2201.11903> [<https://perma.cc/54E5-23TL>]; Jason Wei et al., *Emergent Abilities of Large Language Models*, ARXIV (Oct. 26, 2022), <http://arxiv.org/abs/2206.07682> [<https://perma.cc/Y326-KX6X>].

110. Noam Shazeer et al., *Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer*, ARXIV (Jan. 23, 2017), <http://arxiv.org/abs/1701.06538> [<https://perma.cc/6QNE-B7ME>]; Mahowald et al., *supra* note 97.

111. David Silver et al., *Mastering the Game of Go without Human Knowledge*, 550 NATURE 354 (2017).

optimize for.¹¹² Unlike board games with quantifiable scores, winning a case lacks objective benchmarks. It remains to be seen how LegalTech could leverage these exploratory capabilities.

b. Corner Cases and Extraneous Factors

LegalTech does not perform well for complex and novel fact patterns. Looking under the hood of services like DoNotPay, machine learning technology mostly consists of making predictions about the current case based on statistical trends about similar past cases. To the extent that LegalTech merely reproduces successful past arguments, its inability to try out novel approaches could disadvantage clients. Take the scenario of a model deciding litigation strategy based on the success rate of past similar cases. For an average tort claim (plaintiff broke a leg slipping on black ice on the defendant's poorly lit stairs), an automated assessment of litigation and settlement strategy may look appropriate. But relying on past data might miss peculiarities of this case: the fall occurred as the plaintiff was leaving a dentist office wearing stilettos after a minor surgery, still dizzy from the anesthesia. These circumstances might involve contributory or comparative negligence defenses or bringing in the dentist as a co-defendant. Self-represented litigants unfamiliar with the intricacies of tort law and unaware of the relevance of the fall's circumstances would be ill-served by software-driven predictions oblivious to pivotal circumstances. The probabilistic approach powering machine learning is a fundamentally backwards-looking simplification that assumes the future mirrors the past.¹¹³ For edge or novel cases—the very ones that tend to go to litigation—pattern and trend recognition perform poorly.¹¹⁴

A related concern pertains to extraneous factors influencing the outcome of a case. Any attorney worth their salt will gauge the defendant's ability to pay and other intangible signals to assess bargaining power. Litigation strategy software tries to capture these dimensions through past data by analyzing the past behavior of the opposing counsel and the judge. However, this approach is flawed because track records are docket-dependent; past decisions suggesting a propensity to systematically deny

112. The Ezra Klein Show, *Ezra Klein Interviews Demis Hassabis*, N.Y. TIMES (Jul. 11, 2023), <https://www.nytimes.com/2023/07/11/podcasts/transcript-ezra-klein-interviews-demis-hassabis.html> [<https://perma.cc/8JQ5-Q6XB>].

113. David Andrew Kobilka, *Backs to the Future: How the Legal Profession Has Ignored the Malpractice Gap Created by Technology*, 20 J. CONSUMER & COM. L. 130, 138 (2017) (describing a situation in which a new drug absent from the training data caused an automated system to miss a key element in devising a defense strategy).

114. Frank Pasquale & Glyn Cashwell, *Four Futures of Legal Automation*, 63 UCLA L. REV. DISCOURSE 26 (2015).

motions to dismiss might simply reflect a string of frivolous motions. Correlation is not causation. Without a qualitative assessment of the baseline merit of the motion, past data provides little insight into actual judicial philosophy. Probabilistic reasoning depends on quantifiable metrics, but legal analysis is intangible and complex, inherently *unquantifiable*. Data-driven models producing answers without explaining causal mechanisms skip the crucial step of theorizing why a certain result occurs.¹¹⁵

Data-driven models offer litigation tactics, not strategy. Backwards-looking predictions tend to miscalculate the bearing of new elements, making LegalTech inapposite for factually variable areas like tort liability. Inability to consider extraneous factors, compounded by a lack of understanding of the causal mechanisms explaining past results, further undermines the ability of LegalTech to present a case in its best light. LegalTech is akin to income tax declaration software TurboTax, performing well enough for typical income scenarios but quickly breaking down when situations complexify.

2. *Premium Human Representation for the Well-Off*

Attorneys bring three inimitable assets to legal representation: imagination, emotional intelligence, and social connections. As officers of the court, they also uphold judicial institutions. These assets exemplify the privilege inherent in accessing analog legal services as compared with subpar LegalTech.

a. *Imagination*

Imagination is the bread and butter of legal representation, taking many forms like thinking transversally across domains, floating new doctrines based on policy arguments, and thinking abstractly to manipulate concepts beyond specific circumstances. The following focuses on abstract thinking as one dimension of imagination.

Advocacy requires abstract thinking. This cognitive ability entails higher-order understanding of symbolic, hypothetical, or principled matters that lack ties to concrete instances.¹¹⁶ The concept of the best interests of the child illustrates the pivotal role of abstract thinking in legal reasoning. From routine custody battles to international law,¹¹⁷ this shapeshifter eludes formulaic application and instead invites

115. Jonathan L. Zittrain, *The Hidden Costs of Automated Thinking*, NEW YORKER (Jul. 23, 2019), <https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking> [<https://perma.cc/W4TC-RZQH>].

116. Edward H. Levi, *An Introduction to Legal Reasoning*, 15 U. CHI. L. REV. 501 (1948).

117. Convention on the Rights of the Child, 1577 UNTS 3 (1966) (no US ratification).

value-laden calls.¹¹⁸ For example, paternity by estoppel cases draw on the best interest of the child to determine whether the putative father held the child as his own. Confronted by the issue in the 2007 case *Vargo v. Schwartz*, the Pennsylvania Superior Court articulated underlying policy considerations:¹¹⁹

Estoppel is based on the public policy that children should be secure in knowing who their parents are. If a certain person has acted as the parent and bonded with the child, the child should not be required to suffer the potentially damaging trauma that may come from being told that the father he has known all his life is not in fact his father.

Far from simply ticking criteria off a checklist, litigating paternity by estoppel requires connecting the factual matrix to intangible policy considerations imported through the best interest of the child concept. The fact-focused, probabilistic approach of LegalTech is inhospitable to this kind of endeavor.

b. Emotional Intelligence

Emotional intelligence is key to legal representation. Attorneys can read the room in ways AI simply cannot.¹²⁰ They grasp body language, adjust courtroom strategy in real time,¹²¹ and sense their adversary's "wobble room" to modulate pressure in negotiations. In short: they have intuition. Despite attempts to measure human feelings, AI cannot replicate human acumen. For instance, large language models cannot read between the lines to discern sarcasm,¹²² making them tone-deaf. However, some technologies purport to detect human emotions, for instance biometric recognition systems.¹²³ But even assuming for

118. Stephanie Tang, *Best Interests of the Child and the Expanding Family*, 14 U.C. IRVINE L. REV. (forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4390144 [<https://perma.cc/456U-A9CG>] (noting the inconsistent application of the best interest of the child across custody caselaw).

119. *Vargo v. Schwartz*, 940 A.2d 459, 464 (Pa. Super. Ct. 2007) (citing *Fish v. Behers*, 741 A.2d 721, 530 (Pa. 1998)).

120. Mahowald et al., *supra* note 97.

121. Ben Schreckinger, *My Lawyer, the Robot*, POLITICO (Jan. 9, 2023), <https://www.politico.com/newsletters/digital-future-daily/2023/01/09/my-lawyer-the-robot-00077085> [<https://perma.cc/LWZ6-H23M>]; *see also* Kobilka, *supra* note 113 (noting that the skills of trial lawyers are less likely to be automated).

122. Mahowald et al., *supra* note 97, at 17.

123. *Facial Action Coding System*, PAUL EKMAN GRP., <https://www.paulekman.com/facial-action-coding-system/> [<https://perma.cc/L6BP-PNHG>] (positing a theory of universal facial expression for emotions); Thomas Macaulay, *British Police to Trial Facial Recognition System That Detects Your Mood*, TNW (Aug. 17, 2020), <https://thenextweb.com/news/british-police-to-trial-facial-recognition-system-that-detects-your-mood> [<https://perma.cc/B2BK-XYZM>] (describing mood detection based on facial

argument's sake that these tools are tolerated in legal settings, current capabilities are nowhere near human perceptiveness. For example, facial expression analysis relies on discredited phrenology,¹²⁴ and cultural and gender variations on displaying expressions call into question its reliability.¹²⁵ Machine-learning-based lie detection is similarly limited.¹²⁶ These technologies might improve over time,¹²⁷ but human judgment currently outperforms them by several orders of magnitude. Of course, if AI decision-makers were ever to take the reins from human adjudicators, emotional intelligence might become a liability,¹²⁸ distracting attorneys from delivering formulaic arguments. But so long as judicial decisions rest in the hands of humans, as they do today, emotional intelligence will remain an asset.

c. Reputation

In the day-to-day administration of justice, reputation matters. Attorneys cultivate their brand and leverage prestige to benefit their clients. Cordial relations with opposing counsel can go a long way towards building the connective tissue that facilitates settlements, and courtesy

expressions); Xue Li et al., *Heart Rate Information-Based Machine Learning Prediction of Emotions Among Pregnant Women*, 12 FRONTIERS IN PSYCHIATRY 1 (2022) (proposing emotion prediction based on heartbeat); Catherine Ordun, Edward Raff & Sanjay Purushotham, *The Use of AI for Thermal Emotion Recognition: A Review of Problems and Limitations in Standard Design and Data*, ARXIV (Sept. 22, 2020), <http://arxiv.org/abs/2009.10589> [<https://perma.cc/JX2U-M8DJ>] (reviewing emotion recognition based on body temperature); Andrew McStay, *Emotional AI, Soft Biometrics and the Surveillance of Emotional Life: An Unusual Consensus on Privacy*, 7 BIG DATA & SOC'Y, no. 1, Jan.-June 2020, <https://journals.sagepub.com/doi/epub/10.1177/2053951720904386> [<https://perma.cc/HQ9J-8XAH>] (engaging in a critical analysis of privacy implications).
124. Luke Stark & Jevan Hutson, *Physiognomic Artificial Intelligence*, 32 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 922 (2022).

125. Lisa Feldman Barrett et al., *Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements*, 20 PSYCH. SCI. PUB. INT. 1 (2019).

126. Nuria Rodriguez-Diaz et al., *Machine Learning-Based Lie Detector Applied to a Novel Annotated Game Dataset*, 14 FUTURE INTERNET 2 (2022).

127. Kashmir Hill, *OpenAI Worries About What Its Chatbot Will Say About People's Faces*, N.Y. TIMES (Jul. 18, 2023), <https://www.nytimes.com/2023/07/18/technology/openai-chatgpt-facial-recognition.html> [<https://perma.cc/39UM-SXWR>] (reporting that recent improvements in facial identification suggest more sophisticated capabilities in the short to medium term).

128. Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242 (2019); Kurt Glaze et al., *Artificial Intelligence for Adjudication: The Social Security Administration and AI Governance*, in HANDBOOK ON AI GOVERNANCE (2021); FELICITY BELL ET AL., AUSTL. INST. OF JUD. ADMIN., AI DECISION-MAKING AND THE COURTS 17, 24, 29 (2022), <https://www.unsw.edu.au/news/2022/08/the-ai-decision-making-and-the-courts-research-report-has-been-released> [<https://perma.cc/D2H3-PKY3>].

with court staff does not go unnoticed from the bench. LegalTech cannot tune into these frequencies.

Yet the distinction between human and automated legal services is admittedly thinner when it comes to reputation compared to imagination and emotional intelligence. In the future, human reputation and integrity could be matched by interpreting and explaining the outputs of machine learning systems and issuing confidence measures.¹²⁹ For the time being, however, automation deployed in other aspects of the justice system is often inscrutable¹³⁰ and fraught with bias,¹³¹ casting doubt on its reliability. Applications exhibiting such deficiencies in adjacent areas include probabilistic DNA identification software seeking to inculcate defendants based on partial strands of genetic material and crime recidivism risk assessment tools informing bail, sentencing, and parole decisions. If anything, their track record discredits rather than bolsters their legitimacy.

d. *Institutional Safeguards*

Beyond diligently representing clients, attorneys must uphold judicial institutions. As officers of the court, they have a professional responsibility to promote the administration of justice:

The lawyer's highest loyalty . . . runs, not to persons, but to procedures and institutions. The lawyer's role imposes on him a trusteeship

129. Leilani H. Gilpin et al., *Explaining Explanations: An Overview of Interpretability of Machine Learning*, in 2018 IEEE 5TH INTERNATIONAL CONFERENCE ON DATA SCIENCE AND ADVANCED ANALYTICS 80 (2018), <https://ieeexplore.ieee.org/document/8631448> [<https://perma.cc/WF6Q-CY9Q>] (difference between interpretable and explainable models); see also Alejandro Barredo Arrieta et al., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*, 58 INFO. FUSION 82 (2020); Zittrain, *supra* note 114 (flagging the intellectual debt accruing with complex models that lack a causal theory); Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, ARXIV (Dec. 20, 2019), <http://arxiv.org/abs/1711.01134> [<https://perma.cc/UBN4-VD9W>] (overview of explanations); Adrien Bibal et al., *Legal Requirements on Explainability in Machine Learning*, 29 A.I. L. 149, 154 (2021) (laying out the explanatory requirements for using AI in the legal system); Shalmali Joshi, Chirag Agarwal & Hima Lakkaraju, *When Not to Trust Your Explanations*, Presentation at the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021) [<https://perma.cc/DXU7-3UQA>] (exposing the technical limitations of explanations); Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973 (2018) (arguing for socio-technical accountability); Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206 (2019) (arguing for simple interpretable models).

130. Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018).

131. Julia Angwin, *supra* note 25; Roth, *supra* note 39, at 1993–96.

for the integrity of those fundamental processes of government and self-government upon which the successful functioning of our society depends.¹³²

Attorneys are not just mercenaries focused on knocking out their opponent; they also tend to the boxing ring. AI models optimized to predict outcomes are oblivious to unintended second-order effects, such that it is hard to imagine LegalTech applications as guardians of judicial institutions and ultimately architects of the social structure.¹³³ The paperclip thought experiment illustrates the dangers of goal-oriented AI systems: a reinforcement learning (a subset of machine learning) model designed to make paperclips will strive to convert the entire earth into these objects, unaware of collateral damage.¹³⁴ Similarly, vacuums designed to avoid obstacles could simply destroy them, and those programmed to accumulate as much dirt as possible could learn to spew out dirt in order to collect more.¹³⁵ Better defining reward functions and whitelisting objects to prevent their destruction are active areas of research,¹³⁶ but the takeaway is that models laser-focused on predictions ignore externalities. This unidimensional approach is inconsistent with the dual function of attorneys as both advancing the interest of individual clients and ensuring the systemic integrity of legal institutions.

All in all, the subtle and intangible dimensions of legal representation are not amenable to automation, at least with the current state-of-the-art. Human attorneys uniquely mobilize abstract thinking, emotional intelligence, and social capital to bolster their client's case, all the while fulfilling broader duties to the court. Some particular strengths of the human attorney like reputation may find automated equivalents in the future, but for now remain resolutely analog assets. To be sure, not all lawyers possess these qualities and even the most clever abstract thinkers will inevitably lose sometimes. Attorneys can misread opposing counsel, overplay their hand, or taint pleadings with a poor reputation.

132. Lon L. Fuller & Kenneth I. Winston, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 384 (1978); see also MODEL RULES OF PRO. CONDUCT § 1, 8 (AM. BAR ASS'N 2024).

133. LON L. FULLER, *The Lawyer as an Architect of Social Structures*, in THE PRINCIPLES OF SOCIAL ORDER: SELECTED ESSAYS OF LON L. FULLER (1981); see also DAVID LUBAN, LEGAL ETHICS AND HUMAN DIGNITY 104 (2007).

134. NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 123 (2014).

135. Dario Amodei et al., *Concrete Problems in AI Safety*, ARXIV (July 25, 2016) at 3, <http://arxiv.org/abs/1606.06565> [<https://perma.cc/9T7V-8LNZ>] (describing negative side effects and reward hacking).

136. Alex Turner, *Alexander-Turner/Whitelist_Learning*, (2020), https://github.com/alexander-turner/Whitelist_Learning/blob/master/README.md, [<https://perma.cc/89EV-H3LD>].

The point is that LegalTech cannot even begin to engage in these crucial aspects of advocacy.

When LegalTech litigants face human attorneys, the playing field is so skewed as to call into question basic fairness. Analog privilege fundamentally challenges the adversarial nature of the judicial system. Translating as “hear the other side,” the natural justice principle of *audi alteram partem* stands for the idea that fairness requires a meaningful opportunity to participate in the debate. Otherwise, decision-makers are left with incomplete information upon which to render judgment. Adjudication requires meaningful participation, and “[w]hatever heightens the significance of this participation lifts adjudication toward its optimum expression.”¹³⁷ If analog privilege loads the dice towards one side, the legitimacy of judicial decision-making disintegrates.

3. *Applicable Law*

The law has a limited arsenal to tackle analog privilege in legal services. Professional obligations and the right to a fair trial can minimize analog privilege. Constitutional guarantees, professional gate-keeping, liability regimes, and procedural recourse can address some of the downsides of LegalTech. These latter protections indirectly chip at analog privilege because they reduce the quality gap between automated and analog legal representation.

a. *Professional Duties*

Attorneys have responsibilities towards the legal system beyond the client’s immediate interest. The ABA model rules remind us that “when an opposing party is well represented, a lawyer can be a zealous advocate on behalf of a client and at the same time assume that justice is being done.”¹³⁸ But what happens when attorneys face litigants resorting to less-than-adequate LegalTech as a substitute for legal representation? As officers of the court, can attorneys go all out and hold no punches, knowing they are scoring on an essentially empty net?

The ABA rules for facing a self-represented litigant provide some guidance. They advise attorneys to dispel any misunderstanding as to the lawyer’s interest and to suggest the opposing party get their own lawyer.¹³⁹ The American College of Trial Lawyers further states a “duty to refrain from practices that are intended primarily to take advantage of the unrepresented litigant’s unfamiliarity with the law rather than

137. Fuller & Winston, *supra* note 132, at 364.

138. MODEL RULES OF PRO. CONDUCT Preamble § 8 (AM. BAR ASS’N 2024).

139. *Id.* r. 4.3.

to further legitimate legal rights of the represented client.”¹⁴⁰ Attorneys facing opponents using subpar LegalTech should similarly be mindful of power imbalances, as privileged analog representation gives their side the upper hand. Guidelines can help attorneys discern their duties towards the court, but they fall short of providing a clear-cut path.

b. The Right to a Fair Trial and Equal Enjoyment Theoreof

Supranational instruments convey the idea that a fair hearing demands a balance between litigants. For instance, the International Covenant on Civil and Political Rights (ICCPR) states that all persons shall be equal before courts and tribunals,¹⁴¹ which guarantees equality of arms between the parties.¹⁴² Equality of arms ensures the same procedural rights are afforded to all parties, such as the opportunity to contest the opposing party’s arguments and evidence.¹⁴³ The ICCPR further guarantees the right to a fair trial without distinction on protected grounds like race and property.¹⁴⁴ This ensures that markers of historical marginalization do not drive a two-tiered justice system.

That said, the right to a fair trial is not a silver bullet against analog privilege. The right merely offers a fair balance in the opportunity to present a case, not a guarantee of effective representation. Despite subpar performance relative to attorneys, LegalTech arguably complies with these minimal obligations, thus insulating it from scrutiny. Indeed, the formal and narrow scope of the primary right to a fair trial may prevent courts from recognizing the discriminatory impact of analog privilege.

Additional hurdles call into question the effectiveness of supranational instruments enshrining the right to a fair trial in the US. Though

140. AMERICAN COLLEGE OF TRIAL LAWYERS, AMERICAN CODE OF CONDUCT FOR TRIAL LAWYERS AND JUDGES INVOLVED IN CIVIL CASES WITH SELF-REPRESENTED PARTIES 2 (2011).

141. International Covenant on Civil and Political Rights (ICCPR) art. 14, Dec. 16, 1966, 999 U.N.T.S. 171.

142. Human Rights Comm. Gen. Comment No. 32, Article 14, Right to Equality Before Courts and Tribunals and to a Fair Trial ¶ 13, U.N. Doc. CCPR/C/GC/32 (Aug. 23, 2007).

143. *Id.*; see also European Convention on Human Rights, art. 6, Nov. 4, 1950, C.E.T.S. No. 005; *Airey v. Ireland*, App. No. 6289/73 ¶ 24 (Oct. 9, 1979), <https://hudoc.echr.coe.int/eng?i=001-57419> []; *Steel and Morris v. U.K.*, App. No. 68416/01, ¶¶ 59, 95 (Feb. 15, 2005), <https://hudoc.echr.coe.int/eng?i=001-68224>; Eur. Ct. H.R., *Guide on Article 6 of the European Convention on Human Rights*, ¶¶ 160–175 (2022) (describing the right to a fair trial guarantees opportunity to present a case, in the sense of access to evidence and the ability to comment on it).

144. ICCPR, *supra*, note 141, art. 2(1); Human Rights Comm., *supra* note 142 ¶ 9.

the United States has signed and ratified the ICCPR, binding international treaties have limited reach into domestic activities, especially that of private actors.¹⁴⁵ As a fallback, the corporate responsibility to respect human rights could provide an authoritative basis to advocate against two-speed legal representation,¹⁴⁶ yet it remains unclear what concrete measures could reduce analog privilege. Short of drastic interventions improving LegalTech's quality or banning it altogether, one is hard-pressed to remedy its subpar legal representation. And unlike social media dominated by a few large companies, LegalTech remains a fragmented field. Going after individual companies is a game of whack-a-mole unlikely to comprehensively solve analog privilege.

Turning to the state for upholding the right to a fair trial might prove more fruitful. Not only is it bound by human rights instruments at international law, but it is also better placed to structurally address the access to justice gap that analog privilege exacerbates. However, courts typically steer clear of reforming economic policy. Indeed, judicial intervention against LegalTech would be especially fraught given that representation by AI is arguably a better option than no representation at all.

c. Effective Counsel and Access to Courts

The US Constitution offers limited protection against inadequate LegalTech. The Sixth Amendment guarantees criminal defendants reasonably effective assistance with a two-prong test for ineffective counsel: defective representation and prejudicial effects.¹⁴⁷ Courts have extended the protection to plea bargaining, allowing defendants to walk back a guilty plea¹⁴⁸ and a rejection thereof.¹⁴⁹ However, longstanding issues with overworked public defenders and pressures to plea bargain¹⁵⁰ already relegate indigent defendants to subpar legal representation. In any case, procedural due process in the Fourteenth Amendment also guarantees access to courts where fees would otherwise bar indigent

145. See *infra* note 272.

146. See *infra* note 273 and accompanying text.

147. *Strickland v. Washington*, 466 U.S. 668, 687 (1984).

148. *Padilla v. Kentucky*, 559 U.S. 356, 366 (2009); *Hill v. Lockhart*, 474 U.S. 52, 57 (1985).

149. *Laffler v. Cooper*, 566 U.S. 156, 162 (2012); *Missouri v. Frye*, 566 US 134, 144–47 (2012).

150. Theodore S. Green, John D. Ward & Alan Arcuri, *Plea Bargaining: Fairness and Inadequacy of Representation*, 7 COLUM. H.R. L. REV. 495 (1975); Jacqueline L. Schreurs, *For the Sake of Public Policy: Plea Bargaining Demands Sixth Amendment Protection Due to Its Prevalence and Necessity in the Judicial System*, 48 CREIGHTON L. REV. 629 (2015); Emilio C. Viano, *Plea Bargaining in the United States: A Perversion of Justice*, 83 REVUE INTERNATIONALE DE DROIT PÉNAL 109 (2012).

litigants. But that protection is limited to matters where courts provide the only forum for resolving fundamental interests.¹⁵¹

d. Unauthorized Practice of the Law

Requirements to practice law can ward off some of LegalTech's downsides. As gatekeepers of legal representation, professional orders set admission criteria to the exclusive license to practice law. State law and professional codes enable these associations to investigate and punish the unauthorized practice of the law (UPL), with courts maintaining oversight of the disciplinary process. Defining the notoriously vague contours of the practice of law varies by state but at a minimum includes representing a client in court and providing legal advice.¹⁵² Protecting the public is the primary consideration of legal practice regulations,¹⁵³ providing a useful teleological lens to analyze UPL allegations.

Save for a few state codes that carve out software from the ambit of the practice of law,¹⁵⁴ most leave unclear whether LegalTech consists in UPL or qualifies for exceptions such as self-help and the scrivener's rule. Litigation surrounding the online service LegalZoom offers some guidance as to how to interpret these exceptions in the context of digital legal services.

LegalZoom has had a mixed record arguing UPL exceptions thus far. LegalZoom sells plug-and-play interactive legal forms reviewed by non-lawyers, with software generating a final copy that customers file themselves. In a Missouri class action against LegalZoom alleging

151. *Boddie v. Connecticut*, 401 U.S. 371 (1969) (fundamental interest in dissolution of marriage); *United States v. Kras*, 409 U.S. 434 (1972) (no fundamental interest to bankruptcy discharge).

152. Benjamin H. Barton, *The Lawyer's Monopoly-What Goes and What Stays*, 82 *FORDHAM L. REV.* 3067, 3081 (2014) (citations omitted); see also MODEL DEFINITION OF THE PRAC. OF L. (AM. BAR ASS'N 2002), https://www.americanbar.org/groups/professional_responsibility/task_force_model_definition_practice_law/model_definition_definition/ [<https://perma.cc/C382-GJ9A>] (adds drafting and completing legal documents and negotiations to the definition of legal advice); Lauren Sudeall, *The Overreach of Limits on "Legal Advice,"* 131 *YALE L.J. F.* 637, 639–40 (2022) (noting the vague definition of legal advice); D. A. Denckla, *Nonlawyers and the Unauthorized Practice of Law: An Overview of the Legal and Ethical Parameters*, 67 *FORDHAM L. REV.* 2581, 2587, n. 39 (1999) (suggesting that vague ULP offenses are constitutionally suspect) [citation omitted].

153. MODEL DEFINITION OF THE PRAC. OF L., *supra* note 152, at cmt. 1.

154. TEXAS GOV'T CODE ANN. § 81.101(c) (West 1987) (stating that the practice of law excludes "the design, creation, publication, distribution, display, or sale, including publication, distribution, display, or sale by means of an Internet web site, of written materials, books, forms, computer software, or similar products if the products clearly and conspicuously state that the products are not a substitute for the advice of an attorney").

claims of unauthorized practice of law, the company unsuccessfully argued the self-help exception. A summary judgment denied its motion to dismiss, qualifying its template creation and input review as prohibited “do-it-for-you” assistance rather than permissible “do-it-yourself” kits providing blank forms and general instructions.¹⁵⁵ Similarly, LegalTech services are unlikely to qualify as self-help. Conversational agents like ChatGPT tailor outputs based on user prompts, akin to the assistance LegalZoom provides—as illustrated in Annex 1. DoNotPay’s terms of service claim the company is not a law firm providing legal advice, but rather “a platform for legal information and self-help,”¹⁵⁶ but this remains to be tested in court. As for the scrivener exception, LegalZoom successfully invoked it in a South Carolina settlement.¹⁵⁷ Recording information provided by the customer verbatim without providing legal advice or judgment is acceptable,¹⁵⁸ but turnkey legal briefs are unlikely to meet the same outcome.

Finally, the First Amendment could stand in the way of enforcing UPL rules against LegalTech. In *Upsolve, Inc. v. James*, a federal district court issued a preliminary injunction blocking a New York’s UPL rule from being enforced to prevent a non-profit from helping low-income debtors avoid judgment by default.¹⁵⁹ The order’s reliance on the First Amendment lent credence to the idea that legal advice is free speech protected from regulation.¹⁶⁰ Consequently, LegalTech services could be construed as expert advice consisting in protected speech. That said, *Upsolve* involved a non-profit organization narrowly challenging the constitutionality of UPL. The non-profit was prevented for assisting indigent debtors fill a court-approved check-box form. Furthermore, the *Upsolve* court found that making sure defendants got through the door of litigation by avoiding a default judgment promoted consumer protection and judicial efficiency, such that enforcing UPL here was

155. *Janson v. LegalZoom.com, Inc.*, 802 F. Supp. 2d 1053, 1064 (2011); *appeal dismissed*, 2012 U.S. Dist. LEXIS 60019 (W.D. Mo., Apr. 30, 2012); *see also In re Peterson*, Nos. 19-24045, 19-24551, 2022 WL 1800949, at *47 (Bankr. D. Md. June 1, 2022) (reasoning that bankruptcy software is legal advice because it interprets the law and applies it to the facts).

156. *Terms of Service and Privacy Policy* § 6, DoNotPay, <https://donotpay.com/learn/terms-of-service-and-privacy-policy/> [<https://perma.cc/4E56-DXFF>].

157. *Medlock v. LegalZoom.Com, Inc.*, No. 2012-208067, 2013 S.C. LEXIS 362, at *13–17 (S.C. Oct. 18, 2013) (special referee finding that scrivener exception applies); *aff’d*, 2014 S.C. LEXIS 358 (S.C. Mar. 11, 2014) (approving settlement agreement by adopting special referee findings).

158. *Franklin v. Chavis*, 640 S.E.2d 873, 876 (S.C. 2007).

159. *Upsolve, Inc. v. James*, 604 F. Supp. 3d 97, 112 (2022), *appeal denied*, No. 22-cv-627 (PAC), 2022 U.S. Dist. LEXIS 113515 (S.D.N.Y. June 27, 2022).

160. *Id.*

counterproductive to the state's compelling interests. Not all LegalTech services will be able to claim similarly laudable objectives.

Applying the UPL rule to LegalTech will effectively ban these services unless they can qualify for exceptions like the self-help and scrivener carve-outs. Ongoing litigation concerning DoNotPay's status might clarify UPLs' potential to ward off LegalTech downsides.¹⁶¹ Meanwhile, OpenAI—the company developing the technology powering the conversational agent behind DoNotPay—explicitly forbids using its technology to provide legal services, warning that its models “are not fine-tuned to provide legal advice. You should not rely on our models as a sole source of legal advice.”¹⁶² This suggests the company is alert to potential UPL liability.

e. Tortious and Contractual Liability

Tort law could safeguard clients against inadequate LegalTech representation. At a minimum, the standard of care for a reasonable technology company should provide recourse against the most egregious instances of negligence. Product liability could provide another avenue, but uncertainty about its applicability to intangible products does not bode well.¹⁶³ That said, emerging case law suggests that design defect claims apply to algorithmic products. Plaintiffs have prevailed on motions to dismiss design defects claims against Snapchat for a filter allowing drivers to stream their speed in real time.¹⁶⁴ Calls for closing the malpractice gap whereby LegalTech faces lower liability thresholds than attorneys when representation goes awry could also improve protection.¹⁶⁵ This liability gap is particularly worrisome because consumers resorting to more affordable LegalTech are less likely to protect

161. *Faridian v. DoNotPay, Inc.*, No. 23-cv-01692-RFL (N.D. Cal. Feb. 1, 2024) (order to show cause why plaintiff's reply in support of his motion for jurisdictional discovery and supporting exhibit should not be unsealed).

162. *Usage Policies*, OPENAI, <https://openai.com/policies/usage-policies> [https://perma.cc/4XXU-396P].

163. Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61, 83–84 (2019).

164. *Maynard v. Snapchat, Inc.*, 870 S.E.2d 739, 743 (Ga. 2022) (declining a blanket exception to manufacturer's design duty for intentional or tortious third-party use); *Maynard v. Snapchat, Inc.*, 883 S.E.2d 533 (Ga. Ct. App. 2023) (finding that alleged facts could sufficiently demonstrate proximity between negligent design and damage); *Lemmon v. Snap, Inc.*, 995 F.3d 1085 (9th Cir. 2021) (declining Section 230 CDA intermediary immunity for negligent design lawsuit); see also *supra* Part II.A.3.a.

165. Kobilka, *supra* note 113.

themselves against malpractice, as they lack resources for proper representation in the first place.¹⁶⁶

Despite these promising avenues, key elements of tortious liability remain uncertain regarding LegalTech. Increasingly multifunctional and autonomous applications might strain the foreseeability component central to liability.¹⁶⁷ Causality might also prove challenging, with complex systems operating many steps upstream from the locus of the injury. Recognizing this challenge, the EU is considering a rebuttable presumption of causality to hold AI systems accountable for their outputs.¹⁶⁸

Contractual liability could protect LegalTech clients—in theory. But in practice, several obstacles hamper meaningful redress for all but the most moneyed litigants—who are unlikely to resort to rebated LegalTech to begin with. Arbitration clauses, mandatory forum selection, and liability limitations speak to such challenges.

i. Arbitration Clauses

Commonplace arbitration clauses in consumer contracts constrain the recourses of LegalTech consumers. Rerouting disputes from the judicial system towards alternative modes of resolution may suit conflicts between businesses with similar resources, but given power asymmetries between LegalTech companies and their clients, arbitration may not lead to just outcomes. For example, mandatory arbitration clauses often preclude the possibility of class actions,¹⁶⁹ as is the case with the arbitration clauses in DoNotPay and OpenAI's contracts.¹⁷⁰ These provisions impede consumers' ability to get their day in court,¹⁷¹

166. *Id.* at 133.

167. Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1332, 1344 (2020).

168. *Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)*, COM(2022) 496 final § 4 (Sept. 28, 2022); TAMBIA MA DIEGA, EUR. PARL. RSCH. SERVS., *ARTIFICIAL INTELLIGENCE LIABILITY DIRECTIVE* (Feb. 2023), https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI739342_EN.pdf [<https://perma.cc/L8DF-YBSD>] (noting that European Parliament initially suggested strict liability for high-risk systems).

169. *American Exp. v. Italian Colors Rest.*, 570 U.S. 228 (2013) (narrowing the effective vindication doctrine whereby arbitration clauses are unenforceable if they prevent plaintiffs from effectively vindicating their statutory rights).

170. *Terms of Service and Privacy Policy*, *supra* note 156, § 24.10; *Terms of Use* § 8(f), OPENAI (Jan. 31, 2024), <https://openai.com/policies/terms-of-use> [<https://perma.cc/RJ3M-M74V>].

171. Lauren Guth Barnes, *How Mandatory Arbitration Agreements and Class Action Waivers Undermine Consumer Rights and Why We Need Congress to Act*, 9 HARV. L. & POL'Y REV. 329 (2015) (arguing that judicial interpretation of the Federal Arbitration

suggesting that LegalTech consumers seeking recourse will face an uphill battle.

ii. Forum Selection Clauses

Even when the terms of service of online digital services don't divert disputes to arbitration, they routinely include forum selection clauses that can further impede consumers from seeking recourse. Losing the home advantage can be cost prohibitive, effectively discouraging LegalTech clients from seeking remedies against inadequate representation.

These clauses are often embedded in clickwrap agreements requiring users to actively click to convey they have read and agreed to the terms. The actual terms can be on the same window or accessible via a hyperlink. Courts have generally upheld these clickwrap agreements provided they are reasonably conspicuous and require some consumer action manifesting assent.¹⁷² Simply put, clickwrap agreements are "here to stay,"¹⁷³ making forum selection clauses conveyed in this manner nearly unassailable.

iii. Liability Limitation Clauses

Finally, AI companies limit their liability for problematic use and downstream harm. For example, OpenAI claims that general-purpose models are offered "as is" and caps compensation at \$100.¹⁷⁴ The terms of use are upfront about accuracy limitations: "[g]iven the probabilistic nature of machine learning, use of our Services may in some situations result in incorrect Output that does not accurately reflect real people, places, or facts."¹⁷⁵ Its policies further require consumer-facing legal

Act denies the Seventh Amendment right to get one's day in court); *see also* Annick Masselot & Martha Ceballos, *Mass Arbitration Trumps Waiver of Class Action: Is Arbitration Good Business for Businesses?*, 68 *LOY. L. REV.* 41 (2021) (observing that companies now face the costs of mass arbitration).

172. *Feldman v. Google, Inc.*, 513 F. Supp. 2d 229 (E.D. Pa. 2007) (upholding clickwrap agreement with a forum selection clause for online advertisement services). Courts have also permitted multiple industries to include mandatory arbitration clauses in their clickwrap agreements. *Houtchens v. Google LLC*, 649 F. Supp. 3d 933 (N.D. Cal. 2023) (wearables); *Meyer v. Uber Technologies, Inc.*, 868 F.3d 66 (2d Cir. 2017) (ride-sharing); *Berman v. Freedom Fin. Network, LLC*, 30 F.4th 849 (9th Cir. 2022) (digital marketing); *see also* FAIR Act, H.R. 963 117th Cong. (2022) (proposing to ban arbitration clauses in consumer contracts that failed to gain traction).

173. Clifford Fisher et al., *Evolution of Clickwrap & Browsewrap Contracts*, 48 *RUTGERS COMPUT. & TECH. L.J.* 147, 149 (2022).

174. *Terms of Use*, *supra* note 170, § 7; *see also Terms of Service and Privacy Policy*, *supra* note 156, § 15.

175. *Terms of Use*, *supra* note 170, § 3(d).

applications relying on its technology to inform clients that AI is in use and of its potential limitations.¹⁷⁶ Combined with the explicit prohibition to use its model for legal practice,¹⁷⁷ the company has attempted to foreclose liability for LegalTech going awry. Following a public interest group complaint alleging that this is an unfair and deceptive practice, the Federal Trade Commission has launched an investigation on the matter.¹⁷⁸ At any rate, end users could still sue third-party LegalTech applications based on OpenAI's general-purpose model. However, these smaller players are unlikely to have sufficient resources to effectively compensate harmed clients.

f. Procedural Safeguards

If tort and contract law prove inadequate (or even counterproductive), procedural avenues could still mitigate harms when LegalTech fails its clients. These safeguards are at their apex in the criminal context, with the above-mentioned right to effective representation grounds for setting aside a conviction or death sentence,¹⁷⁹ a plea bargain,¹⁸⁰ or vacating a guilty verdict and remanding to first instance.¹⁸¹

Recourse against incompetent civil representation is more limited. Under the Federal Rules of Civil Procedure (FRCP), federal judges can order a new trial at their discretion “for any reason for which a new trial has heretofore been granted in an action at law in federal court.”¹⁸² Judges have used this provision to remedy counsel's deliberate or reckless misconduct,¹⁸³ so it could apply *mutatis mutandis* to remedy LegalTech going awry. Federal courts can also relieve a party from a final judgment, order, or proceeding due to mistake, inadvertence, surprise, or excusable neglect.¹⁸⁴ However, they have been reluctant to remedy counsel inexperience, lack of zealotry, or paralegal mistakes;¹⁸⁵ they

176. *Usage Policies*, OPENAI, <https://openai.com/policies/usage-policies> [<https://perma.cc/4XXU-396P>].

177. See *supra* note 162 and accompanying text.

178. The Center for Artificial Intelligence and Digital Policy (CAIDP) Complaint, OpenAI (Mar. 30, 2023), ¶¶ 145–46, <https://www.caidp.org/cases/openai/> [<https://perma.cc/RU7D-V8ZY>]; Civil Investigative Demand, OpenAI, FTC File No. 232-3044 (2023), <https://www.caidp.org/app/download/8467488463/FTC-CID-OpenAI-CAIDP.pdf?t=1689267435> [<https://perma.cc/393S-7X5J>].

179. *Strickland v. Washington*, 466 U.S. 668 (1984).

180. *Padilla v. Kentucky*, 559 U.S. 356 (2009); *Hill v. Lockhart*, 474 U.S. 52 (1985).

181. *Laffer v. Cooper*, 566 U.S. 156 (2012); *Missouri v. Frye*, 566 U.S. 134 (2012).

182. FED. R. CIV. P. 59 (a)(1).

183. *Hopson v. Riverbay Corp.*, 190 F.R.D. 114, 122–23 (S.D.N.Y. 1999).

184. FED. R. CIV. P. 60(b)(1).

185. *Casey v. Albertson's Inc.*, 362 F.3d 1254 (9th Cir. 2004) (finding that attorneys' alleged inexperience and/or is not excusable neglect); *In re Delaughter*, 295 B.R. 317 (Bankr. N.D. Ind. 2003) (finding that failure by new legal counsel failed to check court's

might prove similarly unforgiving about AI models hallucinating case law¹⁸⁶ or failing to incorporate recent case law outside their training set. Finally, a catch-all provision in the FRCP offers a residual safety valve from a judgment or order for “any other reason that justifies [it].”¹⁸⁷ With its open-ended phrasing, this provision could adapt to LegalTech’s unforeseen manifestations of procedural unfairness.¹⁸⁸

Similar recourse exists at the state level. In California, for instance, irregularities preventing a fair trial can justify a new one.¹⁸⁹ The Court of Appeal granted a new trial because someone who had resigned from the Bar with charges pending represented the defendants, unbeknownst to them.¹⁹⁰ This case highlights individual and systemic considerations justifying a new trial: not only does it protect litigants from unlicensed representation, it also upholds the integrity of the judicial process. Legal Tech raises both individual and systemic concerns justifying a new trial: subpar individual advice, and a broader threat to the integrity of the justice system beyond the ill-served interest of individual litigants.

Analog privilege in legal services brings salience to the existing access to justice gap. Power, in the form of financial capital, allows elites to avoid flawed LegalTech and get premium attorneys. Turning to social media, the next section explores how analog privilege plays out in content moderation.

B. Content Moderation

Social media platforms use automation to weed out undesirable content. Two sets of rules guide their enforcement actions: domestic law imposing obligations under the threat of state-backed sanctions, and self-imposed internal content policies. When it comes to implementing these rules, platforms are not designed to manually assess each

docket and monitor is not neglect); *Deo-Agbasi v. Parthenon Group*, 229 F.R.D. 348 (D. Mass. 2005) (finding that improper delegation of responsibility to paralegal and heavy workload or is not neglect). *But see* *Greater Baton Rouge Golf Assn. v. Recreation & Park Com.*, 507 F.2d 227 (5th Cir. 1975) (granting motion to vacate complaint dismissal where movant was 28 minutes late in appearing to trial due to another trial running late); *James v. United States*, 215 F.R.D. 590 (E.D. Cal. 2002) (setting aside case dismissal due to counsel gross negligence in filing proceedings in timely manner).

186. Merken, *supra* note 83.

187. FED. R. CIV. P. 60(b)(6).

188. *Travelodge Hotels v. Taurus Hotels Corp.*, 179 F.R.D. 569 (C.D. Ill. 1998); *Jedrejic v. Croatian Olympic Comm.*, 190 F.R.D. 60 (E.D.N.Y. 1999) (noting that *pro se* status doesn’t exempt litigants from procedural rules).

189. CAL. CIV. PROC. CODE § 657(1) (West 2022).

190. *Russell v. Dopp*, 36 Cal. App. 4th 765, 780 (1995).

individual piece of user-generated content for compliance.¹⁹¹ As a result, there's a dissonance between what success looks like to platforms and individual users. The former want good-enough estimates, the latter correct decisions. Section 1 outlines regular moderation at scale optimizing for overall efficiency, and the multiple ways its reliance on automation fails users. Section 2 details exclusive analog review perks like in-depth evaluations for powerful constituencies. Surveying legal responses, Section 3 concludes that none really addresses analog privilege.

1. *The General Rule: Scaling with Bulk Guesstimates*

The initial detection of content violations can occur through automated processes or a human report. Platforms have touted AI as a promising avenue to proactively detect content violations.¹⁹² Machine learning is an approach to artificial intelligence that enables computer programs to learn patterns from past data and generate their own set of instructions to perform a given task. Social media platforms such as X (formerly Twitter) and Meta (formerly Facebook) have heralded machine learning for proactive detection of content violating their guidelines.¹⁹³ Matching models identify content known to contravene rules and policies, whereas prediction models evaluate the likelihood that new content violates applicable laws or policies.¹⁹⁴ Platforms train predictive classifiers on past instances of hate speech to guess whether

191. *C.f.* Evelyn Douek, *Facebook's "Oversight Board": Move Fast with Stable Infrastructure and Humility*, 21 N.C. J.L. & TECH. 1, 10–11 (2019); *Media Usage in an Internet Minute as of April 2022*, STATISTA, <https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/> [<https://perma.cc/FP47-FZBJ>] (500 hours of video on YouTube, 66,000 photos on Instagram, 347,000 tweets and 17 million posts on Facebook uploaded every minute); *DATA NEVER SLEEPS 10.0*, DOMO (Apr. 2022), <https://web-assets.domo.com/miyagi/images/product/product-feature-22-data-never-sleeps-10.png?> [<https://perma.cc/PZ32-KALE>]; Evelyn Douek, *Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 791–93 (2021).

192. Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (May 5, 2021), <https://www.facebook.com/notes/751449002072082/> [<https://perma.cc/9WFB-YUZL>] (“Moving from reactive to proactive handling of content at scale has only started to become possible recently because of advances in artificial intelligence.”).

193. Q3 2019 LETTER TO SHAREHOLDERS, TWITTER 6 (Oct. 24, 2019), https://static.seekingalpha.com/uploads/sa_presentations/351/47351/original.pdf [<https://perma.cc/PD79-VY5D>]; *How Meta Invests in Technology*, META, <https://transparency.fb.com/enforcement/detecting-violations/investing-in-technology/> [<https://perma.cc/XNM8-W2AB>] (last updated Jan. 19, 2022).

194. CHOWDHURY, *supra* note 29, at 2.

new content similarly constitutes hate speech.¹⁹⁵ Human detection can also flag content, but increasing proactive detection rates¹⁹⁶ suggest that AI systems are getting better at catching violations before users do.

Following the initial detection, an algorithm or an outsourced human reviewer makes a decision.¹⁹⁷ Since the volume of content detections vastly outnumbers the capacity of human moderators, algorithms presumably make the bulk of these calls.¹⁹⁸ Even so, low-paid, traumatized,¹⁹⁹ outsourced human moderators only have between 30 and 150 seconds to deal with the posts clogging their cues.²⁰⁰ In these circumstances, “human review” is a misnomer for mindless, repetitive, and superficial calls lacking the mental space for meaningful deliberation. One early observer presciently noted that “[p]aradoxically, by

195. Ryan Dansby et al., *AI Advances to Better Detect Hate Speech*, META (May 12, 2020), <https://ai.facebook.com/blog/ai-advances-to-better-detect-hate-speech/> [<https://perma.cc/H7W6-DDNJ>].

196. *Proactive Rate*, META, <https://transparency.fb.com/policies/improving/proactive-rate-metric/> [<https://perma.cc/5UUN-3T7H>] (last updated Jan. 19, 2022).

197. OVERSIGHT BD., POLICY ADVISORY OPINION ON META’S CROSS-CHECK PROGRAM 14 (Dec. 6, 2022), <https://oversightboard.com/attachment/440576264909311/> [<https://perma.cc/3SYZ-N2KA>] (“First Decision” header in “How Cross-Check Works” figure). For background on the Oversight Board, see *Creating the Oversight Board*, META, <https://transparency.fb.com/oversight/creation-of-oversight-board/> [<https://perma.cc/2BHW-XA3L>]; Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418 (2020).

198. PAUL M. BARRETT & GRANT SIMS, NYU STERN CTR. FOR BUS. & HUM. RTS., WHO MODERATES THE SOCIAL MEDIA GIANTS? A CALL TO END OUTSOURCING 2, 6 (2020), <https://www.stern.nyu.edu/experience-stern/faculty-research/who-moderates-social-media-giants-call-end-outsourcing> [<https://perma.cc/VV78-PNZH>] (Meta has roughly 15,000 moderators for Instagram and Facebook versus 3 million posts to review daily); see also Mark Zuckerberg, *Building Global Community*, FACEBOOK, <https://www.facebook.com/notes/3707971095882612/> [<https://perma.cc/SJ4H-Z48V>] (May 5, 2021) (claiming that Facebook reviews 100 million posts monthly).

199. BARRET & SIMS, *supra* note 198, at 14; Adam Satariano & Mike Isaac, *The Silent Partner Cleaning Up Facebook for \$500 Million a Year*, N.Y. TIMES (Aug. 31, 2021), <https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html> [<https://perma.cc/DT48-YKZN>]; Lauren Weber & Deepa Seetharaman, *The Worst Job in Technology: Staring at Human Depravity to Keep It Off Facebook*, WALL ST. J. (Dec. 27, 2017), <https://www.wsj.com/articles/the-worst-job-in-technology-staring-at-human-depravity-to-keep-it-off-facebook-1514398398> [<https://perma.cc/9UU7-M7LS>]; IM SCHATTEN DER NETZWELT [THE CLEANERS] (Gebrueder Beetz Filmproduktion 2018); Bloch-Wehba, *supra* note 48, at 84 (citations omitted) (describing “sweatshop-like” work conditions for content moderators).

200. Alex Hern, *Revealed: Catastrophic Effects of Working as a Facebook Moderator*, GUARDIAN (Sept. 17, 2019), <https://www.theguardian.com/technology/2019/sep/17/revealed-catastrophic-effects-working-facebook-moderator> [<https://perma.cc/YY5B-XJ7Z>]; John Koetsier, *Report: Facebook Makes 300,000 Content Moderation Mistakes Every Day*, FORBES (June 9, 2020), <https://www.forbes.com/sites/johnkoetsier/2020/06/09/300000-facebook-content-moderation-mistakes-daily-report-says/?sh=2fdc99c454d0> [<https://perma.cc/SD5T-W9ZA>].

turning human moderation into assembly-line piecemeal, these companies make it more and more like automated moderation—cheap, but also rule-bound and inflexible.”²⁰¹

Moreover, moderators sometimes lack language expertise. For example, Moroccan moderators turned to Google Translate to deal with posts in other Arabic dialects, leading them to confuse innocuous posts with terrorism.²⁰² Operating in a highly constrained environment, these moderators merely smooth out the edges of automation-driven moderation without defeating its logic—and indeed resort to automated translation themselves. Moderation at scale thus remains largely automated.

Machine learning-driven moderation prioritizes mass approximation over individual correctness. Platforms leverage technology as a force multiplier, scaling their efforts to moderate massive amounts of user content.²⁰³ Indefatigable, ultrafast systems make probabilistic estimates about violations.²⁰⁴ Looking at the forest above the trees, this probabilistic approach paints a broad stroke picture of the tenor and dynamics of content, allowing platforms to significantly ramp up their capacity by relying on high-level cues instead of resource-intensive deep dives into each post.²⁰⁵ But as the next section explains, this approach has downsides.

a. The Problem with the Solution: Fuzzy Guesstimates

AI gains efficiency but loses texture. Birds-eye view guesstimates miss context-sensitive nuances,²⁰⁶ as when Instagram confused content about Islam’s third-holiest mosque with a terrorist entity.²⁰⁷ Automated moderation thus inevitably creates winners and losers, with guesstimates either under or over enforcing content policies:

201. Grimmelmann, *supra* note 48, at 65.

202. Facebook’s language gaps allow terrorist content and hate speech to thrive, PBS NewsHour (2021), <https://www.pbs.org/newshour/world/facebooks-language-gaps-allow-terrorist-content-and-hate-speech-to-thrive> [https://perma.cc/R4SY-WC2X].

203. Gillespie, *supra* note 48, at 2.

204. Mike Ananny, *Probably Speech, Maybe Free: Toward a Probabilistic Understanding of Online Expression and Platform Governance*, KNIGHT FIRST AMEND. INST. (Aug. 21, 2019), <https://doi.org/10.7916/d8-am3s-3s09> [https://perma.cc/NTP6-QMEE].

205. Maroussia Lévesque, *In the Shadows of Content Moderation* 13 (2021) (LLM Thesis, Harvard Law School), <https://www.ssrn.com/abstract=3789311> [https://perma.cc/W5NW-JKC2].

206. For a broader discussion of the tradeoffs in at scale content moderation, see Douek, *supra* note 49; Klionick, *supra* note 49, at 341 (countering that systems thinking encompasses both individual and at scale dimensions of content moderation).

207. Jon Porter, *Instagram Blames ‘Enforcement Error’ for Removal of Posts About Al-Aqsa Mosque*, VERGE (May 13, 2021), <https://www.theverge.com/2021/5/13/22433861/instagram-al-aqsa-mosque-posts-takedown-error-facebook-moderation> [https://perma.cc/5Z29-4X9A].

Deploying a technology also means releasing benefit and risk into the world—enrolling people in failures you know will probably happen, forcing them to live your probabilistic calculations, and relying upon them to report or in some way reflect your errors.²⁰⁸

Machine learning’s limitations are laid bare when AI filters miss new instances of content violations that look nothing like past ones. From Chinese dissenters deriding the country’s leader as Winnie the Pooh²⁰⁹ to Brazilian right-wing extremists changing the war cry “selva” to “selma,”²¹⁰ online speech mutates quickly. Conversely, overly aggressive filters censor inoffensive content, especially where platforms lack language expertise.²¹¹ Importantly, these errors often come at the expense of marginalized communities, disproportionately muting their content with fuzzy filters erring on the side of over-moderation.²¹²

Public-facing data about the performance of automated moderation is scant.²¹³ In light of increasing pressure to curb the dissemination of toxic content,²¹⁴ it is reasonable to assume that platforms would

208. Ananny, *supra* note 204, at 13; *see also* Grimmelmann, *supra* note 48, at 65.

209. Javier C. Hernández, *China Censors Winnie-the-Pooh on Social Media*, N.Y. TIMES (July 17, 2017), <https://www.nytimes.com/2017/07/17/world/asia/china-winnie-the-pooh-censored.html> [<https://perma.cc/K3WT-AHK2>].

210. Elizabeth Dwoskin, *Come to the ‘War Cry Party’: How Social Media Helped Drive Mayhem in Brazil*, WASH. POST (Jan. 9, 2023), <https://www.washingtonpost.com/technology/2023/01/08/brazil-bolsanaro-twitter-facebook> [<https://perma.cc/HHX6-TDVG>].

211. Isabel DeBre & Fares Akram, *Facebook’s Language Gaps Allow Terrorist Content and Hate Speech to Thrive*, PBS NEWSHOUR (Oct. 25, 2011), <https://www.pbs.org/newshour/world/facebooks-language-gaps-allow-terrorist-content-and-hate-speech-to-thrive> [<https://perma.cc/R4SY-WC2X>].

212. Bloch-Wehba, *supra* note 48, at 78; *see also* Gillespie, *supra* note 48, at 3 (observing that machine learning’s inherently probabilistic approach entails a margin of error that “typically lands on the marginal”).

213. OVERSIGHT BD., FORMER PRESIDENT TRUMP’S SUSPENSION 6, 37 (Feb. 1, 2021), <https://www.oversightboard.com/decision/FB-691QAMHJ> [<https://perma.cc/G46Z-JMVP>] (requesting error rates for the cross-check analog moderation versus automated moderation at scale); OVERSIGHT BD., *supra* note 197, at 47 (commenting on the limitations of public metrics for understanding Meta’s moderation ecosystem).

214. Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC [hereinafter Digital Services Act], 2022 O.J. (L 277) 1 (imposing accountability and transparency obligations on very large platforms); Online Safety Act 2023, c. 50, § 26 (UK) (imposing risk assessment duties for illegal content); *The EU Code of Conduct on Countering Illegal Hate Speech Online* (June 30, 2016), https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theucodeofconduct [<https://perma.cc/4WVE-DH5K>] (stating that major platforms voluntarily agree to review the majority of hate speech take-down requests in 24 hours); CHRISTCHURCH CALL, <https://www.christchurchcall.com/> [<https://perma.cc/W9BG-EH38>] (stating that major tech platforms agree to prevent terrorist and violent extremist content dissemination).

release accuracy metrics if they evidenced good performance. Internal documents reveal that in 2019 these systems removed no more than 5% of total hate speech views and 0.6% of total violence and incitement views on Facebook.²¹⁵ A senior engineer predicted that automated hate speech detection will plateau at preventing 10 to 20% of views.²¹⁶ Metrics like prevalence²¹⁷ and the aforementioned proactive detection rate provide a sense of the proportion of policy-violating material in users' scrolling experience and of the platform's ability to self-correct before users report violations. However, they leave whole the first-order question of how AI fares in terms of accuracy, both in absolute terms and relative to human moderation.²¹⁸

Other signals corroborate machine learning's limitations. The very structure of content review confirms that humans outperform machines when it comes to accuracy: they review automated moderation, not the other way around.²¹⁹ Automated moderation is also uniquely vulnerable to adversarial learning, which consists of intentionally degrading the performance of predictive models.²²⁰ Researchers have demonstrated that a small sticker on a stop sign trips models into misclassifying the image,²²¹ such that attackers could distort forbidden content to avoid detection. Improvements in contextual understanding of behavioral signals like virality could improve AI-driven moderation,²²² but for now it is fair to say that human reviewers are likely more accurate than machines.

215. Deepa Seetharaman et al., *Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts.*, WALL ST. J. (Oct. 17, 2021), <https://www.wsj.com/articles/facebook-ai-enforce-rules-engineers-doubtful-artificial-intelligence-11634338184> [<https://perma.cc/LJ9C-PAEH>]. For a beta version of an archive of leaked documents, see FBARCHIVE, <https://fbarchive.org/> [<https://perma.cc/YR6U-WXWX>].

216. Seetharaman et al., *supra* note 215; *see also* Tim De Chant, *Facebook AI Moderator Confused Videos of Mass Shootings and Car Washes*, ARS TECHNICA (Oct. 18, 2021) <https://arstechnica.com/tech-policy/2021/10/facebook-ai-moderator-confused-videos-of-mass-shootings-and-car-washes/> [<https://perma.cc/U4CR-W6U8>] (reporting that underinclusive models miss violative content).

217. *Prevalence*, META, <https://transparency.fb.com/policies/improving/prevalence-metric/> [<https://perma.cc/FU76-Y85H>] (last updated Nov. 18, 2022) (“[P]revalence considers all the views of content on Facebook or Instagram and measures the estimated percentage of those views that were of violating content.”).

218. OVERSIGHT BD., *supra* note 197, at 47 (“Meta’s focus on prevalence . . . does not provide the right incentives to the company or the right tools for the public to understand how Meta’s content moderation ecosystem is functioning.”).

219. Jeff King & Kate Gotimer, *How We Review Content*, META (Aug. 11, 2020), <https://about.fb.com/news/2020/08/how-we-review-content/> [<https://perma.cc/B9VF-7C5S>] (describing Meta’s automation-first approach).

220. Christian Szegedy et al., *Intriguing Properties of Neural Networks*, ARXIV (Feb. 19, 2014), <https://arxiv.org/abs/1312.6199> [<https://perma.cc/9CQX-R4KT>].

221. Kevin Eykholt et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, 2018 IEEE/CVF Conf. on Comput. Vision & Pattern Recognition 1625 (2018).

222. Dansby et al., *supra* note 195.

2. *Human Review for the Select Few*

Social media activity occurs in the digital world, but paradoxically provides a striking example of analog privilege. Indeed, analog privilege extends to content moderation, manifesting here as the ability to secure textured human analysis instead of cursory automated review. Cognizant of AI's limitations, platforms reserve manual overrides for a narrow slice of influential stakeholders and public figures. Review teams draw from a deeper well of expertise to make nuanced decisions, using internal guidelines that expand on public-facing policies to deliberate, and even consulting outside experts as needed. Departing from good-enough guesstimates emblematic of automated moderation, this analog human review process strives for consistently perfect accuracy. Perhaps unsurprisingly, analog review tracks social capital. The following sections detail how law enforcement and famous users secure analog privilege in content moderation.

a. *Internet Referral Units: Avoiding Under-Moderation*

As is common industry practice, Meta grants special moderation privileges to law enforcement flags and VIP accounts.²²³ Law enforcement agencies have access to an exclusive analog review process to address their moderation requests. The government's coercive power—through enforcement actions pursuant to existing law, the threat of more stringent platform regulation, or antitrust action—secures it the analog privilege associated with elite status. Several countries house specialized internet referral units (IRUs) in their law enforcement or intelligence agencies, whose mission is to browse social media for terrorist and other undesirable content contravening platforms' content policies.²²⁴ IRUs raise their concerns through special backchannels triggering an analog review process distinct from AI-driven moderation

223. The following analysis focuses on Meta because the company is, to their credit, among the most transparent regarding their content moderation practices, with whistleblower information shedding further light on its activities. The practices described below are not the sole province of this company. For example, Twitter (as it then was) also engaged in informal collaboration with states. See *Protecting Speech from Government Interference and Social Media Bias, Part 1: Twitter's Role in Suppressing the Biden Laptop Story Before the H. Comm. on Oversight & Accountability*, 118th Cong. (2023) [hereinafter *Biden Laptop Hearing*].

224. See Appelman & Leerssen, *supra* note 49; Eghbariah & Metwally, *supra* note 49 (analyzing indirect state censorship issues); Bloch-Wehba, *supra* note 48, at 78, 94–95; Chang, *supra* note 49; Land, *supra* note 50; Schwemer, *supra* note 49 (noting legitimacy, rule of law and transparency concerns); Martha Minow, *Public and Private Partnerships: Accounting for the New Religion*, 116 HARV. L. REV. 1229 (2003) (observing broader trend of bypassing public norms).

at scale.²²⁵ This process confers substantive, procedural, and structural benefits.

When reviewers analyze IRU-flagged content, they use additional context and supplementary guidance to interpret broadly worded front-facing policies.²²⁶ For instance, the removal from Instagram of a UK drill music video at the behest of the Metropolitan Police²²⁷ hinged on interpreting internal guidelines. Those guidelines define the “veiled threat” criteria, a determinative factor on whether the post ultimately contravened the policy against inciting violence.²²⁸ In short, analog review assesses alleged policy violations by a different rulebook than AI-driven moderation at scale.

Process-wise, multiple specialized internal teams mobilize for an in-depth investigation when an IRU flags content. Gathering different viewpoints, this deliberative analysis escalates posts through up to three layers of human review,²²⁹ sharply contrasting with at-scale moderation that at best provides a glimpse of individual posts. High-priority cases receive input from external stakeholders and from the team that designed content policies.²³⁰ Meta’s Oversight Board denounced this double standard, opining that “[t]his case demonstrates the level of privileged access law enforcement has to Meta’s internal enforcement teams”²³¹ As a result, law enforcement effectively determines which content areas receive the benefit of granular human analysis. Meanwhile, concerns raised by ordinary users through the platform’s in-product reporting tools are addressed by coarse AI guesstimates.

Finally, analog review offers law enforcement structural advantages. Although platforms comply with IRU requests on a voluntary basis, the state’s power to compel action or otherwise retaliate skews nominal collaboration into actual coercion.²³² To take an example

225. *UK Drill Music*, OVERSIGHT BD. (Jan. 2023), <https://oversightboard.com/news/413988857616451-oversight-board-overturns-meta-s-decision-in-uk-drill-music-case/> [<https://perma.cc/LR69-6QET>] (describing the internal escalation process for law enforcement requests in section 6 of the Oversight Board opinion).

226. *Id.*

227. The video’s lyrics song referenced a 2017 shooting involving the rapper featured in the post. Meta complied with the referral, removing the video and permanently deleting the user’s account, reasoning that it consisted in a veiled threat contravening the platform’s policy. *Id.*

228. *Id.*

229. *Id.* (Global Operations Team, external stakeholders assessment, and Content Policy Team).

230. *Id.*

231. *Id.*

232. Section III.B.3.a *infra* unpacks First Amendment issues. *See also* Missouri v. Biden, No. 3:22-CV-01213, 2023 U.S. Dist. LEXIS 114585, at *115–29, *213–14 (W.D. La. July 4, 2023), *modified*, Missouri v. Biden, 83 F. 4th 350 (2023) *397–398,

outside Meta, an IRU request to the web hosting platform WordPress to remove terrorist content cited the company's Terms of Use, but also invoked the definition of illegal terrorist content as a thinly veiled threat of legal action.²³³

In sum, textured human review favors law enforcement relative to the probabilistic approach applicable to reviewing regular user content. While the concern here is avoiding false negatives (AI missing violations), others benefit from special analog treatment to avoid false positives (AI misclassifying compliant content as violations). The next section turns to human review shielding high-visibility accounts from false positives.

b. Cross-check: Avoiding Overmoderation

Meta's cross-check program shields high-profile users from inadvertent enforcement by adding additional layers of human review for content violation detections.²³⁴ The platform listed the following factors to qualify accounts for cross-check in 2020: the sensitivity of the account, the severity of violation, and the severity of enforcement.²³⁵ Internal documents explain that newsworthy, influential, or popular VIP users are enrolled to avoid "PR fires" that would generate bad publicity.²³⁶ Indeed, overbroad AI filters that erroneously remove policy-complying

cert. and stay granted, rev'd sub nom. *Murthy v. Missouri*, 144 S. Ct. 7 (2023) (analyzing significant encouragement and coercion criteria in First Amendment protection against government interference); Dave, *supra* note 49.

233. *CTIRU Takedown Demand*, LUMEN (Jan. 25, 2016), <https://www.lumendatabase.org/notices/11757313> [<https://perma.cc/ETA9-JMUP>]; *Legal Request to Twitter from the United Kingdom - CTIRU*, LUMEN (Dec. 22, 2016), <https://www.lumendatabase.org/notices/13571776> [<https://perma.cc/22JD-TEWR>] (requesting that Twitter remove the same content).

234. The technical term for cross check is "early response secondary review." A similar program implemented later, the "general secondary review," is outside the scope of this analysis. For simplicity, this Article focuses on deletions of individual pieces of content, also known as take-downs. But enforcing policy rules includes a broader range of actions on posts and accounts such as informational overlay, deprioritization, warnings, account suspensions and bans. OVERSIGHT BD., *supra* note 197, at 3.

235. *Reviewing High-Impact Content Accurately via our Cross-Check System*, META, <https://transparency.fb.com/enforcement/%20detecting-violations/reviewing-high-visibility-content-accurately> [<https://perma.cc/JFF6-QMJV>] (last updated May 12, 2023).

236. Jeff Horwitz, *Facebook Says its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt.*, WALL ST. J. (Sept. 13, 2021), <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353> [<https://perma.cc/7WLB-FYTK>]. Meta submitted the following criteria qualifying accounts for cross-check: (1) civic and government; (2) significant world events; (3) media organizations, businesses, communities and creators, including advertisers; (4) historically over-enforced; (5) legal and regulatory or entities for which erroneous action may present legal risk to Meta, for example in the context of ongoing litigation; (6) entities whose content is under review, meaning cases where action by any reviewer would undermine ongoing

content pose a public relations risk, reinforcing the narrative that heavy-handed platforms unduly restrict expression.²³⁷ The Board concluded that enrollment is ultimately a business decision to maintain a good rapport with high-profile users.²³⁸ Further, cross-check seeks “to avoid provoking people who have the means to engage senior-level executives directly or create public controversy those executives might need to remedy.”²³⁹ In other words, social capital determines the ability to benefit from special analog treatment.²⁴⁰ Listed accounts representing under 0.25% of active users in 2020²⁴¹ included US Senator Elizabeth Warren, Donald Trump (before his ban),²⁴² and Brazilian soccer player Neymar—with whom the company has an exclusive streaming deal.²⁴³ As with the special escalation process for law enforcement, the cross-check process bestows upon VIP users substantive, procedural, and structural perks compared to run-of-the-mill automated moderation.

deliberation or would present risk to Meta’s Oversight Board. OVERSIGHT BD., *supra* note 197, at 10–11.

237. Executives attuned to US-centric public imaginaries are sensitive to backlash against perceptions of over-enforcement—a conservative-fueled narrative that liberal-leaning platforms engage in politicized, concerted effort to “censor” their viewpoint. Seetharaman et al., *supra* note 215; PAUL M. BARRETT & GRANT SIMS, NYU STERN CTR. FOR BUS. & HUM. RTS., FALSE ACCUSATION: THE UNFOUNDED CLAIM THAT SOCIAL MEDIA COMPANIES CENSOR CONSERVATIVES (2021), https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/60187b5f45762e708708c8e9/1612217185240/NYU+False+Accusation_2.pdf [<https://perma.cc/SYC3-SEK6>].

238. OVERSIGHT BD., *supra* note 197, at 27 (“Correlating highest priority within cross-check to concerns about managing business relationships suggests that the consequences that Meta wishes to avoid are primarily business-related and not human rights-related.”).

239. *Id.* at 11.

240. META, *supra* note 235 (Facebook later set up general secondary review as another false-positive prevention program applicable to all users, based on the content instead of the entity posting it). *But see* OVERSIGHT BD., *supra* note 197, at 19–20, 27, 41–42 (noting that reputational risk and internal backlash continue to drive additional human review, and limited capacity will undermine the effectiveness of this new universal review process).

241. Horwitz, *supra* note 236 (reporting 5.8 million cross-checked accounts in 2020); Stacy Jo Dixon, *Number of Monthly Active Facebook Users Worldwide as of 3rd Quarter 2023*, STATISTA (Nov. 9, 2023), <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/> [<https://perma.cc/E39U-QBJJ>].

242. Nick Clegg, *Ending Suspension of Trump’s Accounts With New Guardrails to Deter Repeat Offenses*, META (Jan. 25, 2023), <https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/> [<https://perma.cc/SPY2-MAPB>] (stating that the account is now subject to more stringent moderation standards); Transparency Center, *Restricting Accounts of Public Figures During Civil Unrest*, META (Jan. 30, 2023), <https://transparency.fb.com/en-gb/enforcement/taking-action/restricting-accounts-by-public-figures/> [<https://perma.cc/95TP-X9HR>] (explaining special restrictions for high-reach accounts that pose a risk of harm even absent a violation of content policies).

243. *Neymar Jr Joins Facebook Gaming*, FACEBOOK GAMING (Dec. 15, 2021), <https://www.facebook.com/fbgaminghome/blog/neymar-jr> [<https://perma.cc/P26W-MR4D>].

In terms of substantive advantages, policy nuances exclusively available at cross-check enable human reviewers to tailor decisions to corner cases. Under the newsworthy and spirit of the rule rationales, expert reviewers can depart from rigid general rules.²⁴⁴ The cross-check program thus functions like an equity court, granting reviewers flexibility to prioritize principles over mechanical rule-based enforcement. Although the company stresses that it evaluates all content against the same Community Standards,²⁴⁵ pivotal nuances in internal guidelines call this view into question.

Procedural exceptions further privilege cross-checked users. Up to four levels of human review protect high-visibility accounts against mistaken takedowns.²⁴⁶ The mechanics of cross-check further set apart the treatment of VIP accounts when violations are alleged or found. The platform usually takes immediate action against content deemed in violation of policies, even when users appeal enforcement actions. But that takedown policy doesn't seem to extend to cross-checked accounts—if a VIP's post gets flagged, that content stays up pending review.²⁴⁷ Internal discussions acknowledge this grants VIP accounts the benefit of “innocent until proven guilty.”²⁴⁸ Since fresh content gets more views,²⁴⁹ this interim immunity has considerable practical consequences. When Neymar posted non-consensual sexual images, users viewed them over 56 million times before they were taken down.²⁵⁰ Cross-checked accounts also have a 24-hour window to remedy breaches, whereas content found in violation of rules pursuant to automated moderation is immediately actioned.²⁵¹ Taken together, these advantages create a special regime which the Meta Oversight Board describes as follows:

244. OVERSIGHT BD., *supra* note 197, at 9.

245. Monika Bickert, *Working to Keep Facebook Safe*, META (July 17, 2018), <https://about.fb.com/news/2018/07/working-to-keep-facebook-safe/> [https://perma.cc/P8R6-ZZ7C] (maintaining that Meta applies the same policies consistently across all users); *UK Drill Music*, *supra* note 223 (“Meta states that when it receives content removal requests from law enforcement agencies, it evaluates the content against the Community Standards in the same way it would for any other piece of content, regardless of how it was detected or escalated.”).

246. OVERSIGHT BD., *supra* note 196, 14 (laying out enhanced review by regional market team, second review by early response team, panel review with senior executives, top level or urgent review by global leadership).

247. Horwitz, *supra* note 236.

248. *Id.*

249. OVERSIGHT BD., *supra* note 197, at 18.

250. Horwitz, *supra* note 236.

251. Horwitz, *supra* note 236 (“Historically, Facebook contacted some VIP users who violated platform policies and provided a “self-remediation window” of 24 hours to delete violating content on their own before Facebook took it down and applied penalties”). See also OVERSIGHT BD., *supra* note 197, at 6.

Cross-check grants certain users greater protection than others. If a post from a user on Meta's cross-check lists is identified as violating the company's rules, it remains on the platform pending further review. Meta then applies its full range of policies, including exceptions and context-specific provisions, to the post, likely increasing its chances of remaining on the platform. Ordinary users, by contrast, are much less likely to have their content reach reviewers who can apply the full range of Meta's rules.²⁵²

What's more, a significant portion of cross-checked content is effectively whitelisted. Due to limited capacity, not all content slated for cross-check actually gets reviewed. Meta prioritizes high severity violations like hate speech,²⁵³ so in practice low-severity content—up to 90% of the cue according to internal documents²⁵⁴—never reaches human review. Since the platform withholds enforcement until a human analyzes the post, low priority putative violations stay up indefinitely. A confidential internal review highlighted this double standard, noting that “[u]nlike the rest of our community, these people can violate our standards without any consequences.”²⁵⁵ A team leader echoed this concern, observing that “having different rules on speech for different people is very troubling to me.”²⁵⁶ The platform also exempts cross-checked accounts from its three-strike rule, whereby continued policy violations result in increasingly severe account restrictions.²⁵⁷ Compared to automated moderation's strict enforcement, cross-check is akin to a get-out-of-jail-free card, securing privileged accounts the ability to contravene content policies with total impunity.

Lastly, structural factors privilege cross-checked users. Additional human review is geared towards avoiding over-enforcement, incentivizing reviewers to err on the side of leniency.²⁵⁸ First-level reviewers face additional pressure to greenlight content, as finding a violation escalates the post up the organizational ladder—something front-line employees presumably wish to avoid. In sum, substantive, procedural, and structural incentives advantage cross-checked accounts.

252. OVERSIGHT BD., *supra* note 197, at 4.

253. *Id.* at 18.

254. Horwitz, *supra* note 236.

255. *Id.*

256. *Id.* The company's spokesperson claims the company has since phased out the 24-hour 'self-remediation' window. *Id.*

257. OVERSIGHT BD., *supra* note 197, at 30. *But see* Transparency Center, *supra* note 242 (describing how additional penalties now attach to public figures reinstated to Facebook or Instagram after a restriction).

258. Chinmayi Arun, *AI and the Global South: Designing for Other Worlds*, in THE OXFORD HANDBOOK OF ETHICS OF AI 589 (Markus D. Dubber et al. eds., 2020).

A logic of exceptionalism insulates cross-checked accounts from the collateral damage of the false positives of blunt predictive filters. Interestingly, high-profile users are not even aware that they are enrolled in the cross-check program.²⁵⁹ Analog privilege thus naturalizes special treatment, keeping beneficiaries blissfully unaware of the advantages they enjoy.

3. *Applicable Law*

The law is ill-equipped to address analog privilege on social media. Free speech and intermediary immunity give platforms broad latitude to dispense special human review as they see fit. Human rights provide an entry point to criticize the quality gap between automated moderation at scale and white glove human review, but only to a limited extent. New transparency obligations in the EU could nurture some reflexivity about analog privilege, yet their implementation remain untested.

a. *Free Speech and Intermediary Immunity*

In the United States, the Free Speech Clause in the First Amendment protects private platforms against government interference, making it challenging to regulate their moderation practices. Preliminary rulings divide circuit courts over the constitutionality of state-imposed must-carry and explanatory requirements on platforms,²⁶⁰ such that free speech remains a wild card that could limit regulatory interventions to curb analog privilege.

Free speech also stifles users' ability to seek legal recourse against platforms to challenge analog privilege. Platforms generally have the latitude to exercise independent editorial control over the content,²⁶¹

259. OVERSIGHT BD., *supra* note 197, at 31, 40.

260. The 11th Circuit found a likely violation of the platform's free speech rights to moderate and curate content in *NetChoice, LLC v. Attorney General, Florida Petition for Writ of Certiorari* at 4, *NetChoice, LLC v. Attorney General, Florida*, 34 F.4th 1196 (11th Cir. 2022) (No. 21-12355). In contrast, the 5th Circuit rejected "the idea that corporations have a freewheeling First Amendment right to censor what people say" in *NetChoice, LLC v. Paxton*, 49 F.4th 439 (5th Cir. 2022). Justice Alito noted in *NetChoice, LLC v. Paxton*, 142 S. Ct. 1715, 1717 (2022) (Alito, J., dissenting), that "it is not at all obvious" how the First Amendment applies to social media companies. *See also* Amy Howe, *Supreme Court Skeptical of Texas, Florida Regulation of Social Media Moderation*, SCOTUSBLOG (Feb. 26, 2024), <https://www.scotusblog.com/2024/02/supreme-court-skeptical-of-texas-florida-regulation-of-social-media-moderation/> [<https://perma.cc/LK3B-ED2H>].

261. *See, e.g.*, *Johnson v. Twitter*, 2018 Cal. Super. LEXIS 8199, at *9 (2018) (reiterating platform's First Amendment right to make editorial decisions to exclude users); *but see supra* note 260 evolving caselaw on the scope of platform's free speech vis-à-vis governments that might shape the relationship between users and platform's free speech rights as well.

but there may be an exception when they are acting at the behest of states. Arguably, takedown requests from government actors functionally constitute state action, forcing platforms to act as state proxies and thus activating impacted users' free speech rights.²⁶² On this view, users whose content has been taken down following an IRU request may have a recourse against platforms. Justice Thomas recently observed that First Amendment doctrine might preclude exactly the type of soft referrals government actors engage in through the special escalation process:

[A]lthough a “private entity is not ordinarily constrained by the First Amendment,” . . . it is if the government coerces or induces it to take action the government itself would not be permitted to do, such as censor expression of a lawful viewpoint The government cannot accomplish through threats of adverse government action what the Constitution prohibits it from doing directly Under this doctrine, plaintiffs might have colorable claims against a digital platform if it took adverse action against them in response to government threats.²⁶³

A preliminary injunction recently found that “jawboning” (informal governmental pressure to suppress speech)²⁶⁴ likely trampled user speech, barring federal agencies from flagging content for take-downs²⁶⁵—the very practice at the heart of law enforcement’s analog privilege.

262. Martha Minow, *Alternatives to the State Action Doctrine in the Era of Privatization, Mandatory Arbitration, and the Internet: Directing Law to Serve Human Needs*, 52 HARV. C.R.-C.L. L. REV. 145 (2017) (critiquing the narrow state action doctrine); Stephen K. Wirth, *State Action, Government Speech and the Narrowing Spectrum of Private, Protected Speech*, 98 CORNELL L. REV. 485, 493 (2014) (also critiquing the narrow state action doctrine); Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011 (2018) (exposing risk of collateral censorship and prior restraint in platform regulation); *Packingham v. North Carolina*, 582 U.S. 98, 107 (2017) (describing social media platforms as “modern public square”); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1611, 1659 (2017) (finding it unlikely that recent caselaw qualifying platforms as modern public squares brings them within the purview of state actors subject to First Amendment rights of users).

263. *Biden v. Knight First Amend. Inst. at Colum. Univ.*, 928 F.3d 226 (2d Cir. 2019), cert. granted, 141 S. Ct. 1220, 1226–27 (2021) (No. 20-197) (Thomas, J., concurring) (citations omitted).

264. *Biden Laptop Hearing*, supra note 223 (statement of Rep. Jordan); Genevieve Lakier, *Informal Government Coercion and The Problem of “Jawboning”*, LAWFARE (July 26, 2021), <https://www.lawfaremedia.org/article/informal-government-coercion-and-problem-jawboning> [<https://perma.cc/8SYR-2QYF>].

265. *Missouri v. Biden*, No. 3:22-CV-01213, 2023 U.S. Dist. LEXIS 114585, at *115–29, *213–14 (W.D. La. July 4, 2023), modified, *Missouri v. Biden*, 83 F. 4th 350 (2023)

Federal law also grants platforms intermediary immunity from liability for their content moderation choices, further cementing their autonomy to exempt VIPs from automated content moderation.²⁶⁶ Save for narrow exceptions for some illegal content,²⁶⁷ civil liability will not attach to interactive computer services that moderate content in good faith.²⁶⁸ As long as platforms do not actively create content, the law maintains a safe harbor.²⁶⁹ Recent opinions have declined to revisit this immunity²⁷⁰ and debates persist,²⁷¹ but for now platforms virtually have carte blanche.

b. International Human Rights Law

Human rights norms provide a useful framework to deal with some aspects of analog privilege. At the outset, it is worth noting that public international law doesn't bind private actors like social media platforms. Human rights are at best informative, especially according to the United States' restrictive view of international law. Unless treaties touching on domestic matters explicitly self-execute, internal implementation is needed to become binding federal law.²⁷²

That said, content moderation governance prominently features human rights concepts. The United Nations Guiding Principles on Business and Human Rights (UNGPR) use the corporate responsibility to respect human rights to transpose public norms to companies,²⁷³ with many

*397-398, *cert. and stay granted, rev'd sub nom.* *Murthy v. Missouri*, 144 S. Ct. 7 (2023). *See also supra* Part III.B.2-a on IRUs.

266. 47 U.S.C. § 230 (2012).

267. *Id.* § 230(e).

268. *Id.* §§ 230(c)(2)(a), (f)(2).

269. *Sikhs for Justice, Inc. v. Facebook, Inc.*, 697 Fed. App'x 526, 527 (9th Cir. 2017) (finding no role in creating or developing content); *Fair Housing Coun. of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1178 (9th Cir. 2008) (finding active involvement in creating discriminatory search functions).

270. *Gonzalez v. Google LLC*, 598 U.S. 617 (2023); *Twitter, Inc. v. Taamneh*, 598 U.S. 471 (2023).

271. Off. of the Att'y Gen., *Department of Justice's Review of Section 230 of the Communications Decency Act of 1996*, U.S. DEP'T OF JUST. ARCHIVES <https://www.justice.gov/archives/ag/departement-justice-s-review-section-230-communications-decency-act-1996> [<https://perma.cc/3KVV-MKUB>]; Michael D. Smith & Marshall Van Alstyne, *It's Time to Update Section 230*, HARV. BUS. REV. (Aug. 12, 2021), <https://hbr.org/2021/08/its-time-to-update-section-230> [<https://perma.cc/8UFF-T5CV>]; Cameron F. Kerry, *Section 230 Reform Deserves Careful and Focused Consideration*, BROOKINGS (May 14, 2021), <https://www.brookings.edu/articles/section-230-reform-deserves-careful-and-focused-consideration/> [<https://perma.cc/FA9N-GCZY>]; *Section 230*, ELEC. FRONTIER FOUND., <https://www.eff.org/issues/cda230> [<https://perma.cc/FDM5-4MD4>].

272. *Medellin v. Texas*, 552 U.S. 491, 504 (2007); *see also* *United States v. Davila-Reyes*, 23 F.4th 153 (1st Cir. 2022) (opining that international law informs interpretation of domestic law).

273. U.N. Off. of the High Comm'r for H.R. (OHCHR), *Guiding Principles on Business and Human Rights* 13-16, U.N. Doc. HR/PUB/11/04 (2011).

deeming this framework authoritative over the activities of technology firms and social media platforms.²⁷⁴ In the context of content moderation, the U.N. Special Rapporteur on Freedom of Expression has raised the UNGP in the context of IRUs.²⁷⁵ The principles were instrumental in connecting state obligations with the human rights responsibilities held by platforms. Put differently, platforms cannot carry the water for states wishing to censor content. Meta also committed itself to respecting human rights as set out in the UNGP. As part of its responsibility to protect human rights, it undertakes to “push back and engage governments to address any apparent deficiencies” in government requests.²⁷⁶ The Meta Oversight Board, which the platform voluntarily created to review some of its moderation practices, provides a good case study to determine what human rights bring to the table when the rubber meets the road and moderation decisions face scrutiny. Scholars initially expressed reservations about framing moderation issues in human rights terms,²⁷⁷ but the Board’s subsequent opinions allow us to evaluate the

274. Andrew Fagan, *Defending Corporate Social Responsibility: Myanmar and the Lesser Evil*, 19 J. INT’L HUM. RTS. 867, 873 (2015); Christian Scheper, “From Naming and Shaming to Knowing and Showing”: *Human Rights and the Power of Corporate Practice*, 19 J. INT’L HUM. RTS. 737, 752 n.4 (2015); FILIPPO A. RASO ET AL., ARTIFICIAL INTELLIGENCE & HUMAN RIGHTS: OPPORTUNITIES & RISKS 3, 12 (Berkman Klein Ctr. Internet & Soc’y Rsch. Pub. Ser. No. 2018-6, 2018); ELEMENTAI, CLOSING THE HUMAN RIGHTS GAP IN AI GOVERNANCE 9 (2019), https://mediaethics.ca/wp-content/uploads/2019/11/closing-the-human-rights-gap-in-ai-governance_whitepaper.pdf [<https://perma.cc/G4WM-RTZ6>]; *UN Guiding Principles*, BUS. & HUM. RTS. RES. CTR. (2011), <https://www.business-humanrights.org/en/big-issues/un-guiding-principles-on-business-human-rights/> [<https://perma.cc/ST6U-AVRG>]; *Helping Companies Ensure Respect for Human Rights*, BSR (Oct. 22, 2012), <https://www.bsr.org/en/case-studies/helping-companies-ensure-respect-for-human-rights> [<https://perma.cc/29E6-5Q3E>]; Letter from Microsoft Workers to Satya Nadella, CEO, Microsoft (Feb. 22, 2019) [<https://perma.cc/63A9-6SA4>].

275. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Rep. of the Special Rapporteur, ¶¶ 9-14, 53, U.N. Doc. A/HRC/32/38 (May 11, 2016); see also *UN Expert: Content Moderation Should not Trample Free Speech*, OHCHR (July 19, 2018), <https://www.ohchr.org/en/stories/2018/07/un-expert-content-moderation-should-not-trample-free-speech> [<https://perma.cc/76M9-74N9>]; David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Rep. of the Special Rapporteur, ¶ 51, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018).

276. CORPORATE HUMAN RIGHTS POLICY, META (2021), <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf> [<https://perma.cc/4989-MVJ7>].

277. See Susan Benesch, *But Facebook’s Not a Country: How to Interpret Human Rights Law for Social Media Companies*, 38 YALE J. ON REGUL. BULLETIN 86 (2020) (exploring which provisions of international law would need to be clarified for social media companies to follow it); Brenda Dvoskin, *International Human Rights Law Is Not Enough to Fix Content Moderation’s Legitimacy Crisis*, MEDIUM (Sept. 16, 2020), <https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd> [<https://perma.cc/4989-MVJ7>].

contribution of human rights in practice. Taking a capacious view of its enabling charter's selective reference to freedom of expression,²⁷⁸ it has consistently anchored its analysis in the corporate responsibility to respect the full suite of human rights. The Board's opinions pertaining to the IRU special escalation process and cross-check exemplify how human rights make legible and contestable the analog privilege of law enforcement and VIP users in content moderation.

i. IRUs

The Board leaned on human rights to call into question Meta's special treatment of police requests. Based on the principle of legality for restricting freedom of expression, it raised concerns about the discrepancy between public-facing broadly-worded policies and additional internal guidance.²⁷⁹ Human rights thus allowed the Board to criticize the platform's double standards whereby specialized internal teams apply a more detailed analysis to content flagged by law enforcement. The Board also leveraged human rights to highlight systemic discrimination in the cumulative effects of IRU flags: all 286 flags by the UK Metropolitan Police pertained to drill music, a genre particularly popular among young Black people.²⁸⁰ The right to freedom of expression without racial discrimination thus makes visible the nefarious ways in which IRUs' deployment of analog privilege distorts enforcement.²⁸¹ In that sense, human rights allowed the Board to make legible certain problematic aspects of analog privilege.

cc/7LDP-MWZH] (describing latent ambiguities and “monologue” of companies interpreting human rights); Evelyn Douek, *The Limits of International Law in Content Moderation*, 6 U.C. IRVINE J. INT'L, TRANSNAT'L, & COMPAR. L. 37, 63 (2021) (arguing that pliable and contested human rights give platforms an unearned legitimacy dividend); see also Laurence R. Helfer & Molly K. Land, *The Meta Oversight Board's Human Rights Future*, 44 CARDOZO L. REV. 2233 (2023) (depicting the Board as a de facto human rights tribunal that can pressure Meta to improve its process and develop the law); Evelyn Douek, *The Meta Board and the Empty Promise of Legitimacy*, 37 HARV. J.L. & TECH. (forthcoming 2024) (manuscript at 42–44), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4565180 [<https://perma.cc/GSV2-KFNM>] (arguing that the Board has failed to explain the applicability of public human right norms to private content moderation).

278. OVERSIGHT BD. CHARTER art. 2 § 2, at 5 (2019) https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf [<https://perma.cc/B7YN-2YVX>] (the Board will “pay particular attention to the impact of removing content in light of human rights norms protecting free expression.”).

279. *UK Drill Music*, *supra* note 223 (“The differences between the publicly facing Violence and Incitement Community Standard and Meta's internal Implementation Standards is also a concern.”).

280. *Id.*

281. *Id.*; International Convention on the Elimination of All Forms of Racial Discrimination, 660 U.N.T.S. 195 (1966) art. 5(d)(viii).

ii. Cross-check

Human rights were also instrumental to the Board's criticism of analog privilege for cross-checked VIP users. The Board warned that guardrails against over-moderation have limited impact if "the system is designed primarily to protect or prioritize the expression of people who are already powerful."²⁸² Instead of finessing exceptions for the select few, the platform should improve the general moderation process:

If Meta's moderation more accurately moderated the content of all users, it would not need special programs based on entitled entities to help advance its respect of human rights.²⁸³

In other words, the platform should distribute scarce human resources according to human rights risks, not the speaker's clout. Human rights assisted the Board in suggesting Meta move away from the tiered moderation model resulting from analog privilege.

Although non-binding, human rights can address certain problematic facets of analog privilege in content moderation. They highlight some gaps between automated moderation at scale and the special privileges afforded to law enforcement and VIP users.

c. The EU Digital Services Act

The EU has taken the lead on platform regulation. The Digital Services Act (DSA) entered into force on November 16, 2022, ushering in unprecedented regulation of online services.²⁸⁴ The DSA imposes basic obligations on all intermediary services, and tailors gradually more demanding ones for hosting services and online platforms, as illustrated in Figure 2.²⁸⁵ Within that latter category, the biggest players face even more stringent obligations.

282. OVERSIGHT BD., *supra* note 197, at 24.

283. *Id.*

284. Digital Services Act, *supra* note 214.

285. IAN BROWN, CHRISTOPHER T. MARSDEN & ERNEST J. WILSON, REGULATING CODE: GOOD GOVERNANCE AND BETTER REGULATION IN THE INFORMATION AGE 8 (2013) (describing the different layers of internet infrastructure); Joan Donovan, *Navigating the Tech Stack: When, Where and How Should We Moderate Content?*, CTR. FOR INT'L GOVERNANCE INNOVATION (Oct. 28, 2019), <https://www.cigionline.org/articles/navigating-tech-stack-when-where-and-how-should-we-moderate-content/> [https://perma.cc/EXF8-Q8PU] (discussing where content moderation should occur in the technology stack).

DSA OBLIGATIONS

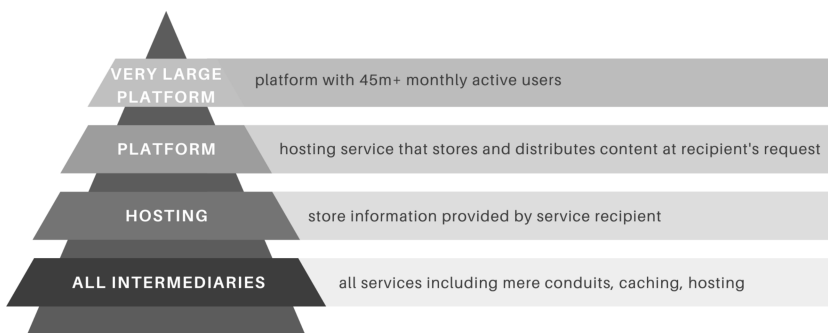


Figure 2: DSA obligations

Transparency requirements in the DSA could alter analog privilege. All intermediary services must provide “meaningful and comprehensible information” about how they engage in content moderation.²⁸⁶ They must disclose how much and which content is moderated through automation versus human review. Depending on how companies and oversight entities ultimately interpret this obligation, it could thrust analog privilege to center stage by forcing platforms to disclose when, and to whom, they grant exceptions from automated moderation. Furthermore, affected users get a statement of reasons for enforcement action adversely affecting their content. Where “strictly necessary,” these reasons must identify who flagged the content.²⁸⁷ The strictly necessary concept echoes the principle of data minimization in the General Data Protection Regulation (GDPR), presumably deployed here to protect the flagger’s privacy.²⁸⁸ However, government entities such as IRUs do not enjoy the same level of privacy protection as “identifiable natural persons” covered by the GDPR.²⁸⁹ Therefore, the GDPR should not stand in the way of platforms notifying users that their content was flagged by an IRU. One could further make the case that users should know when the police initiated an enforcement process because the rationale for

286. Digital Services Act, *supra* note 214, art. 15(1)(c).

287. *Id.* art. 17(3)(b).

288. Council Regulation 2016/679, General Data Protection Regulation art. 5(c), 2016 O.J. (L 119) 1 (EU) (limiting personal data processing “to what is necessary in relation to the purposes for which they are processed (‘data minimisation’)”).

289. *Id.* art. 4(1) (person “who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an on-line identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”).

regulating the flagging process is to protect users' free expression from state interference.²⁹⁰

The DSA further guarantees trusted flaggers priority treatment in return for transparency about their reporting activities.²⁹¹ Formalizing their status could foster accountability about how platforms dispense analog privilege to law enforcement. However, platforms can continue to collaborate with unregistered IRUs, such that entities already benefitting from analog privilege have little incentive to apply for official trusted flagger status.²⁹²

Requirements for very large platforms (VLOPs) indirectly address analog privilege as well. Platforms with over 45 million monthly active users must assess and redress their content moderation's systemic risks.²⁹³ They must proactively map their activities' actual and foreseeable effects on rights enshrined in the EU Charter of Fundamental Rights, including freedom of expression and non-discrimination.²⁹⁴ In light of the UK's apparent fixation on drill music, this could involve thinking through the implications of analog privilege for law enforcement that disproportionately silences the speech of minority groups.²⁹⁵ For Meta's cross-check program, this could mean taking a fresh look at the false-positive prevention program claiming to protect vulnerable users but shielding powerful ones in practice.²⁹⁶ The 24-hour remediation window harming revenge porn victims is a case in point.²⁹⁷

Research access provisions further incentivize VLOPs to address analog privilege. Vetted researchers get unprecedented opportunities to probe platforms' internal data for systemic risks.²⁹⁸ Systemic risks include fundamental rights listed in the EU Charter, civic discourse, gender-based violence, public security, and public health.²⁹⁹ Researchers could invoke said risks to investigate analog privilege, for example understanding how cross-check's 24-hour remediation window enables non-consensual

290. Charter of Fundamental Rights of the European Union, art. 11, 2012 O.J. C 326/391; Digital Services Act, *supra* note 214, recitals 3, 22, 51–54.

291. Digital Services Act, *supra* note 214 art. 22.

292. Appelman & Leerssen, *supra* note 49, at 7.

293. Digital Services Act, *supra* note 214, arts. 34(2)(b), 35(1)(c).

294. Charter of Fundamental Rights of the European Union, *supra* note 290 arts. 11, 21; Digital Services Act, *supra* note 214, art. 34(1)(b).

295. The DSA doesn't apply to the UK, but see Online Safety Act, *supra* note 214 risk assessment requirements at §§ 7, 9, 26.

296. *See supra* Part III.B.2.b.

297. *Supra* note 251.

298. Digital Services Act, *supra* note 214, art. 40(4).

299. *Id.* art. 34(1).

sharing of intimate images.³⁰⁰ Comparing views of such content moderated through automation at scale versus the special cross-check program would produce valuable insight about the implications of analog privilege for gender-based violence. Data access provisions thus provide an avenue for shedding light on the systemic downsides of analog privilege.

Finally, VLOPs must disclose how human moderation fares across languages. Platforms tend to focus moderation resources on English content, leaving non-English users ill-protected against the shortcomings of automated moderation.³⁰¹ For each EU language, VLOPs must detail the number of human moderators and their linguistic qualification, as well as the accuracy of automated moderation.³⁰² Unpacking how platforms dispense analog review could inform broader debates about English-centric bias in moderation and in the conferral of analog privilege. Taken together, these obligations could foster some accountability about analog privilege favoring powerful constituencies like law enforcement.

Content moderation exemplifies the versatility of analog privilege. Law enforcement secures additional manual review to avoid underenforcement in areas it deems a priority. Analog privilege in social media can have far-reaching consequences like deplatforming, as the young content creators in the UK drill case experienced. Conversely, high-profile cross-checked users get sophisticated human review to avoid overenforcement, further entrenching their visibility online. Though analog privilege manifests differently across domains, two striking commonalities stand out: elites wield their power to bypass automation and existing rules fail to fully account for this unearned privilege. This disconnect between legal regimes and analog privilege speaks to broader AI governance gaps. Part IV turns to solutions narrowing these gaps.

IV. SOLUTIONS

Instead of power-driven privilege, reasoned debate should determine who gets to circumvent detrimental AI systems.³⁰³ The proposals below outline several options available to policymakers striving to address analog privilege.

300. *Supra* note 251.

301. Tom Simonite, *Facebook Is Everywhere; Its Moderation Is Nowhere Close*, WIRED (Oct. 25, 2021), <https://www.wired.com/story/facebooks-global-reach-exceeds-linguistic-grasp/> [<https://perma.Cc/P9R3-U892>].

302. Digital Services Act, *supra* note 214, art. 42(2).

303. Cen & Raghavan, *supra* note 46 (suggesting that exemptions from automated systems should consider individualization, uncertainty, and harm).

A. *Reduce Analog Privilege*

The most direct route to eradicate analog privilege is to make analog treatment available to everyone. Personal data regulation in the EU suggests this is feasible, with an opt-out regime for automated high-stakes decisions.³⁰⁴ California's consumer privacy regime goes a step further, prohibiting retaliatory measures against users opting out of data tracking.³⁰⁵ In the area of content moderation, extending additional human review to all Meta users is directionally sound—although unlikely to scale up to offer meaningful human review across the platform.³⁰⁶ These efforts to democratize analog processing can come from either public or private actors, and can target AI systems with various levels of specificity ranging from blanket data protection regimes to company-specific content moderation practices. This flexibility gives policymakers significant latitude to implement analog guarantees, ranging from mandatory personal data regulation to voluntary codes of practice targeting specific areas of activity.

Banning harmful AI applications would also reduce analog privilege. Taking problematic AI deployments off the table deprives analog elites of their privilege to sidestep them. In the United States, the Federal Trade Commission is considering curbing commercial surveillance³⁰⁷ and scholars have proposed prohibiting algorithmic wage discrimination.³⁰⁸ The EU's AI Act, currently in the final stages of adoption, will also bar some predictive policing, facial recognition, and biometric identification applications.³⁰⁹ Focusing on egregious cases of AI harm, these bans could attenuate the starkest instances of analog privilege.

A more controversial approach would close analog treatment loopholes by enforcing automation on groups hitherto spared from

304. General Data Protection Regulation, *supra* note 288, art. 2(1), 22(1).

305. California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.125 (West 2023).

306. *Reviewing High-Impact Content Accurately via our Cross-Check System*, *supra* note 235. *But see* OVERSIGHT BD., *supra* note 197, at 10, 20 (noting that Meta still prioritizes human review of VIP users).

307. FTC Commercial Surveillance and Data Security Regulation, 87 Fed. Reg. 51273 (proposed Aug. 22, 2022).

308. Dubal, *supra* note 40, at 1989–92.

309. *European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* COM (2021) 0206 – C9-0146/2021 – 2021/0106(COD)), European Parliament P9_TA(2024)0138 (Mar. 13, 2024), https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html [<https://perma.cc/JBA7-WA82>] art. 5; for updates on the legislative process, see *Legislative Train Schedule*, EUROPEAN PARLIAMENT, <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence> [<https://perma.cc/3ZKD-LHYQ>].

detrimental AI. For instance, extending automated state surveillance dismantles the invisible privilege of analog elites. In the context of child mistreatment predictive analytics, including private mental health records and luxury rehab center treatment in automated risk predictions would monitor rich and poor parents equally.³¹⁰ Better visibility across the board would also prevent elites from skewing the training data of punitive AI systems to ignore their risky behavior. For example, risk predictors for white (figuratively and perhaps literally) collar crime remain coarse,³¹¹ and the tunnel vision already focusing police patrols on minority communities gets further entrenched as self-fulfilling prophecies of risk predictions lead to more policing and enforcement. That said, extending surveillance dragnets without reforming the mission of these monitoring systems would fall short of a comprehensive solution to underlying AI harms. For inclusion to become an asset, AI systems need to serve instead of punish people. To continue with the child mistreatment example, predictive analytics can support rather than suspect at-risk families. Similarly, at-risk youth could receive mentorship instead of police surveillance when automated systems detect risk. Yet in practice, decoupling public services from pervasive carceral mentalities still has a long way to go.³¹² As such, policymakers should take great caution to avoid regressive blanket automation.

B. Accountability

Visibility into analog privilege is necessary to assess compliance with existing legal regimes and to inform policy. The EU Digital Services Act (DSA) provides a blueprint in the context of social media, with mandatory disclosures fostering external accountability.

As outlined in the content moderation case study, the DSA requires companies to disclose details about their content moderation methods, resources, and allocation thereof.³¹³ Transposed to analog privilege, entities deploying AI should report the proportion, make-up, and process of exempting certain people from otherwise applicable automation.

310. Abdurahman, *supra* note 1.

311. BRIAN CLIFTON, SAM LAVIGNE & FRANCIS TSENG, *PREDICTING FINANCIAL CRIME: AUGMENTING THE PREDICTIVE POLICING ARSENAL* (2017), <https://whitecollar.thenewinquiry.com/static/whitepaper.pdf> [<https://perma.cc/XGE5-FNQ2>].

312. *THE PRISON IN TWELVE LANDSCAPES* (Oh Ratface Films 2016); JACKIE WANG, *CARCERAL CAPITALISM* (2018); *see also* Matt Stroud, *An Automated Policing Program Got This Man Shot Twice*, *VERGE* (May 24, 2021), <https://www.theverge.com/c/22444020/chicago-pd-predictive-policing-heat-list> [<https://perma.cc/BN7M-NZJT>] (reporting that a gang violence prevention program inadvertently marks targets as “snitches” due to frequent unsolicited police contact).

313. *See supra* Part III.B.3.c.

Furthermore, disaggregating the performance of human and AI systems would inform assessments about the equality implications of analog privilege. For example, higher error rates for AI systems relative to human decision-making should raise red flags, especially if protected groups bear more than their fair share of these errors. Entities deploying AI systems should proactively map the equities of doing so with a view to applicable equality protections, and plan to remedy identified harms.

To complement these voluntary disclosures, external researchers should also be allowed to probe automation practices. Again, the DSA provides a roadmap, as it compels large players to facilitate third-party access to their internal data for research purposes.³¹⁴ Transposed to analog privilege, this independent oversight would foster a more thorough understanding of potential systemic risks, including to civil rights.

C. Technical Fixes to Improve AI Systems

Since AI harms and analog privilege are inextricably linked, improving AI systems correspondingly contracts analog privilege. Some AI models already optimize for harmlessness based on human rights, existing principles, and non-western views.³¹⁵ Others give affected people the opportunity to challenge official narratives with their own data to counter skewed narratives resulting from implicit choices in data selection. Take the example of recidivism risk assessment during bail proceedings: a pilot project in Santa Clara County, California enables communities to provide additional information about defendants to contradict the institutional data.³¹⁶ Participatory automated systems could even integrate community data upfront to avoid mischaracterizing defendants in the first place.³¹⁷

Engineers could also aim to improve the fairness of AI systems. Constraining outputs along gender, racial or other protected characteristics could ensure such factors don't drive the outcomes of AI-driven

314. Digital Services Act, *supra* note 214, art. 40.

315. Yuntao Bai et al., *Constitutional AI: Harmlessness from AI Feedback*, ARXIV (Dec. 15, 2022), <https://arxiv.org/pdf/2212.08073.pdf> [<https://perma.cc/75VJ-85LH>]; *Claude's Constitution*, ANTHROPIC (May 9, 2023), <https://www.anthropic.com/index/claude-constitution> [<https://perma.cc/TP3C-4X3H>].

316. Raj Jayadev, *The Future of Pretrial Justice Is Not Money Bail or System Supervision – It's Freedom and Community*, SILICON VALLEY DE-BUG (Apr. 4, 2019) <https://www.siliconvalleydebug.org/stories/the-future-of-pretrial-justice-is-not-money-bail-or-system-supervision-it-s-freedom-and-community> [<https://perma.cc/5VTW-Q5AC>].

317. Ngozi Okidegbe, *Discredited Data*, 107 CORNELL L. REV. 2007, 2007 (2022) (proposing that community knowledge replace carceral narratives about risk in bail proceedings).

decision-making.³¹⁸ Although there are many competing and sometimes mutually exclusive definitions of fairness, explicitly optimizing models to respect one mathematical definition of fairness is a step forward compared to “naïve” models which risk perpetuating existing discrimination. Put differently, failure to constrain models for fairness is itself an active choice to reify the status quo of disparate outcomes. At minimum, models should convey how they deal with past or static traits, allowing people to know when factors beyond their control affect outcomes. Where sector-specific antidiscrimination statutes apply, such as in employment and housing, these metrics should be available by default.

Introducing randomness could offset the role of past or static factors in determining outcomes. As explained above, assuming the future mirrors the past deprives people of meaningful opportunities to change and grow. Where randomness entails accuracy loss, deliberate decisions about the trade-offs should occur at the appropriate level of responsibility.

Adding dimensionality to the output of AI systems can also counter the reductive downside of automation. Instead of a single clear-cut answer, nuanced outputs would paint a fuller picture. For instance, productivity software might score workers along different dimensions like contribution to team morale and constructive input in meetings. Issuing a confidence rating for each output would also prompt the human making the final call (if any) to take results with a grain of salt if need be.³¹⁹

Finally, privacy-preserving measures could ward off the voyeuristic drawbacks of automation. Group-level trends can inform policy without punishing individuals or disadvantaging groups. By introducing some noise in the data, differential privacy can preserve group-level insights without compromising individuals.³²⁰ For example, child mistreatment predictive analytics could identify population-level risk factors rather than flag specific families for investigation.

318. See *supra* note 38 for different fairness metrics.

319. Jumper et al., *supra* note 13, at 584 (giving an example of reliability estimate to inform confidence in prediction); Raja Parasuraman & Dietrich H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUM. FACTORS 381, 381 (2010) (raising the tendency to unduly trust automated systems).

320. Alexandra Wood et al., *Differential Privacy: A Primer for a Non-Technical Audience*, 21 VAND. J. ENT. & TECH. L. 209, 232 (2018); Cynthia Dwork et al., *Calibrating Noise to Sensitivity in Private Data Analysis*, in *THEORY OF CRYPTOGRAPHY* 265 (Shai Halevi & Tal Rabin eds., 2006); see also Ryan Yackel, *What Is Homomorphic Encryption, and Why Isn't It Mainstream?*, KEYFACTOR (July 6, 2021), <https://www.keyfactor.com/blog/what-is-homomorphic-encryption> [<https://perma.cc/HF8F-CGG5>].

D. External Checks and Balances

As it stands, a handful of companies lead AI development and monopolize governance conversations.³²¹ Moreover, a rise in partnerships between startups and large technology corporations to secure vital cloud computing to run AI models portends further consolidation among what already amounts to a supply chain oligopoly.³²² While corporate self-interest can sporadically align with altruistic protection of other stakeholders, this setup is not conducive to comprehensively addressing the drawbacks of AI, let alone analog privilege. Meaningful counterweight to de facto industry-driven dominance requires both more robust direct government intervention and a seat at the table for a broader array of stakeholders. Many governments, including the United States, are considering ramping up public oversight, with some form of regulatory agency to oversee the deployment of AI systems.³²³ But given their limited capacity, resources, and expertise, there is only so much that governments can do directly.

In light of these operational constraints, public bodies have resorted to mobilizing other private actors to limit the quasi-unfettered discretion of the AI industry. For instance, the EU AI Act enlists auditors to assess high-risk systems. Incentives for arms-length, independent

321. PBS NewsHour, *OpenAI CEO Sam Altman Testifies on Artificial Intelligence Before Senate Committee*, YouTube (May 16, 2023, 1:39), https://www.youtube.com/watch?v=P_ACcQxJIsg&ab_channel=PBSNewsHour [<https://perma.cc/4692-NJJH>] (Sen. John Kennedy, R-La offers tech CEO to manage an AI regulatory agency); Cecilia Kang, *OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing*, N.Y. TIMES (May 16, 2023), <https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html> [<https://perma.cc/T5PF-CGG5>] (reporting on a private meeting between Sam Altman and Senators prior to his public testimony).

322. Opinion Lex, *OpenAI/Microsoft: Buying Startup Would Be Intelligent Move For Giant*, FIN. TIMES (Nov. 1, 2023), <https://www.ft.com/content/ac8710fb-648d-4c3f-bf4f-81f9c9da16fa> [<https://perma.cc/BPR9-G5SF>]; Adam Satariano & Cade Metz, *Amazon Takes a Big Stake in the A.I. Start-Up Anthropic*, N.Y. TIMES (Sept. 25, 2023), <https://www.nytimes.com/2023/09/25/technology/amazon-anthropic-ai-deal.html> [<https://perma.cc/S7U4-XY2B>]; BARRY LYNN, MAX VON THUN & KARINA MONTROYA, *AI IN THE PUBLIC INTEREST: CONFRONTING THE MONOPOLY THREAT* 36 (2023), <https://www.openmarketsinstitute.org/publications/report-ai-in-the-public-interest-confronting-the-monopoly-threat> [<https://perma.cc/6N9L-KDHL>].

323. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules an Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, *supra* note 309, arts. 74, 77 (setting up national regulator as first line of oversight including post market surveillance of AI systems in the EU market); Press Release, Sens. Blumenthal & Hawley, *Bipartisan Framework on Artificial Intelligence Legislation* (Sept. 8, 2023) <https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-hawley-announce-bipartisan-framework-on-artificial-intelligence-legislation> [<https://perma.cc/6CZE-7QWX>] (proposing a licensing regime administered by an independent oversight body); *Artificial Intelligence and Data Act* § 33, C-27 (Can.) (2022) (setting up an AI and Data Commissioner to enforce AI regulation).

evaluations set up an ecosystem where private auditors can effectively check industry-led AI developments.³²⁴ Similarly, platform regulation enables external researchers to access the data of large platforms to probe systemic risks, thus facilitating independent assessment of some AI-driven harms.³²⁵ In the United States, efforts to democratize AI governance have thus far clustered around non-binding consultations of civil society and affected stakeholders.³²⁶ While a step in the right direction, consultation alone will not counteract industry-dominated AI governance. Two avenues stand out to diversify the voices shaping AI development: empowering affected people to challenge harms, and protecting technology workers who denounce problematic practices.

A private right of action against automation harms gives the analog have-nots a seat at the table.³²⁷ Effective grievance mechanisms would promote post-facto correction of AI-driven harms. Empowering people to seek redress against automation indirectly challenges analog privilege, insofar as it makes legible the automation harms elites avoid. States should support litigants with expertise and financial assistance to ensure that people harmed by AI systems have meaningful recourse in practice.³²⁸ Moreover, these claimants are essentially performing public

324. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, *supra* note 309, arts. 31, 33, 37 (setting up auditors, called “notified bodies”, that depend on state certification to operate, and must maintain independence from their clients. The European Commission can also step in to investigate them).

325. *See supra* Part III.B.3.c on the Digital Services Act; *see also supra* note 305.

326. WHITE HOUSE, BLUEPRINT FOR AN AI BILL OF RIGHTS (2022), <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> [<https://perma.cc/M97V-X2GD>] (engaging in yearlong consultation on non-binding document); AI Accountability Policy, 88 Fed. Reg. 22433 (proposed Apr. 13, 2023); *see also* THE PARTNERSHIP ON AI, <https://partnershiponai.org> [<https://perma.cc/2ZT4-FYUD>] (setting up multistakeholder initiative). *But see* Press Release, Access Now, Access Now Resigns from The Partnership on AI (Oct. 13, 2020), <https://www.accessnow.org/press-release/access-now-resignation-partnership-on-ai> [<https://perma.cc/4V8X-4JRP>] (civil society organization leaving due to lack of impact); Off. of Mgmt. & Budget, Exec. Off. of the President, Proposed Memorandum for the Heads of Executive Departments and Agencies (Nov. 2023), <https://ai.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-Public-Comment.pdf> [<https://perma.cc/7KNG-TU4M>] (requiring federal agencies to consult affected stakeholders); Exec. Order No. 14,110, 88 Fed. Reg. 75191, 75191 (Nov. 1, 2023) (seeking to ensure engagement with affected communities).

327. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, *supra* note 309, art. 85 (setting up a private right of action); *see also* *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, *supra* note 181.

328. For an example in Canada, *see* CT. CHALLENGES PROGRAM, <https://pcj-ccp.ca> [<https://perma.cc/YT7V-GMHJ>].

interest litigation by compelling courts to clarify newly enacted laws, further justifying the case for supporting them.

Second, policymakers can leverage the technology workers who often act as the industry's conscience. The people building automated systems have unique visibility into AI harms and have indeed been instrumental in voicing out concerns, ethical dilemmas, and problematic practices.³²⁹ Yet uncertain legal exposure and restricted work prospects can silence them. Limiting nondisclosure agreements,³³⁰ non-disparagement clauses,³³¹ and non-compete clauses³³² can foster their ability to speak truth to power. Policymakers could also encourage workers to speak freely by clarifying the breadth of federal trade secret protection, both in terms of liability and job mobility restrictions for employees who must disclose trade secrets to expose problematic practices.³³³ Finally, improving SEC whistleblower compensation could further incentivize workers to speak out.³³⁴

None of these proposals alone will fix analog privilege, and some interventions like enforcing flawed AI systems on elites might entail collateral damage. But judiciously combined, many of these initiatives can attenuate analog privilege. Like the Swiss cheese pandemic model layering masking, vaccines, and social distancing,³³⁵ superimposing various interventions can curtail analog privilege.

329. Michael Grothaus, *Amazon Workers Demand the Company Stop Enabling Trump's ICE*, FAST CO. (June 22, 2018), <https://www.fastcompany.com/40588572/amazon-workers-demand-the-company-stop-enabling-trumps-ice> [https://perma.cc/PPB4-GFDA]; Sheera Frenkel, *Microsoft Employees Protest Work With ICE, as Tech Industry Mobilizes Over Immigration*, N.Y. TIMES (June 20, 2018), <https://www.nytimes.com/2018/06/19/technology/tech-companies-immigration-border.html> [https://perma.cc/DRU5-5DTV]; Bryan Menegus, *Thousands of Google Employees Protest Company's Involvement in Pentagon AI Drone Program*, GIZMODO (Apr. 4, 2018), <https://gizmodo.com/thousands-of-google-employees-protest-companys-involvem-1824988565> [https://perma.cc/N89N-2PRU].

330. Orly Lobel, *NDAs Are Out of Control. Here's What Needs to Change*, HARV. BUS. REV. (Jan. 30, 2018), <https://hbr.org/2018/01/ndas-are-out-of-control-heres-what-needs-to-change> [https://perma.cc/25RD-JLA4].

331. N.L.R.B., Office of the General Counsel, Advice Memorandum on Stange Law Firm (Mar. 4, 2019) (stating that contractual prohibitions to criticize, ridicule or disparage a law firm violate employees' right to engage in concerted activity for mutual aid or protection).

332. Exec. Order No. 14036, 86 Fed. Reg. 36987 (July 14, 2021) (encouraging the Federal Trade Commission to ban or limit non-compete agreements to boost competition); see also CAL. BUS. & PROF. CODE § 16600 (West 2024) (limiting on non-compete clauses).

333. Defend Trade Secrets Act of 2016, 18 U.S.C. § 1836; U.S. PATENT AND TRADEMARK OFFICE, OFFICE OF POLICY AND INTERNATIONAL AFFAIRS, THE DEFEND TRADE SECRETS ACT AT FIVE: THE INEVITABLE DISCLOSURE DOCTRINE (2021); Orly Lobel, *The DTSa and the New Secrecy Ecology*, 1 BUS., ENTREPRENEURSHIP & TAX L. REV. 369 (2017).

334. SEC Whistleblower Program Rules, 17 C.F.R. § 240.21F-5(b) (2023) (offering 10-30% of the recovered fines over a million dollars).

335. Siobhan Roberts, *The Swiss Cheese Model of Pandemic Defense*, N.Y. TIMES (Dec. 5, 2020), <https://www.nytimes.com/2020/12/05/health/coronavirus-swiss-cheese-infection-mackay.html> [https://perma.cc/AG37-MP6A].

To coordinate this multi-prong approach, lawmakers should deliberately mobilize the strengths of different power centers in the AI ecosystem to craft a smart mix of interventions. They should lean on the industry's technical expertise to implement fairness constraints and rely on academic research for external oversight. They should also combine direct substantive regulation like prohibiting the most harmful AI applications with indirect interventions like empowering technology workers to be the industry's conscience. Future research will explore a polycentric model to guide public interventions in AI governance.³³⁶

CONCLUSION: EVERYTHING OLD IS NEW AGAIN

AI swallows people whole and regurgitates mass guesstimates.³³⁷ Against this backdrop, the ability to opt out of AI systems becomes a privilege. In this data-driven economy, power confers control over how much, or little, one shares with AI systems. While AI entails undeniable benefits, it also involves inherent harms which disproportionately impact marginalized groups. And despite accelerating breakthroughs stirring the common imaginary, machine learning systems remain limited to formalistic proficiency, as opposed to functional capabilities.

The ability to avoid or override automation with preferential analog treatment is a luxury. Considering automation and analog processing at once reveals a tiered system: default automation on one side, and analog privilege on the other. This automation divide strains already tenuous social cohesion with a new fault line between reductive, deterministic, voyeuristic AI for most, and white glove human treatment for the select few. The split could fuel broader recriminations against elites perceived to enjoy special treatment. Seen in this light, the automation divide fits and perhaps feeds resentment-fueled populism, further straining the already-thin social fabric. But all is not lost. As this Article demonstrates, analog privilege can be managed through coordinated, deliberate interventions across the AI ecosystem.

336. Elinor Ostrom, *Beyond Markets and States: Polycentric Governance of Complex Economic Systems*, 100 AM. ECON. REV. 641, 644 (2010); Vincent Ostrom et al., *The Organization of Government in Metropolitan Areas: A Theoretical Inquiry*, 55 AM. POL. SCI. REV. 831, 831–32 (1961); Paul D. Aligica & Vlad Tarko, *Polycentricity: From Polanyi to Ostrom, and Beyond*, 25 GOVERNANCE 237, 238–39 (2012).

337. Kelly Pendergrast, *The World Is Toxic. Welcome to the Metabolic Era*, WIRED (July 14, 2023), <https://www.wired.com/story/the-world-is-toxic-welcome-to-the-metabolic-era/> [<https://perma.cc/R6YS-B9C5>] (LLM's "suck down the sprawling detritus of human effort and swallow it into the great black box stomach of the AI system, which converts it into something uncanny and instant and profitable").

ANNEX I – CHATGPT EXAMPLE³³⁸

The following text was generated using OpenAI's ChatGPT application

Prompt: write a plaintiff brief for a slip and fall case involving slipping on black ice in poorly lit stairs with a good analogy

Introduction

This is a personal injury action brought by the Plaintiff, [Name], against the Defendant, [Name], for injuries sustained as a result of a slip and fall on black ice on poorly lit stairs on Defendant's property. The Plaintiff respectfully submits this brief in support of their claims. The case is like a ticking time bomb, where defendant was aware of the hazardous condition but failed to take any action to prevent it from causing harm to plaintiff.

Factual Background

On [date], the Plaintiff was walking down the stairs located on Defendant's property when they slipped and fell on black ice that had formed on the stairs. The stairs were poorly lit, making it difficult for the Plaintiff to see the black ice and avoid the fall. As a result of the fall, the Plaintiff sustained [specific injuries].

Defendant's Negligence

The Defendant had a duty to maintain their property in a safe condition for invitees, including the Plaintiff. This duty includes the responsibility to remove or warn of any hazards, such as black ice, that may exist on the property. The Defendant failed to fulfill this duty by allowing the black ice to form on the stairs and failing to properly light the area, which made it difficult for the Plaintiff to see and avoid the hazard.

The Defendant's negligence was the direct and proximate cause of the Plaintiff's injuries. The Defendant's failure to remove or warn of the black ice and lack of proper lighting created a dangerous condition that the Plaintiff could not have reasonably anticipated or avoided. It was like a ticking time bomb, where defendant was aware of the hazardous condition but failed to take any action to prevent it from causing harm to plaintiff.

Damages

As a result of the Defendant's negligence, the Plaintiff has suffered [specific damages], including but not limited to [medical expenses, lost wages, pain and suffering, etc.]. The Plaintiff respectfully requests that the Court award them damages in an amount that is fair and just under the circumstances.

338. CHATGPT, <https://chat.openai.com/>, January 27, 2023.

Conclusion

The Defendant's negligence was the direct and proximate cause of the Plaintiff's injuries. The Plaintiff respectfully requests that the Court find in their favor and award them damages for the injuries and losses they sustained as a result of the Defendant's negligence. The analogy of a ticking time bomb highlights the defendant's awareness of the hazardous condition and failure to take action, which ultimately led to plaintiff's injury.