ESSAY

ARTIFICIAL INTELLIGENCE IS LIKE A PERPETUAL STEW

NATHAN REITINGER*

Artificial intelligence is inescapable. It is in our phones, fridges, and most of the businesses we engage with use it to "improve" their services. From deciding on what YouTube video to watch next to driving vehicles or firing weapons, artificial intelligence is a linchpin in our society. But what is artificial intelligence? And, more importantly, why does that matter? It matters because we are currently unprepared to deal with the paradigm-shifting legal issues brought about by artificial intelligence. And without this understanding, we are nearly certainly going to make mistakes. The bright side is that artificial intelligence is not complicated. Artificial intelligence—more accurately, machine learning—is built on simple and intuitive concepts revolving around: (1) a particular type of machine known as a neural network; and (2) getting that machine to learn through a process of, or similar to, gradient descent. The aim of this Essay is to provide a low-level, accurate, and easy-to-digest explanation of what artificial intelligence is, providing a novel metaphor to aid in that explanation. Like a perpetual stew, models built with artificial intelligence start with a recipe (neural network architecture), are tweaked to a particular palate (training), and are set to live forever as long as they continue to produce tasty (accurate) results. In turn, the Essay provides the legal community with an accurate vantage point from which to analyze the many artificial-intelligence-based issues that will not be on the horizon for much longer.

* Ph.D. Candidate, *University of Maryland*, Department of Computer Science; M.S., *Columbia University*, Department of Computer Science; J.D., *magna cum laude*, *Michigan State University College of Law*.

TABLE OF CONTENTS

INTRODUCTION

We are in an artificial intelligence (AI) summer[1]: ChatGPT is a household name,[2] a large portion of our work would not be possible

---

1. Winter is coming—ChatGPT (the Chat Generative Pre-Trained Transformer) may be king right now, but, after a period of quelled AI interest, another form of AI will rise in its stead, just as ChatGPT replaced alphaGo replaced deepBlue. *See* JOSH HARGUESS & CHRIS M. WARD, *Is the Next Winter Coming for AI? Elements of Making Secure and Robust AI*, *in* APPLIED IMAGERY PATTERN RECOGNITION WORKSHOP (2022) (discussing the periodicity of AI "hype"); Henry A. Kautz, *The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture*, 43 A.I. MAG. 105, 105 (2022) (explaining how "[w]e are now in AI's third summer, a period of rapid scientific advances, broad commercialization, and exuberance" but each prior summer has been "followed by a winter of collapse"); Luciano Floridi, *AI and Its New Winter: From Myths to Realities*, 33 PHIL. & TECH. 1, 1 (2020) ("AI has had several winters. Among the most significant, there was one in the late 1970s, and another at the turn of the 1980s and 1990s. Today, we are talking about another predictable winter.") (internal citations and footnotes omitted); Max Roser, *The Brief History of Artificial Intelligence: The World Has Changed Fast—What Might be Next?*, OUR WORLD IN DATA (Dec. 6, 2022), https://ourworldindata.org/brief-history-of-ai [https://perma.cc/9QUU-QPCD].

2. *See* Bernard Marr, *A Short History of ChatGPT: How We Got to Where We Are Today*, FORBES (May 19, 2023, 1:14 AM), https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=12992b7f674f [https://perma.cc/DV6X-4NMX] (noting how ChatGPT "quickly went viral" after an early demo was released on November 30, 2022); *see also* Rohit Shewale, *ChatGPT Statistics—User Demographics (February 2024)*, DEMANDSAGE (Jan. 12, 2024), https://www.demandsage.com/ChatGPT-statistics [https://perma.cc/87FA-7377] (discussing the popularity of ChatGPT); Jack Kelly, *Goldman Sachs Predicts 300 Million Jobs Will Be Lost or Degraded by Artificial Intelligence*, FORBES (Mar. 31, 2023, 10:48 AM), https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/?sh=68e9bdad782b [https://perma.cc/7VT5-MWMK] ("ChatGPT surpassed one million users in its first five days of launching, the fastest that any company has ever reached this benchmark.").

without AI,[3] and it seems like every few months a new discipline or industry is disrupted.[4] AI—or what may be more easily conceived of as "perpetual stews" given their recipe-following creation, endlessly adjustable optimizations, and the continual care needed to maintain

3. From Grammarly to GitHub, AI-based helper tools are everywhere. *See* Brian Kennedy, Alec Tyson & Emily Saks, *Public Awareness of Artificial Intelligence in Everyday Activities*, PEW RSCH. CTR. (Feb. 15, 2023), https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities [https://perma.cc/2R5B-2TJT] (finding that a quarter of Americans report nearly constant use of AI tools); Steven Vaughan-Nichols, *92% of Programmers Are Using AI Tools, Says GitHub Developer Survey*, ZDNET (June 14, 2023, 10:06 AM), https://www.zdnet.com/article/github-developer-survey-finds-92-of-programmers-using-ai-tools [https://perma.cc/S5ZG-YP97] (surveying 500 US-based developers and finding that over 90% of programmers rely on AI when coding).

4. *See* Pranshu Verma & Gerrit De Vynck, *ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners*, WASH. POST (June 2, 2023, 6:00 AM), https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs [https://perma.cc/3AP7-8TSV] ("'Whenever people brought up ChatGPT, I felt insecure and anxious that it would replace me,' [Olivia Lipkin, a twenty-five-year-old copywriter] said. 'Now I actually had proof that it was true, that those anxieties were warranted and now I was actually out of a job because of AI.'"). Hollywood is perhaps the most recent example of this. *See* Michael Russell Gunn, *POV: I'm a Hollywood Screenwriter. This Is Why Our Strike Matters to You*, BU TODAY (May 4, 2023), https://www.bu.edu/articles/2023/pov-writers-guild-strike [https://perma.cc/QRP8-NZBX] ("We writers are the canaries in the coal mine. What's happening here in Hollywood is going to happen across the nation, and in most other industries, it's going to happen without a union to fight it."); Kevin Collier, *Actors vs. AI: Strike Brings Focus to Emerging Use of Advanced Tech*, NBC NEWS (July 14, 2023, 2:20 PM), https://www.nbcnews.com/tech/tech-news/hollywood-actor-sag-aftra-ai-artificial-intelligence-strike-rcna94191 [https://perma.cc/M2GK-9EDV] ("'[A]rtificial intelligence poses an *existential threat* to creative professions, and all actors and performers deserve contract language that protects them from having their identity and talent exploited without consent and pay.'" (emphasis added)); Jan Hatzius, Joseph Briggs, Devesh Kodnani & Giovanni Pierdomenico, *The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)*, GOLDMAN SACHS (Mar. 26, 2023, 9:05 PM), https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html [https://perma.cc/2KAH-FTV7] ("Extrapolating our estimates globally suggests that generative AI could expose the equivalent of 300mn full-time jobs to automation."); *see also* Tyna Eloundou, Sam Manning, Pamela Mishkin & Daniel Rock, *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*, ARXIV 1, 1–4 (Aug. 22, 2023), https://arxiv.org/pdf/2303.10130.pdf [https://perma.cc/7JM6-ASEE] (describing the likely broad and significant economic impact that large language models will have on the labor market). *But see* Jacques Bughin, *Why AI Isn't the Death of Jobs*, 60 MIT SLOAN MGMT. REV. 1, 1 (2018) (discussing the job-adding abilities of AI).

the production of accurate outputs—is inescapable.[5] Whether you are a judge making a sentencing decision,[6] a radiologist detecting the presence of a disease,[7] a lawyer filing a brief,[8] or a YouTuber making a

5. Kennedy et al., *supra* note 3 (listing examples of common AI integration: fitness trackers posting habit insights, chatbots answering customer questions, product recommendations based on prior purchases, security camera sending an "unrecognized person" alert, music playlist recommendations, and email spam filtering). As with many new technologies, integration to the point of necessity is a historic trend, reaching all the way back to Jacquard's Loom. *See* JAMES ESSINGER, JACQUARD'S WEB: HOW A HAND LOOM LED TO THE BIRTH OF THE INFORMATION AGE 40–43 (2004) (discussing the rise in popularity of Jacquard's loom).

6. *See* John Villasenor & Virginia Foggo, *Artificial Intelligence, Due Process, and Criminal Sentencing*, 2020 MICH. ST. L. REV. 295, 311 (2020) (discussing the use of AI in judicial risk assessment).

7. *See generally* Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz & Hugo J.W.L. Aerts, *Artificial Intelligence in Radiology*, 18 NATURE REVS. CANCER 500, 500–10 (2018) (investigating AI's impact on the practice of radiology); Dakai Jin, Adam P. Harrison, Ling Zhang, Ke Yan, Yirui Wang & Jinzheng Cai et al., *Artificial Intelligence in Radiology, in* A.I. MED. 265, 265 (2021) (explaining that the advances of AI deep learning has spurred interest in AI within radiology, because "AI has the potential to recognize and localize complex patterns from different radiological imaging modalities, many of which even achieve comparable performance to human decision-making in recent applications").

8. Even besides the infamous lawyer who filed a brief with ChatGPT's assistance, unknowingly citing fake cases, Westlaw has been using AI publicly since 2018. *See* Press Release, *Thomson Reuters Unveils New Legal Research Platform with Advanced AI: Westlaw Edge*, PR NEWSWIRE (July 12, 2018, 8:28 AM), https://www.prnewswire.com/news-releases/thomson-reuters-unveils-new-legal-research-platform-with-advanced-ai-westlaw-edge-300680095.html [https://perma.cc/4LCC-8686]; *see also* Christopher A. Suarez, *Disruptive Legal Technology, COVID-19, and Resilience in the Profession*, 72 S.C. L. REV. 393, 394–95 (2020) (discussing AI's role for lawyers); Sara Merken, *New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief*, REUTERS (June 26, 2023, 4:28 AM), https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22 [https://perma.cc/2ZYB-CD7Q] (discussing how two New York lawyers received sanctions for submitting a legal brief with "six fictitious case citations generated by . . . ChatGPT").

decision on what video to watch next[9]—AI is playing a but-for role.[10] But what actually is AI, and more importantly, why does it matter?[11]

To answer the latter question, holding the former for now, consider this: by failing to understand what AI is, those who interpret, make, or enforce our laws will be woefully unprepared to handle the many AI-infused issues headed our way. These issues include *liability diffusion* (e.g., who is responsible when an AI itself makes a bad decision?[12]), *evidentiary mistrust* (e.g., if fake media is indistinguishable from real

9. *See* Paul Covington, Jay Adams & Emre Sargin, *Deep Neural Networks for YouTube Recommendations*, *in* ACM CONF. SERIES RECOMMENDER SYS. 191, 192 (2016) (discussing the neural network architecture giving rise to recommendations); *see also* Sundar Pichai, *Opening Keynote | Google I/O 2023*, YOUTUBE (May 11, 2023), https://www.youtube.com/watch?v=ixRanV-rdAQ [https://perma.cc/BS56-5EW6] (describing at 00:26, how Alphabet plans on embedding AI throughout its spectrum of products).

10. But for the AI, a judge would approach sentencing differently, a radiologist would look at a picture differently, a lawyer would draft a brief differently, and a user would pick which video to watch next differently. AI has transformed the way we live—just like saying "hello" was not a social norm until the telephone became widely popular. *See* AMMON SHEA, THE PHONE BOOK: THE CURIOUS HISTORY OF THE BOOK THAT EVERYONE USES BUT NO ONE READS 18–19 (2010) (internal citation omitted) ("'Hello,' as a greeting, seems as long-standing as the English language itself, and one can easily imagine that it was in use back when our linguistic ancestors were still living in mud huts and pestering the Romans. Incredibly, the advent of the telephone seems to mark the beginning of a great surge in this word's popularity in our language . . . . Thomas Edison is popularly credited with instigating the practice of saying 'hello' when answering the telephone [though h]is rival, Alexander Graham Bell, preferred 'ahoy' to be used.'").

11. From the legislative to the executive, knowing how to handle AI remains a question mark. *See* Cecilia Kang & Adam Satariano, *As A.I. Booms, Lawmakers Struggle to Understand the Technology*, N.Y. TIMES (Mar. 3, 2023), https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html [https://perma.cc/PXA4-YH7M] (discussing the light touch Congress has used to handle AI so far, leaving us with little to no answers); Will Douglas Heaven, P*redictive Policing Algorithms are Racist. They Need to be Dismantled*, MIT TECH. REV. (July 17, 2020), https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice [https://perma.cc/UC8E-7R4U] (discussing AI tools built to help policing, but which end up perpetuating systematic racism); Karen Hao, *AI is Sending People to Jail—and Getting it Wrong*, MIT TECH. REV. (Jan. 21, 2019), https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai [https://perma.cc/G6WG-56MK] (discussing the bias inherent in recidivism scoring).

12. *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 894 (2018) (discussing various liability problems caused by AI).

media, does this threaten a baseline of trust structuring society?[13]), and *creativity erosion* (e.g., if an AI can create beautiful, profitable art, do we need to reconsider the 'intelligence' of artistic creation?[14]). Although seemingly farfetched, some of these issues are more rooted in reality than we realize.

13. *See* Riana Pfefferkorn, *"Deepfakes" in the Courtroom*, 29 B.U. PUB. INT. L.J. 245, 245–46 (2020) (arguing that rules of evidence will withstand unauthentic synthetic data); *see also* Nathan Reitinger, Nathan Malkin, Omer Akgul, Michelle L. Mazurek & Ian Miers, *Is Cryptographic Deniability Sufficient? Non-Expert Perceptions of Deniability in Secure Messaging*, *in* PROCEEDINGS OF THE 44TH IEEE SYMPOSIUM ON SECURITY AND PRIVACY 274, 285–87 (2023) (finding that real-world deniability in secure text messaging is possible even without AI—how much more convincing would AI-assisted chat history spoofing be?). *But see* Agnieszka McPeak, *The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood*, 23 VAND. J. ENT. & TECH. L. 433, 450 (2021) (arguing for authenticity adjustments in the face of this new technology).

14. *See* Justin E. H. Smith, *The Philistine War on AI Art: Genius Can Appear in the Silliest of Places*, UNHERD (Jan. 10, 2023), https://unherd.com/2023/01/the-philistine-war-on-ai-art [https://perma.cc/4AE9-APSE] (discussing AI art as a collaboration between user and machine, in which the AI can produce novel results from the same prompt, and the user can learn to manipulate the results).

For example, if an autonomous vehicle commits a traffic violation, who should get the ticket?[15] The status quo—no one gets a ticket[16]—surely cannot persist. Or if AI creates a novel piece of artwork based loosely[17] on learning from hundreds or thousands or millions of other artists' work, should we consider this a financial windfall, copyright infringement, or something else entirely?[18] Botto, the decentralized

---

15. *See, e.g.*, Jeffrey K. Gurney, *Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles*, 2013 U. ILL. J.L. TECH. & POL'Y 247, 271 (2013) (arguing to place liability on the developer in order to incentivize safer autonomous technology); Matthew Blunt, *Highway to A Headache: Is Tort-Based Automotive Insurance on A Collision Course with Autonomous Vehicles?*, 53 WILLAMETTE L. REV. 107, 131 (2017) (looking to a federal insurance program to fill the liability gaps by proposing "a modified no-fault insurance system, which would treat a fully autonomous vehicle's manufacturer the same as a pure no-fault jurisdiction would treat an at-fault driver when the fully-autonomous vehicle's malfunctioning technology causes an accident"); John W. Zipp, Note, *The Road Will Never Be the Same: A Reexamination of Tort Liability for Autonomous Vehicles*, 43 TRANSP. L.J. 137, 171 (2016) (arguing that the autonomous vehicle, the "algorithm," should be considered the driver and therefore liable). Although the idea of issuing the ticket to the algorithm itself exists, AI today lacks the moral agency that would allow this situation to make sense. *See* Zipp, *supra* note 15, at 171; *see also* Nathan Reitinger, *Algorithmic Choice and Superior Responsibility: Closing the Gap Between Liability and Lethal Autonomy by Defining the Line Between Actors and Tools*, 51 GONZ. L. REV. 79, 90–92 (2016) (explaining that because robots do not currently possess moral agency and thus lack responsibility, there are now only three Lethal Autonomous Weapons Systems: "(1) in the loop—a human is in control of the machine; (2) on the loop—a human may be involved in the machine's actions but does not have to be; and (3) out of the loop—the autonomous robot selects and engages targets itself." In the not-yet-developed fourth loop, "the impetus for the machine's actions would not come from others, but from itself").

16. *See* Bigad Shaban & Michael Bott, *Driverless Cars Immune from Traffic Tickets in California Under Current Laws*, NBC NEWS (Dec. 30, 2023, 9:11 AM), https://www.nbcnews.com/business/business-news/can-driverless-cars-get-tickets-california-law-rcna131538 [https://perma.cc/4XDP-JXDC] ("An internal memo from San Francisco Police Chief Bill Scott. . . instructs officers that 'no citation for a moving violation can be issued if the [autonomous vehicle] is being operated in a driverless mode.'").

17. *See generally* Salman Kahn, Hossein Rahmani, Syed Afaq Ali Shah & Mohammed Bennamoun, *Convolutional Neural Network*, *in* A GUIDE TO CONVOLUTIONAL NEURAL NETWORKS FOR COMPUTER VISION 65–68 (2022) (discussing various loss functions for convolutional neural networks).

18. *See, e.g.*, Haochen Sun, *Redesigning Copyright Protection in the Era of Artificial Intelligence*, 107 IOWA L. REV. 1213, 1226 (2022) (noting how the debate of copyrightability starts on a more fundamental note, namely, can a machine claim copyright given the human-centered purpose of copyright); Yvette Joy Liebesman & Julie Cromer Young, *Litigating Against the Artificially Intelligent Infringer*, 14 FIU L. REV. 259, 261–63, 273 (2020) (engaging with the question of AI-as-litigant given a claim of

autonomous artist, banked over three million dollars during the recent non-fungible token craze.[19] The technology Botto used to make this happen, Stable Diffusion, is being sued for copyright infringement right now.[20]

Reasoned judgments on AI-based issues—more accurately, *machine learning*[21]—will not be had without a deeper understanding of both the machine and the learning happening under the hood.[22] Solutions to

---

copyright infringement); Jessica L. Gillotte, Note, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655, 2669–79 (2020) (discussing copyright infringement issues).

19. Botto, *Botto is a Decentralized Autonomous Artist*, https://www.botto.com [https://perma.cc/SNT2-YZQP]; Harsh Notariya, *How This AI Artist Made $3 Million as NFT Markets Fell*, BEINCRYPTO (Apr. 25, 2023, 11:00 AM), https://beincrypto.com/ai-artist-made-3-million-selling-nfts [https://perma.cc/P2TU-HDC5] (discussing how an AI, Botto, helps produce four- to eight-thousand AI-generated images per week, which are then hand-picked and sold online: "Botto minted its genesis artwork, Asymmetrical Liberation, in October 2021. Since then, the project has minted 75 creations, generating approximately $3 million in revenue").

20. *See* Botto, *Botto's Art Engine*, GITBOOK, https://docs.botto.com/details/bottos-art-engine [https://perma.cc/UPC7-6CTX] (explaining that Botto uses a combination of several software models including Stable Diffusion); *see also* James Vincent, *Getty Images Sues AI Art Generator Stable Diffusion in the US for Copyright Infringement*, VERGE (Feb. 6, 2023, 11:56 AM), https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion [https://perma.cc/M7Z5-2AWG] (reporting that Getty Images filed a lawsuit against the creators of Stable Diffusion for generating artwork using stock photography without the consent of original creators).

21. The current AI craze is not fueled by a new technology but is primarily driven by breakthroughs in a type of AI known as machine learning. Machine learning would fall into the "acting humanly" taxonomy of AI. *See* STUART JONATHAN RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1–5 (3d ed. 2010) (discussing the four buckets of AI: thinking humanly which involves the automation of human thinking, acting humanly which involves the automation of performing functions requiring intelligence, thinking rationally which involves automating reasoning and logic skills, and acting rationally which involves agents acting optimally).

22. *See, e.g.*, Deborah Tussey, *Technology Matters: The Courts, Media Neutrality, and New Technologies*, 12 J. INTELL. PROP. L. 427, 429–30 (2005) (discussing new technology under the lens of intellectual property rights); James J. Tomkovicz, *Technology and the Threshold of the Fourth Amendment: A Tale of Two Futures*, 72 MISS. L.J. 317, 322–24 (2002) (addressing how advancements in technology can induce unreasonable searches under the Fourth Amendment); Meg Leta Jones, *Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw*, 2018 U. ILL. J.L. TECH. & POL'Y 249, 253 (2018) (analyzing different approaches to cyberlaw and how they deal with new technology).

these "new technology"[23] problems exist, but not without a nuanced understanding of how AI works. Looking at the above examples in more detail will help solidify this need.

First, if an autonomous vehicle causes an accident, who should be responsible? On the face of it, this seems like one of the most important and most difficult problems for autonomous vehicles. Without a driver, we lack a liable actor sufficiently tied to the cause of an accident.

The fallacy here is that we are failing to fully appreciate how autonomous drivers are, in part, simply recipes. Looking at the driver as a tool, rather than something approaching a culpable actor, demonstrates that liability is more of a red herring than a solution to a problem.[24]

Machines (neural networks) are inescapably mathematical.[25] Although humans may speed to get to work on time, robots[26] will only act as they are programmed to act.[27] Therefore, in and of itself, moving

23. *See* Jaron Lanier, *There is No A.I.*, NEW YORKER (Apr. 20, 2023), https://www.newyorker.com/science/annals-of-artificial-intelligence/there-is-no-ai [https://perma.cc/48JU-ACEZ] (discussing the recent AI craze using a historical perspective).

24. *See* Zipp, *supra* note 15, at 168–70 (explaining the disadvantages that accompany holding the autonomous vehicle's manufacturer liable); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 127–28, 146 (2014) (arguing to hold manufacturers responsible, even in the case of an inexplicable event, in which the author argues that strict liability rules would come into play).

25. Computers, for the most part, are deterministic, which is why producing true randomness is hard. This also means that inexplicable events, given enough attention, are unlikely to be truly inexplicable. *See* Simson L. Garfinkel & Philip Leclerc, *Randomness Concerns When Deploying Differential Privacy*, *in* 19TH WORKSHOP ON PRIVACY IN THE ELECTRONIC SOCIETY 73, 76 (2020) ("The history of random numbers is littered with the corpses of methods that were known not to be truly random when they were deployed but were incorrectly thought to be good enough for the task at hand.").

26. Autonomous vehicles are robots. *See* Lance Eliot, *Self-Driving Cars and Asimov's Three Laws About Robots*, FORBES (Jan. 5, 2021, 11:30 AM), https://www.forbes.com/sites/lanceeliot/2021/01/05/self-driving-cars-and-asimovs-three-laws-about-robots [https://perma.cc/5YL2-XZ7F]; *see also* Reitinger, *supra* note 15, at 90 (internal citations omitted) ("The word 'robot' comes from the Czech word 'robata,' first used in the 1921 play *Rossum's Universal Robots*. The word means servitude, drudgery, or forced labor and was picked to connote ideations of a slave that eventually finds power and overthrows its human master.").

27. This is true even for buggy software: the actions of the machine may be unintentionally programmed, but programmed nonetheless (e.g., Apple's 'goto' fail bug was a bug, but the computer was simply doing what it was told). *See* David Lazar,

vehicle law is ill-equipped to handle autonomous drivers: the objects that will be regulated are tools, and tools do not care about legal forms of punishments (i.e., punishment regimes) unless they are programmed to care about legal forms of punishment.[28] Traffic violations, like speeding, are also *penumbral* crimes, crimes that lack social stigma and see a high level of noncompliance with a low level of enforcement.[29] Mathematical models, however, are none the wiser. These machines will blindly engage in whatever type of driving their programming instructs, regardless of what society has implicitly agreed on.[30] Finally, although autonomous vehicles may appear magical, in the Arthur Clarke sense,[31] these machines' core intelligence is only a mathematical function fit to data.[32] These are not reasoned decision-makers carefully weighing the moral, ethical, or legal consequences of their actions.

These are machines, metal and silicone, and this should directly impact our understanding of liability: someone is using the autonomous vehicle as a tool to achieve a goal, and that someone is likely a good candidate for culpability. This individual may be a manufacturer attempting to train a vehicle in an unsafe manner. It may

---

Haogang Chen, Xi Wang & Nickolai Zeldovich, *Why Does Cryptographic Software Fail? A Case Study and Open Problems*, *in* PROCEEDINGS 5TH ASIA-PACIFIC WORKSHOP ON SYSTEMS 1, 1 (2014) (explaining that Apple disclosed in 2014 a "'goto' bug in its SSL/TLS implementation" that "existed for over a year, during which millions, if not billions, of devices were vulnerable to man-in-the-middle attacks" and believed to be caused by a programmer accidentally adding "one redundant goto statement"); *see also* Reitinger, *supra* note 15, at 92 (discussing the idea of "fourth loop" machines: machines—not yet in existence, though would break this statement—that are fully volitional).

28. *See* Michele Cotton, *Back with a Vengeance: The Resilience of Retribution as an Articulated Purpose of Criminal Punishment*, 37 AM. CRIM. L. REV. 1313, 1313 (2000) (describing how states began to articulate the purposes of criminal punishment around the 1970s).

29. *See* Margaret Raymond, *Penumbral Crimes*, 39 AM. CRIM. L. REV. 1395, 1395 (2002) (introducing the concept of penumbral crimes and arguing for updates in the law to reflect current-day norms).

30. *See id.* at 1439 (arguing that the law should be adjusted to reflect social norms); *see also* Stefan Magen, *Philosophy of Law*, *in* INTERNATIONAL ENCYCLOPEDIA SOCIAL & BEHAVIORAL SCIENCES 24, 24 (2015) (discussing theories of legal norms, and noting lawyers' tendency to "talk about legal norms as if they were entities with an existence of their own" rather than as "a mere expression" of legislative intent).

31. ARTHUR C. CLARKE, PROFILES OF THE FUTURE: AN INQUIRY INTO THE LIMITS OF THE POSSIBLE 12, 21 n.1 (rev. ed. 1973) ("Any sufficiently advanced technology is indistinguishable from magic.").

32. *See infra* Part I (discussing the core idea behind machine learning: fitting a mathematical function to data).

also be an individual who decided it was a good idea to use the autonomous vehicle in terrain or conditions the individual knew or should have known to be dangerous.[33]

In short, while liability is one of the most pressing issues for autonomous vehicles, this problem is nuanced. It is nuanced because we need to appreciate that these machines are, at the bottom, merely following instructions to produce "correct" outputs based on input.[34] And this has implications beyond solving problems caused by the law, like liability, and warrants consideration of our laws themselves, which may not incentivize or protect as well as when humans were behind the wheel.

Second, consider the seemingly difficult question of copyright infringement due to AI-generated artwork. An AI necessarily learns how to create from somewhere and that somewhere is very likely protected by copyright.[35] However, the legal approach we use today to

---

33. As with any technology, the tool may be used appropriately or inappropriately, and this often does not, in fact, cause a problem for liability. Take automatic or "autonomous" headlights for example. These are vehicle headlights that are designed to turn on without human intervention when lighting in an area is dim. Several popular vehicle manufacturers offer this feature, which has given the courts a chance to debate whether liability is a showstopper when traffic stops are made pursuant to failed engagement of these autonomous headlights. To be sure, courts have unanimously considered this a non-issue; stops made pursuant to failed headlights, even when those headlights are supposedly automated, are nonetheless routine traffic stops. There are no zebras here. *Cf.* United States v. Collier, No. EDCR 13-19 JGB, 2015 WL 11123302, at *3 (C.D. Cal. Nov. 10, 2015) (finding no error in a police officer's stop when defendant's headlights were not in use, although the headlights were claimed to be in "automatic" mode); United States v. Lawrence, No. 1:17-CR-126-LMM-CMS, 2018 WL 2275557, at *3 (N.D. Ga. Apr. 27, 2018) (same). This Essay is too short to handle the larger issue of autonomous liability effects, but notes that, similar to other areas where partial autonomy is involved, the human controlling (even in a loose sense) the machine should likely be responsible for the machine's misgivings. *Cf.* Reitinger, *supra* note 15, at 118–19 (arguing for holding superiors responsible for the misgivings of autonomous weapons systems, rather than allowing autonomous weapons systems to create a liability cavern).

34. Kristin Rheins, *The Debate over Liability for AI-Generated Content*, PROGRESSIVE POL'Y INST. (Aug. 8, 2023), https://www.progressivepolicy.org/blogs/the-debate-over-liability-for-ai-generated-content [https://perma.cc/PN9W-Q662].

35. *See* Tremblay v. OpenAI, Inc., No. 23-cv-03223-AMO, 2024 WL 557720, at *1 (N.D. Cal. Feb. 12, 2024) (explaining that authors claimed in a class action that ChatGPT's use of training data violated their copyrights); *see also* Teresa Xie & Isaiah Poritz, *ChatGPT Creator OpenAI Sued for Theft of Private Data in 'AI Arms Race'*, BLOOMBERG (June 28, 2023, 7:15 PM),

identify copyright infringement does not address the novel problems created by AI.

As background, copyright infringement exists when a plaintiff can show (1) access and (2) substantial similarity.[36] The current way to analyze these cases, given possible AI infringement, would be to look at the generative AI's training data to prove access, and then visually compare the generated content with the protected content to prove similarity.[37] The shortsightedness of this approach, however, is that it fails to fully appreciate the learning part of machine learning.

The AI model's architecture is already conveniently optimized to address issues concerning similarity.[38] As discussed below,[39] in order to gain the skill of image generation, the AI will have undergone training and learned how to create good images, and "good" was very likely quantified using a test of similarity.[40] In fact, a typical loss function (i.e., a measure of goodness) for a generative, image-based AI works by considering more dissimilarity to training data as more loss and

https://www.bloomberg.com/news/articles/2023-06-28/chatgpt-creator-sued-for-theft-of-private-data-in-ai-arms-race?in_source=embedded-checkout-banner [https://perma.cc/2NTH-538W] (discussing a lawsuit claiming that OpenAI, the company responsible for creating of ChatGPT, improperly created ChatGPT by using stolen, personal data in the training process).

36. *See* 17 U.S.C. § 501(a) (infringing copyright); 18 C.J.S. Copyrights § 122 (citing Kay Berry, Inc. v. Taylor Gifts, Inc., 421 F.3d 199, 207–08 (3d Cir. 2005); Fogerty v. MGM Grp. Holdings Corp., 379 F.3d 348, 352 (6th Cir. 2004)).

37. *See, e.g.*, CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW, 4–5 (2023) (discussing training data and substantial similarity at a surface level).

38. A common approach for image generation is to use a generative adversarial network. *See* Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta & Anil A. Bharath, *Generative Adversarial Networks: An Overview*, *in* IEEE SIGNAL PROCESSING MAG. 53 (2018) (discussing backpropagation (i.e., the learning of machine learning) as a competitive process involving a set of deep neural networks); *see also* Teodora Baluta, Ivica Nikolić, Racchit Jain, Divesh Aggarwal & Prateek Saxena, *Unforgeability in Stochastic Gradient Descent*, *in* PROCEEDINGS OF THE ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY (2023) (developing the concept of unforgeability to prove, in a practical application, that copyrighted material was used by a machine learning model—a more verifiable and incontrovertible proof of similarity than the idea posed above).

39. *See infra* Section I.B.

40. *See* Yingjie Tian, Duo Su, Stanislao Lauria & Xiaohui Liu, *Recent Advances on Loss Functions in Deep Learning for Computer Vision*, 497 NEUROCOMPUTING 129, 129 (2022) (stating that loss functions for machine vision typically measure "accuracy, similarity, or goodness of fit between the predicted value and ground-truth").

therefore something to be improved upon.[41] Knowing what exact loss function was being used and how similar the offending image is to the AI's training data, per the AI's own architectural terms, would be invaluable to a plaintiff and to the court.

Without understanding the nuances of AI, we are primed to inadequately address AI-based issues. And these issues have consequences, both for AI and for those parties involved. This Essay teaches the underlying technological makeup of AI to give readers the ability to accurately reason about any problem that involves an AI. Although the field of AI is advancing at a rapid pace, making any per-technique, per-architecture, or per-model analysis obsolete in a few years,[42] just as ChatGPT replaced alphaGo replaced deepBlue,[43] there are two key concepts underlying the entire discipline which will persist for years to come: neural networks (the machine) and gradient descent (the learning). Moreover, these two concepts, so far, have been given short shrift in legal scholarship.[44]

---

41. Jake Snell, Karl Ridgeway, Renjie Liao, Brett D. Roads, Michael C. Mozer & Richard S. Zemel, *Learning to Generate Images with Perceptual Similarity Metrics*, *in* 2017 IEEE INT'L CONF. IMAGE PROCESSING PROCEEDINGS, 4277, 4278 (2017).

42. It would be a mistake to frame this Essay around a tool like ChatGPT, which may rise or fall in popularity in the coming years. *See* Steven M. Bellovin, Preetam K. Dutta & Nathan Reitinger, *Privacy and Synthetic Datasets*, 22 STAN. TECH. L. REV. 1, 30 (2019) (discussing recurrent neural networks, which have been replaced by transformers in recent years); *see also* Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi et al., *Transformers: State-of-the-Art Natural Language Processing*, *in* PROCEEDINGS OF THE 2020 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING 38 (2020) (discussing the "transformer" technique used by ChatGPT—a Generative Pretrained Transformer).

43. *See supra* note 1 and accompanying text (discussing the historical evolution of AI). This replacement scheme is inexact, but only offered to prove that AI is a continually advancing field, one which has had many ups and downs. ChatGPT for example, is currently on its fourth iteration. *See* Complaint at 5, Tremblay v. OpenAI, No. 3:23-cv-03223-AMO (N.D. Cal. 2023) ("OpenAI has released a series of large language models, including GPT-1 (released June 2018), GPT-2 (February 2019), GPT-3 (May 2020), GPT-3.5 (March 2022), and most recently GPT-4 (March 2023).").

44. Although many have written extensively about AI generally, or machine learning generally, few have provided the necessary details that permit an understanding of AI that is correct, yet simple. *See, e.g.*, Emile Loza de Siles, *AI, on the Law of the Elephant: Toward Understanding Artificial Intelligence*, 69 BUFF. L. REV. 1389, 1429–31 (2021) (introducing machine learning using patent applications); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 717 (2017) (discussing phases of machine learning like data collection and training). For example, the canonical way of teaching a neural

These two components are taught in this Essay as they would be taught in most computer science curricula: a machine can be thought of as a single-layer perceptron and learning can be thought of as stochastic gradient descent. Once this framework is understood, everything else in AI—from ChatGPT to Bard,[45] or to whatever fad comes next—can be thought of as an optimization, a simple tweak to make the AI a little bit faster, smarter, or lighter.[46]

Taking this understanding one step further, the Essay pivots to set these two concepts in context by analyzing the recent Ever affair.[47] In this case, the Federal Trade Commission (FTC) debuted its use of model disgorgement as a remedy for ill-trained AI models.[48] While it

---

network is through the single-layer perceptron. However, this concept is either introduced hastily or taught in detail, but without connecting single-to-multi-layered concepts, which is the reason these networks are being used in the first place. *See, e.g.*, John Nawara, *Machine Learning: Face Recognition Technology Evidence in Criminal Trials*, 49 U. LOUISVILLE L. REV. 601, 608 (2011) (discussing a single-layer perception for the purpose of stating that it is only able to "deal with" 2-D patterns); Brian S. Haney, *The Optimal Agent: The Future of Autonomous Vehicles & Liability Theory*, 30 ALB. L.J. SCI. & TECH. 1, 10 (2020) (introducing a single-layer perceptron only in terms of how it may be layered to form a deeper neural network, but failing to give breathing room to teaching the single-layer network); Tzipi Zipper, *Mind over Matter: Addressing Challenges of Computer-Generated Works Under Copyright Law*, 22 WAKE FOREST J. BUS. & INTELL. PROP. L. 129, 153–55 (2022) (teaching the single-layer perceptron, but: (1) using a more complex mathematical representation which may be broken down further to aid understanding; (2) failing to connect the singer layer to a multi-layer architecture; and (3) most importantly, failing to show how the perceptron usefully solves a problem such as accurately representing a logical "and" gate).

45. Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang et al., *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT*, ARXIV 1, 4 (May 1, 2023), https://arxiv.org/pdf/2302.09419.pdf [https://perma.cc/HNM4-PZCJ] (discussing pretrained models, like BERT and ChatGPT, which aim "to train a general model using large amounts of data and tasks that can be fine-tuned easily in different downstream applications").

46. *See* Jeremy Howard, *Lesson 3: Practical Deep Learning for Coders 2022*, YOUTUBE (July 21, 2022), https://youtu.be/hBBOjCiFcuo?t=2870 [https://perma.cc/C3SM-2ZPP] (describing, at 47:53, the core of deep learning). Lighter means less data use, faster means compute faster, smarter means more accurate.

47. *See* Complaint, Everalbum, Inc., No. C-4743 (F.T.C. May 6, 2021) [hereinafter Everalbum Complaint], https://www.ftc.gov/system/files/documents/cases/1923172_-_everalbum_complaint_final.pdf [https://perma.cc/ZDL9-EZRK]; *infra* notes 105–119 and accompanying text. Everalbum is the offending photo-storage app owned by Ever. *Id.*

48. *See* Rebecca Kelly Slaughter, Janice Kopec & Mohamad Batal, *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*,

may come as a surprise to some, this settlement agreement was not as far-reaching as it seemed.

The limitations of the Ever affair and a more refined understanding of AI may be made lucid via analogy: machine learning is like a perpetual stew.[49] Machine learning models are created using a recipe and then left to live their lives—simmer—and make predictions on new inputs, doing this accurately for as long as they are maintained. In this way, these models may be thought of as perpetual, mathematical Frankensteins, moving beyond their humble training-loop-inspired creations and taking on a life of their own.

Turning back to the FTC with this analogy, it is easy to see the limitations inherent in the settlement: the FTC did not tell Everalbum to throw away its recipe, nor did the FTC prohibit Everalbum from making another stew.[50] Everalbum would have been free to use what it learned when making its original stew, tweak the recipe with a few new ingredients, and make the exact same stew again. Moreover, given the industry standard for model maintenance,[51] this option would not have been too far-reaching.

In the first Part, this Essay turns to introducing AI as it really is: part machine and part learning. The second Part discusses novel legal issues created by AI, intuitive solutions, and the Everalbum affair as a case study.

---

23 YALE J.L. & TECH. 1, 38–39 (2021) (discussing FTC's "innovative remedy" of "algorithmic disgorgement" that required Everalbum to "delete not only the ill-gotten data but also any facial recognition models or algorithms developed with users' photos or videos"); *see also* Tiffany C. Li, *Algorithmic Destruction*, 75 SMU L. REV. 479, 498–99 (2022) (noting the "rather radical" remedy and discussing model training and weights persistence).

49. In short, both a perpetual stew and a machine learning model undergo a creation phase and then a life phase. In creation, the model is born, being iteratively updated, typically following a recipe, until it is eventually considered complete. From there, it moves on to the life phase, where, like the stew, it lives on, interacting with a world far removed from its initial training input until its maintainer abandons it. *See infra* Section II.A.

50. Everalbum Complaint, *supra* note 47.

51. *See infra* note 118 and accompanying text.

I. WHAT IS AI?

AI, today,[52] is machine learning, and machine learning is a way of fitting a function to data.[53] While this definition sounds overly simplified, the concept in context is surprisingly basic. A hypothetical will be illustrative.

Imagine we are trying to predict whether someone would have survived the Titanic disaster.[54] Further imagine the data[55] we have tells us who did or did not survive, and provides us with a list of attributes associated with each individual (e.g., age, sex, ticket class, ticket cost, number of siblings aboard, number of parents aboard, cabin number, and point of embarkment).[56] The function for this data would be a mathematical model that tries to make a predictive output given a provided set of attribute inputs—e.g., is an individual with this class ticket and this age likely or not likely to survive? Modeling the scenario with mathematics is intuitive: some attributes matter a lot (e.g., individuals with a costly ticket, a proxy for wealth, might have a higher chance of surviving) while others matter a little (e.g., the number of siblings aboard may not mean much).[57]

52. AI may be said to refer to a broad number of tools. *See generally* RUSSELL & NORVIG, *supra* note 21, at 1. The latest, and current, trend for these tools is the more specific category of machine learning. *See* MEHRYAR MOHRI, AFSHIN ROSTAMIZADEH & AMEET TALWALKAR, FOUNDATIONS OF MACHINE LEARNING 1 (2018) ("Machine learning can be broadly defined as computational methods using experience to improve performance or to make accurate predictions.").

53. *See* Howard, *supra* note 46 (describing these concepts at 23:46).

54. *See* Will Cukierski, *Titanic—Machine Learning from Disaster*, KAGGLE (2012), https://www.kaggle.com/competitions/titanic [https://perma.cc/JJM9-PJT2] (introducing the legendary Titanic competition that: "use[s] machine learning to create a model that predicts which passengers survived the Titanic shipwreck").

55. Without data, there would be no machine learning. *See* JEREMY HOWARD & SYLVAIN GUGGER, DEEP LEARNING FOR CODERS WITH FASTAI & PYTORCH: AI APPLICATIONS WITHOUT A PHD 25 (2020) (listing some of the limitations of machine learning: models do not exist without data; models can only see patterns in the data they learn on; models make predictions, not actions; and it is not enough to have data, labels are needed to give that data meaning (e.g., in image classification of dogs and cats, models require labels to indicate which images are dogs and which are cats)).

56. Weights are most often referred to as "parameters" in typical machine learning parlance. *See id.* at 22. For consistency, these parameters are referred to as weights throughout this Essay.

57. *See* Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman & C.V. Jawahar, *Cats and Dogs*, *in* IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 3498, 3498 (2012) (discussing improvements in the computer visions task of animal breed

Machine learning is a way of using what is already known and fitting it to a function: can the machine figure out on its own which attributes matter a lot and which attributes matter a little? Once the machine has learned how much something matters (i.e., the weights[58]), then the machine can predict survival.[59] The machine uses what it learned from existing data to make an educated guess about new individuals who have different, but similar, attributes. The next Section dives deeper into how this concept is represented with mathematics so that computers can understand it.

A. *The Machine—e.g., Single-Layer Perceptron*

The "machine" in machine learning is something that takes in input and produces out output.[60] Imagine we have a series of (x,y) data points. This hypothetical machine will serve[61] a very simple purpose: given an input, in the form of (x,y) pairs, produce an output of either *zero* or *one* as defined in the following table.

---

recognition—a notably difficult problem because, unlike the Titanic example from above, "animals, particularly cats, are very deformable and there can be quite subtle differences between the breeds").

58. *See* Howard, *supra* note 46 (demonstrating how inserted parameters ensure the precision of machine learning); HOWARD & GUGGER, *supra* note 55, at 22 (clarifying that "weights" is another term for model parameters).

59. *See* HOWARD & GUGGER, *supra* note 55, at 22 (explaining that a machine learning model is "trained" and able to deliver results once the developer selects a weight assignment—ideally one that optimizes performance—and this weight assignment is embedded in the model); *Mathematical Functions Power Artificial Intelligence*, NAT'L ACAD. OF SCI, https://nap.nationalacademies.org/resource/other/deps/illustrating-math/interactive/pdf/Machine-learning.pdf [https://perma.cc/AU77-LVVQ] ("A machine learning model can be understood as a mathematical "function" that depends on "parameters" learned from training data. Data scientists must choose the appropriate mathematical function, often trying many different possibilities.").

60. Similar to any algorithm. Nathan Reitinger & Amol Deshpande, *Epsilon-Differential Privacy, and a Two-Step Test for Quantifying Reidentification Risk*, 63 JURIMETRICS 263, 279 Fig. 1 (2023) (providing an algorithm for making a peanut butter and jelly sandwich).

61. *See supra* note 15 and accompanying text.

Table 1. Truth Table: "And"[62]

Line no.	<x>	<y>	Output
1:	0	0	0
2:	0	1	0
3:	1	0	0
4:	1	1	1

Table 1 represents a truth table.[63] For instance, if you see, as line one shows, an input of (0,0), then you know the answer should be 0. The truth table lets us know the correct (i.e., true) output based on certain inputs.

How could a machine complete this, the task of producing the *right* answer according to the truth table? One very simple way might be for the machine to hard-code the answers.[64] Whenever the machine sees an input, it looks in the truth table and finds the answer.[65] This may be a quick process with only four possible values, but what if we had a nearly infinite set of input numbers? The truth table would quickly become massive and cumbersome to carry around. Equally as important, the lookup process would slow as more and more values are added to the table.[66]

A better solution is to let the machine, by manipulating the data, arrive at the correct answer. This idea was conceptualized by Frank

62. This is the logical "and" gate. *See* JOHN E. SAVAGE, MODELS OF COMPUTATION: EXPLORING THE POWER OF COMPUTING 35–39 (Creative Commons ed. 2008) (discussing logical circuits).

63. *Id.*

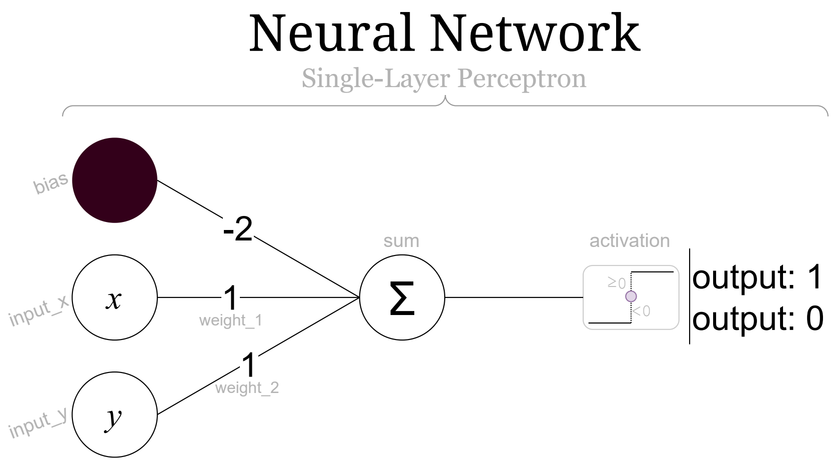
64. *See* Nathan Reitinger, *Perpetual Stew Codebook*, GITHUB (Apr. 20, 2024), https://github.com/nathanReitinger/perpetualStew/blob/main/Perpetual_Stew_Codebook.ipynb [https://perma.cc/TV5C-D3RH] (providing, in the first "CODE A" section, Python code to create this type of dictionary); *see also* Peeratham Techapalokul, *Sniffing Through Millions of Blocks for Bad Smells*, *in* SIGCSE '17: 48TH ACM TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION 781, 781–82 (2017) (discussing the frowned-upon practice of hard coding in the programming context).

65. *See, e.g.*, Athena Ozanich, *A Beginner's Guide to Python Dictionaries*, HUBSPOT (May 16, 2023), https://blog.hubspot.com/website/python-dictionary [https://perma.cc/7ULA-YTK2] (discussing how Python dictionaries can be used to store and lookup information).

66. *See* THOMAS H. CORMEN, CHARLES E. LEISERSON, RONALD L. RIVEST & CLIFFORD STEIN, INTRODUCTION TO ALGORITHMS 43–44 (2009) (discussing asymptotic efficiency: the study of how the "running time of an algorithm increases with the size of the input").

Rosenblatt in the 1950s.[67] This is called a neural network. The following neural network is known as a single-layer perceptron.[68]

Figure 1. Single Layer Neural Network: Logical "And" Gate [69]



The network receives a set of inputs, an *x* value and a *y* value, processes the input in a repeatable way (i.e., an algorithm[70]), and then produces an output that is correct according to the truth table. This is the machine part of machine learning. Figure 2, below, walks step-by-step through how to read the neural network shown in Figure 1.

Figure 2. Input and Output: How to Apply the Weights

How To

(x,y)	*bias*	+	(*input_x* * *weight_1*)	+	(*input_y* * *weight_2*)		*activation*	*output*
(0,0):	-2	+	(0 * 1)	+	(0 * 1)	= -2	<0	0
(0,1):	-2	+	(0 * 1)	+	(1 * 1)	= -1	<0	0
(1,0):	-2	+	(1 * 1)	+	(0 * 1)	= -1	<0	0
(1,1):	-2	+	(1 * 1)	+	(1 * 1)	= 0	≥0	1

67. *See* F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, 65 PSYCH. REV. 386, 389 (1958). A simpler form of this network was envisioned by McCulloch and Pitts in 1943. *See* Warren S. McCulloch & Walter Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, 5 BULL. MATHEMATICAL BIOPHYSICS 115, 115 (1943) (conceptualizing the nervous system as a "net of neurons" whose behavior can be explained in terms of propositional logic).
68. *See* Rosenblatt, *supra* note 67, at 387.
69. *See* SAVAGE, *supra* note 62, at 35–39.
70. *See* Reitinger & Deshpande, *supra* note 60, at 278–79.

In summary, the neural network in Figure 1 is given a set of (x,y) input pairs (noted as *input_x* and *input_y*), adds to these inputs a series of its own numbers (bias is added and weights are multiplied to each input value), sums everything together, and then checks whether the resulting sum meets a predefined threshold (i.e., an activation function, here, less than zero or equal to or greater than zero).[71] And it works! The machine correctly reproduces the truth table noted above for all possible inputs.[72] The output—the prediction—is 100% accurate, given that we know all possible (x,y) pairs and what the correct output should be for each pair.[73]

This simple idea is what gives rise to even the most complex AI architectures today, like ChatGPT: a "deep" neural network. To see where the deep part comes from, imagine that instead of the truth table noted above, we wanted to create another neural network for a different truth table.

Table 2. Truth Table: Logical "XOR" Gate[74]

Line no.	<x>	<y>	Output
1:	0	0	0
2:	0	1	1
3:	1	0	1
4:	1	1	0

Running the same weights and bias value from the "and" neural network will not produce a correct output. Even updating the weights and bias value with new numbers will not produce a correct output.

What we can do instead is combine the "and" perceptron with a few others (i.e., an "or" perceptron and a "not and" perceptron) and form

71. The only input to receive a "one" as output is the (1,1) pair. This is the "AND" gate and requires each x and y input to be true (i.e., one) for a "true" output to occur.

72. See the "How To" section of Figure 2 *infra* for all possible inputs in long-form calculation. *See also* Reitinger, *supra* note 64 (providing, in CODE B, the code for this type of network).

73. Most of the problems being solved by AI today have no clear answer, which is why accuracy—and bias, because some inputs are treated differently than other inputs—play a large role. *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 675 (2016) (discussing how various aspects of machine learning can lead to unfair outcomes).

74. *See* SAVAGE, *supra* note 62.

a layered architecture.[75] This layering in a neural network is where the term "deep" comes from. The network is deep in that it processes input in successive layers, combining these layers to produce a single output in the end. The figure below visualizes a two-layer, deep neural network.

Figure 3. Deep Neural Network: "XOR"

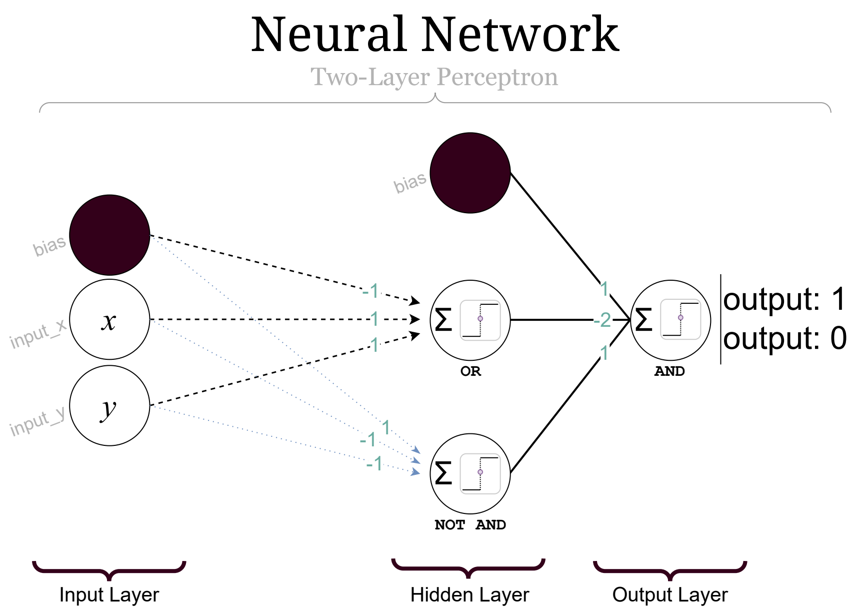


Figure 3 illustrates a *deep* neural network. Deep neural networks may sound complicated, but are simply a series of layered single-layer neural networks. In practice, these networks can have a nearly unlimited number of layers, move data forward through the network, backward through the network, or a combination of each, and each of these layers can have simple or complex jobs[76]—but the same idea underlies all networks. Some initial input is received, the network

75. *See* Alan Lapedes & Robert Farber, *How Neural Nets Work*, *in* NIPS'87: PROCEEDINGS OF THE 1987 INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS 442, 443–44 (1987).

76. *See, e.g.*, Kahn et al., *Examples of CNN Architectures*, *supra* note 17, at 101–15 (discussing various computer vision, particularly convolutional neural network, architectures).

processes that input and produces an output, which is then evaluated against some measure of truth.

This leads to the second foundational component of machine learning: the learning. Learning involves the evaluation of output and the process of updating the network's variables (bias, weights, and many other variables) to achieve correct, better predictions.[77]

B. The Learning—e.g., Gradient Descent

This Part began by claiming that machine learning is the process of fitting a function to data.[78] Section B discusses the fitting part. Fitting helps solve for the variables that were hard-coded (i.e., the weights stated as known) in the previous Section's visualization of neural networks.

What if the weights and bias terms were unknown, but the truth table was known? Can the machine learn on its own what the bias and weights should be to produce the correct outputs? Yes.

The "learning" part of machine learning is, at the bottom, basic math: an iterative process of running inputs through the network and checking the error of the outputs—how far off from the truth are the outputs? The network then updates its variables in a systematic way that will produce more and more correct output on successive loops. This process is formally known as backpropagation.[79] The magic here occurs in knowing how to adjust the weights in a systematic way: gradient descent. An example will be helpful.

Given some data, like the Titanic dataset from above,[80] the attributes (e.g., age, sex, etc.) can be represented as a quadratic function (a way

---

77. Yulong Lu & Jianfeng Lu, *A Universal Approximation Theorem of Deep Neural Networks for Expressing Probability Distributions*, *in* PROCEEDINGS OF THE 34TH INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 1, 1 (2020) (discussing the "learning" of machine learning, which is, at bottom, the ability of networks to approximate functions). For an example of a deep neural network trained to learn the simple "AND" gate from Table 1, see Reitinger, *supra* note 64 (providing, in CODE C, the code for a simple machine learning task—weight print outs and architecture also included).

78. *See* Howard, *supra* note 46.

79. *See* Timothy P. Lillicrap, Adam Santoro, Luke Marris, Colin J. Akerman & Geoffrey Hinton, *Backpropagation and the Brain*, 21 NATURE REVS.: NEUROSCIENCE 335, 335 (2020).

80. *See supra* note 54 and accompanying text.

for the machine to recognize it).[81] This allows the problem to be framed in a way that a computer can understand it and attempt to "fit" it by learning the weights. The function[82] would look like this.

(age * x1) + (sex * x2) + (ticket-class * x3) + . . . = known output

The *x* values, the variables, will be the weights, applied to each attribute of age, sex, and ticket class (these attributes would also be turned into numbers, e.g., one-hot encoding[83]). With a *set* of data, there would be many rows of these equations.

(age_1 * x1) + (sex_1* x2) + (ticket-class_1 * x3) + . . . = known output

(age_2 * x1) + (sex_2 * x2) + (ticket-class_2 * x3) + . . . = known output

(age_3 * x1) + (sex_3 * x2) + (ticket-class_3 * x3) + . . . = known output

. . .

We need to start somewhere when picking the initial weights (*x*1, *x*2 . . .), and one very simple method, which actually happens in practice, is to pick the weights randomly at first.[84]

(age * .24) + (sex * .20) + (ticket-class * .90) + . . . = known output

The output of this calculation is the prediction, using some type of activation function to produce a 0 or 1. For example, if the value of this calculation is more than 0, then the output might be called "survived"; otherwise "not survived."

The first prediction is most likely incorrect given the randomness used to pick the initial weight values, but it provides information on how to update the *x* values. With the predictions from the initially

81. *See* KHAN ACADEMY, *The Quadratic Formula*, https://www.khanacademy.org/math/algebra/x2f8bb11595b61c86:quadratic-functions-equations [https://perma.cc/GDS9-BJRQ] (discussing quadratic functions).

82. *See* Howard, *supra* note 46 (discussing, at 1:04, the process of coding this example as a Microsoft Excel spreadsheet).

83. One-hot encoding is the process of adding numerical values to categorical data. *See* Pau Rodríguez, Miguel A. Bautista, Jordi Gonzàlez & Sergio Escalera, *Beyond One-Hot Encoding: Lower Dimensional Target Embedding*, 75 IMAGE & VISION COMPUTING 21, 21 (2018).

84. *See* Howard, *supra* note 46 (discussing, at 1:08–1:10 the process of using random numbers to start the calculations).

random model, a loss function[85] would then be used to assess how far off the predictions are from a truth table[86] (i.e., the output of survived or not survived).[87] Loss would be taken as a whole, assessing the total loss on all equations in the dataset, and using gradient descent to figure out how much to adjust each of the weights to reduce the loss, each of the *x*1, *x*2, and *x*3 values. The exact details of gradient descent are not covered in this short Essay,[88] but what is important is that the learning is not occurring randomly; there is a precise way to adjust the weights to reduce the total loss on successive loops. This is where the term *epoch* comes from: neural networks are trained in loops, where one loop is referred to as an epoch.[89] This process of setting weights and evaluating output to update the weights is run over and over again until the performance (i.e., the amount of loss) of a final model is deemed satisfactory.[90] The final model's architecture and weights are

85. Many loss functions exist. One of the simplest is mean squared error. In this loss function, the difference between a predicted value and the actual value is squared and then averaged across all predictions made. *See* Zhou Wang & Alan C. Bovik, *Mean Squared Error: Love it or Leave it? A New Look at Signal Fidelity Measures*, IEEE SIGNAL PROCESSING MAG., Jan. 2009, at 98, 99 (discussing the mean squared error and why it is helpful). Some loss functions are particular to data types, like the loss functions associated with images, as noted above.

86. In supervised learning, the learning is done by training from known answers, similar to having a truth table. Because truth is hard to come by, other techniques exist which try to learn truth instead of building from it. *See* Fabrizio Carcillo, Yann-Aël Le Borgne, Olivier Caelen, Yacine Kessaci, Frédéric Oblé & Gianluca Bontempi, *Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection*, 557 INFO. SCI. 317, 317–18 (2021) (discussing the relative strengths of each approach in the context of detecting credit card fraud).

87. At a high level, what we know (i.e., our labeled data) would typically be split into sub datasets, one used for training the network and one used for testing the network. *See* HOWARD & GUGGER, *supra* note 55, at 28–29 (discussing the test-train split of 20% to 80% of the data, respectively).

88. There are many types of gradient descent algorithms. *See, e.g.*, Sebastian Ruder, *An Overview of Gradient Descent Optimization Algorithms*, ARXIV 1, 2 (June 15, 2017), https://arxiv.org/pdf/1609.04747.pdf [https://perma.cc/F65Y-J6WS] (discussing popular gradient descent algorithms like stochastic gradient descent and adaptive moment estimation).

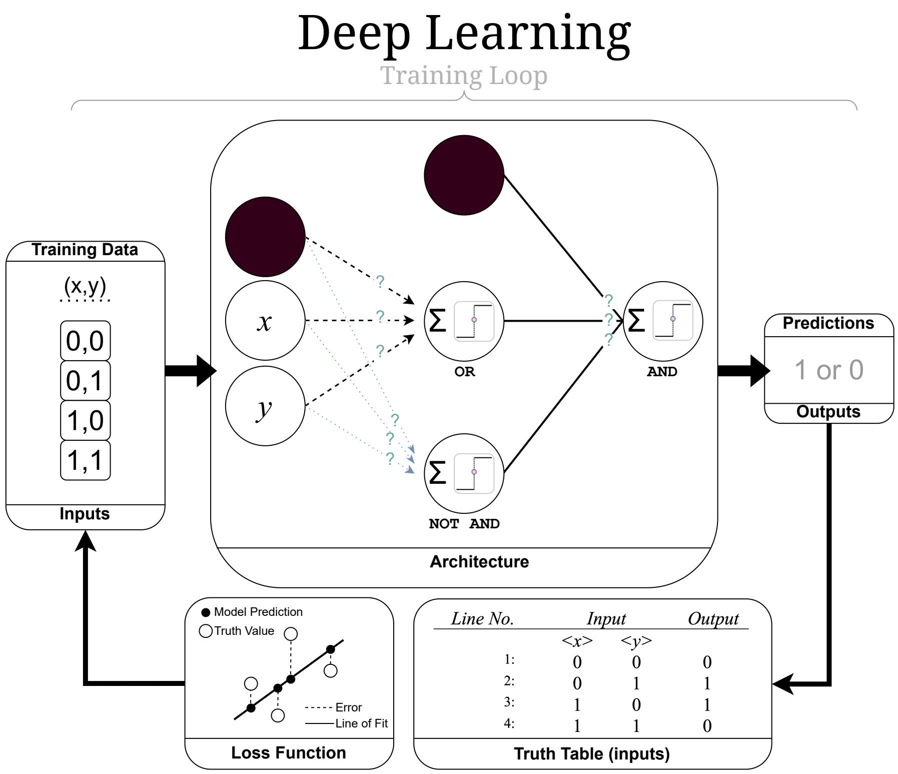
89. *See* Yasusi Kanada, *Optimizing Neural-Network Learning Rate by Using a Genetic Algorithm with Per-Epoch Mutations*, INT'L JOINT CONF. NEURAL NETWORKS 1472, 1473–74 (2016) (discussing epochs and learning rate algorithms).

90. This process hints at a further consideration, which is when to stop training the network. *See* LAURENCE MORONEY, AI AND MACHINE LEARNING FOR CODERS: A PROGRAMMER'S GUIDE TO ARTIFICIAL INTELLIGENCE 30–32 (2021).

then set, ready to receive new inputs and make a prediction on future inputs using what it learned from training.

As the following figure demonstrates, this process looks very similar to the prior networks shown in Figures 1 and 3. Data is passed in as input, manipulated in a series of layers to bring about a final number, which is then passed through an activation function to produce a well-formed output (e.g., zero or one in the binary classification case discussed so far). The outputs are then compared against the known truth and, according to the loss function and gradient descent, the weights are updated for another round of training.

Figure 4. Deep Learning[91]



91. *Inputs*: input can represent anything, but it needs to be mathematically expressed. *Architecture*: this piece of learning involves the type of network or networks being used; the layers are columns of circles (or nodes, as neurons). *Outputs*: models can return a variety of outputs, sometimes in the binary (0 or 1) but other times in categorical (words) or something like a range of numbers; knowing beforehand what a correct output is and teaching the machine using that *a priori* truth is known as supervised learning. *Truth Table*: in this supervised learning task, we are evaluating the predictions against the known correct answer using the truth table. *Loss Function*: many types of loss functions exist, the one pictured above shows the process of calculating the distance of the error (dotted line) based on the predicted value versus the true value; each of the guesses at predicting a true answer would be called an example, and typically we would look at loss among all errors as a whole. *See generally id.* at 13–18 (discussing the training of a simple neural network).

Running this process over and over again will result in a model that arrives at an accurate output on future, unknown input.[92] Although the training data may now be discarded, leaving the architecture and weights, it is important and necessary to maintain a process of retraining due to stagnating data over time—in fact, production models use pipelines for automated retraining given the need and cost of one-off retraining.[93]

To summarize, machine learning is equal parts machine and learning. Together, this representation of problems allows us to make accurate predictions on any real-world problem that may be represented mathematically, which turns out to be most problems.[94] In the next Part, we use a recent example of problematic AI to illustrate how a more refined understanding of AI helps us reason about how the law does (or should) apply.

II. UNINTUITIVE CONSEQUENCES

A correct understanding of the underlying nature of AI can drive analysis—often in unintuitive ways—of whether the law *does* or *should*

---

92. The field of explainability and interpretability in AI is designed to give explanation to these predictions, which are made on never-before-seen input and are therefore difficult to judge. *See generally* Wojciech Samek & Klaus-Robert Müller, *Towards Explainable Artificial Intelligence*, *in* EXPLAINABLE AI: INTERPRETING, EXPLAINING AND VISUALIZING DEEP LEARNING 5, 7–8 (Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen & Klaus-Robert Müller eds. 2019) (discussing "Clever Hans," a horse in the 1900s that was considered to be able to count until it was uncovered that Hans was merely "predicting" the right numbers based on body language; this illustrates how a "right" answer may be predicted for the "wrong" reason, and how "right" answers should not be taken at face value).

93. Geoffrey I. Webb, Roy Hyde, Hong Cao, Hai Long Nguyen & Francois Petitjean, *Characterizing Concept Drift*, 30 DATA MINING & KNOWLEDGE DISCOVERY 964, 964 (2016) ("Most machine learning models are static, but the world is dynamic, and increasing online deployment of learned models gives increasing urgency to the development of efficient and effective mechanisms to address learning in the context of non-stationary distributions, or as it is commonly called *concept drift*."); *see also* HOWARD & GUGGER, *supra* note 55, at 22–23 ("[O]nce the model is trained—that is, once we've chosen our final, best, favorite weight assignment—then we can think of the weights as being *part of the model*, since we're not varying them anymore.").

94. For example, images can be represented as a set of pixels with each pixel ascribed a color-mapped number value: allowing us to represent image with math. *See* M. I. Jordan & T. M. Mitchell, *Machine Learning: Trends, Perspectives, and Prospects*, 349 SCI. 255, 256 (2015) (discussing the many applications of machine learning); Saptarshi Sengupta, Sanchita Basak, Pallabi Saikia, Sayak Paul, Vasilios Tsalavoutis, Frederick Ditliac et al., *A Review of Deep Learning with Special Emphasis on Architectures, Applications, and Recent Trends*, 194 KNOWLEDGE-BASED SYS. 1, 12–21 (2020) (same).

apply to AI issues. The Everalbum affair is a good example of this.[95] Before describing this case, however, it will be useful to start with an analogy. The analogy both helps concretize the concept of machine learning discussed above and highlights, for the case of Ever, easy-to-miss limitations inherent in the settlement reached between the FTC and Ever. To be sure, although scholars have been quick to praise the far-reaching implications of the FTC's settlement,[96] and it is a step in the right direction, this type of AI-based remedy may not be as revolutionary as claimed.

A. AI Is Like a Perpetual Stew

The perpetual stew is a historic concept involving the creation of a soup that, if kept above a certain temperature, may "live" forever.[97] The oldest perpetual stew is said to have lived 300 years, having an unfortunate farewell during World War II.[98] The concept is helpful when it comes to machine learning given two properties of such a stew: a recipe and perpetuity.

First, although not strictly necessary, most of us cook by following a recipe. It is no different for a perpetual stew. The stew starts like any other cooking adventure, by following a set of instructions. Similarly, although you could build a machine-learning model without following

---

95. *See* Everalbum Complaint, *supra* note 47, at 1–3, 7; *see also* Emma Elder, Note, *Wrongful Improvers as a Guiding Principle for Application of the FTC's IP Deletion Requirement*, 97 WASH. L. REV. 1009, 1010 (2022) (discussing the case in detail).

96. *See supra* note 48 and accompanying text.

97. Arthur Prager, *From, a Pot-au-Feu, Many Happy Returns*, N.Y. TIMES C3 (May 6, 1981) (explaining that "[a] pot-au-feu is a kind of combination soup and beef stew. The dish originated not as the creation of a master chef but as the accidental result of peasant indolence. Starting up an ancient wood-burning stove every morning was a tedious job. It was simpler to keep it burning day and night, to heat your cottage as well as cook your meals . . . . Why prepare a new soup every day when all you have to do is keep one simmering and throw in new ingredients occasionally? . . . What had begun as a peasant expedient became an affectation for snobbish gastronomes"); *see also* Maria Trimarchi, *The Stew You Brew (and Eat) for Years: Perpetual Stew*, HOW STUFF WORKS (Feb. 18, 2021), https://recipes.howstuffworks.com/perpetual-stew.htm [https://perma.cc/8DKV-UPWG] ("A savory stew of lobster shells, fish heads and vegetable scraps gently bubbles in the kitchen of Chef David Santos' Portuguese-influenced restaurant, Louro, in New York City's West Village, and, it's been simmering not for hours, but for months.").

98. *See* Adriana Diaz, *Gen Z Flocks to Eat Bizarre, Weeks-Old 'Medieval' Stew: 'A True NY Experience*,*'* N.Y. POST (July 14, 2023, 3:58 PM), https://nypost.com/2023/07/14/gen-z-eats-weeks-old-medieval-stew-a-true-ny-experience [https://perma.cc/5K4J-KCN4] (explaining that the stew can continue cooking for decades if properly maintained).

any instructions, the easiest way is to follow a pre-built recipe. These recipes are like the neural networks discussed above, representing the basic building blocks used in AI. For example, if you have image data, you could use a recipe for a type of neural network known as a convolutional neural network; if you have text data, you might lean toward a type of neural network known as a transformer.

Second, the "thing" produced in machine learning, the model, is a final product made via iterative tweaks and updates, like adding a pinch of salt or just a little more butter. As noted above, the model uses an iterative loop to maximize an objective goal, like accuracy based on ground truth.[99] Tweaks here come in the form of adjusting what works (i.e., performance metrics[100]), what does not work (e.g., overfitting[101] or dropout[102]), and what amounts of ingredients are needed (e.g., how

---

99. Fairness or bias issues are often dealt with in this training phase. For example, in 2015 Google released an image classification AI. Somehow missed in testing, the model labeled African American individuals as "gorillas." Google's solution—the updated model—removed all primate labels from its dataset, effectively pruning the architecture of the model. Alex Hern, *Google's Solution to Accidental Algorithmic Racism: Ban Gorillas*, GUARDIAN (Jan. 12, 2018, 11:04 AM), https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people [https://perma.cc/LDS3-Q2NL]. Along these same lines, it is highly unlikely that this recipe for perpetual stew is the only recipe to exist. While some recipes ae better than others, one can tweak the architecture underlying a model by modifying an activation function or adding a layer to improve the model on some specific task.

100. *See* HOWARD & GUGGER, *supra* note 55, at 169 (discussing the importance of focusing on metrics of performance rather than model loss, because it leads to more accessible improvement strategies). Performance typically refers to how accurate the model is on training data, but in a much more sophisticated way, taking into consideration many of the different ways a model might return a correct or incorrect result. *See* Howard, *supra* note 46 (discussing, at 14:50, popular architectures for machine learning).

101. HOWARD & GUGGER, *supra* note 55, at 30 ("Overfitting is the single most important and challenging issue when training for all machine learning practitioners, and all algorithms . . . it is much harder to make accurate predictions on data the model has never seen before."); *see also* David Lazer & Ryan Kennedy, *What We Can Learn from the Epic Failure of Google Flu Trends*, WIRED (Oct. 1, 2015, 7:00 AM), https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends [https://perma.cc/A79X-3DCW] (discussing real-world overfitting problems—Google Flu Trends was susceptible to non-flu trends).

102. *See* HOWARD & GUGGER, *supra* note 55, at 395 (discussing dropout as a means of generalizing networks so that they do not fit to existing data by randomly removing information from the network, otherwise the model would not work well on unseen inputs).

much data does the network need to make accurate predictions[103]). Once complete, however, the model is alive.

Finally, it is worth noting that just like a perpetual stew is to be kept alive long after its birth, machine learning models act on the world's input long after they are done learning—as long as they are maintained.[104] If a stew is kept above a certain temperature, then the creation is safe to eat, but if it is not kept above that temperature, then it will spoil. Likewise, models live on data, and because data is constantly evolving, so too do models need continual learning. With these concepts in mind, we can turn to see how these nuances play out in practice.

B. The Ever Affair

Ever, owned by Everalbum,[105] was a photo storage and organization application that rose to popularity in 2018 due to its free cross-device photo sharing, via cloud storage, and automatic album creation, via facial recognition.[106] What was not known until 2020, however, was that Everalbum had been collecting user-uploaded photos to feed them to its facial recognition AI.[107] Outside of a few select states with heavy-handed biometric protection laws, like Illinois, users were automatically enrolled in this program and unable to opt out.[108]

In January 2021, the FTC filed a Section 5(a) complaint against Everalbum, and by May, a settlement had been reached. In part, that settlement read:

---

103. *See* Howard, *supra* note 46 (discussing, at 50:45, how one of the most common mistakes beginners to machine learning make is trying out their projects on large, accurate architectures, when slightly-less accurate but faster architectures would work just as well or better for testing, a time when you are learning whether your project will show promise with the existing data you have without needing new data).

104. *See supra* note 95 and accompanying text (discussing the necessity of retraining pipelines given the stagnant nature of data).

105. *See* Everalbum Complaint, *supra* note 47, at 1.

106. *See id.* at 2 ("In February 2017, Everalbum launched its 'Friends' feature, which operates on both the iOS and Android versions of the Ever app. The Friends feature uses face recognition to group users' photos by faces of the people who appear in the photos.").

107. *See id.* ("When Everalbum launched the Friends feature, . . . [it] did not provide users of the Ever mobile app an option to turn off or disable the feature.").

108. *Id.* (noting that users in Texas, Illinois, Washington, and the European Union had to consent to the use of facial recognition technology).

> (A) [D]elete or destroy all photos and videos that Respondent collected from Users who requested deactivation of their Ever accounts . . .
>
> (B) [D]elete or destroy all Face Embeddings [i.e., data, such as a numeric vector, derived in whole or in part from an image of an individual's face] derived from Biometric Information Respondent collected from Users who have not, by that date, provided express affirmative consent for the creation of the Face Embeddings, and provide a written statement to the Commission, sworn under penalty of perjury, confirming that all such information has been deleted or destroyed;
>
> (C) [D]elete or destroy any Affected Work Product [i.e., any models or algorithms developed in whole or in part using Biometric Information Respondent collected from Users of the "Ever" mobile application].[109]

What is most interesting about this settlement is not the intuitive order to destroy all media that was obtained by Everalbum without proper consent. Although this is a commonsense, if controversial,[110] remedy—divesting someone of the "thing" improperly obtained—the remedy would have been wholly inadequate on its own, as discussed above. We learned that most models go through a training phase to learn parameters like weights, but once that phase is over, the training data is no longer needed.[111] The model can make predictions (i.e., recognize faces) without the training data because the model has learned how to arrive at the correct answer. If the FTC did not go one step further and also require the destruction of the "face embeddings" and the model itself, this remedy would not have prohibited Everalbum from continuing to offer the offending facial recognition AI in their products. The FTC did indeed go one step further.[112]

---

109. *See* Decision and Order, Everalbum, Inc., FTC Docket No. C-4743, at 4–5 (May 6, 2021), https://www.ftc.gov/system/files/documents/cases/1923172_-_everalbum_decision_final.pdf [https://perma.cc/H6KX-R6PX].

110. *See, e.g.*, Katherine C. Skilling, Note, *Coverage for Ill-Gotten Gains? Discussing the (Un)Insurability of Restitution and Disgorgement*, 72 WASH. & LEE L. REV. 1077, 1135 (2015) (arguing for disgorgement). *But see* James Tyler Kirk, *Deranged Disgorgement*, 8 J. BUS. ENTREPRENEURSHIP & L. 131, 133 (2014) (questioning disgorgement).

111. *See supra* Section I.B (discussing training, training data, and then moving past that (discarding the training data)).

112. *See* discussion accompanying *supra* note 50 (discussing the FTC's imposition of a remedy so that Everalbum did not profit from the data they should not have obtained).

The model disgorgement provisions make it clear that the FTC intended to prohibit Everalbum from profiting from their ill-gotten gains.[113] Despite the nearly unanimous praise for this provision,[114] however, a nuanced understanding of machine learning lifts the cover on how this may have been more of a compromise than it initially appears.

Considering Everalbum's model as a perpetual stew, the FTC made two demands: (1) throw away the improperly obtained ingredients used to make the stew; and (2) throw away the stew itself.[115] The FTC did not prohibit Everalbum from making another stew. The FTC did not demand that Everalbum throw away their recipe. The FTC merely said, 'throw away any leftover turmeric you have' and 'get rid of this stew.'[116]

This is a step in the right direction, but not the complete eradication of any profitable knowledge gained from the devious Ever app. For example, if Everalbum had enough data from those individuals who explicitly consented to retrain[117] their model, then a new stew could easily be created, allowing Everalbum to come out of this settlement with little loss. Likewise, the performance metrics[118] and architectural, model-specific details learned from the prior model—the recipe—would have been permissibly used as baseline knowledge in a re-training process, providing a sense of how much loss the improper training data's omission caused and speeding up the process of revised model generation.[119] Finally, Everalbum could have attempted to

113. *See* Slaughter et al., *supra* note 48, at 54–57 (discussing options, and limitations, for remedies).

114. *See supra* note 48 and accompanying text.

115. *See supra* note 109 and accompanying text.

116. *Id.*

117. *See supra* note 93 and accompanying text. Training is part and parcel to model creation. Going back to the drawing board using the same procedure with updated data would, in many ways, be trivial. *Id.*

118. *See* Webb, *supra* note 93; *see also* Sasu Mäkinen, Henrik Skogström, Eero Laaksonen & Tommi Mikkonen, *Who Needs MLOps: What Data Scientists Seek to Accomplish and How Can MLOps Help?*, 2021 IEEE/ACM 1ST WORKSHOP ON AI ENG'G 109, 109 (2021) (discussing the necessity of re-training pipelines).

119. A great deal of effort goes into fine tuning a model, and knowing how good a model can be cuts down on a large majority of that work. *See* HOWARD & GUGGER, *supra* note 55, at 75–78; *see also* Binghui Wang & Neil Zhenqiang Gong, *Stealing Hyperparameters in Machine Learning*, 39TH IEEE SYMP. SEC. & PRIV. (2019) (discussing how to protect the effort that goes into fine tuning the model: "Hyperparameters are

"sanitize" the prior data and create a synthetic dataset that maintained the privacy of the original individuals[120] or even attempted to train a new model based directly on outputs from the old model, creating a shadow model that does not use the face embeddings from the prior model directly, but only the prior model's outputs.[121] Although these two workarounds very likely fly in the face of Section C and might receive pushback from the FTC, it is a less clear-cut case and may have been an arguable position for compliance within the four corners of the settlement order.

In short, if the FTC wanted to strip Everalbum of any benefit gained from the improperly created model, further restrictions were needed. Directly prohibiting Everalbum from using this recipe or a substantially similar recipe or prohibiting Everalbum from making any other facial recognition AI would have been necessary. These are more extreme measures, but possibly reasonable given that: (1) a great deal of effort goes into recipe building (i.e., architecture specifics), information itself which could then have been sold; and (2) going to the data store to get more turmeric (i.e., consenting individuals) may have been easy with today's *failed* notice and choice regime, leaving a visible, easy-to-follow path between Everalbum and a "legal" version of their AI.

CONCLUSION

AI moves fast. Yesterday's recurrent neural networks (Apple's Siri[122]) are today's transformers (ChatGPT[123]). And yet, the details matter. Technological details, in many cases, shape and are dispositive to legal outcomes. This short Essay hopes to lay a foundation for

---

critical for [machine learning] algorithms . . . hyperparameters are often learnt through a computationally expensive cross-validation process, which may be implemented by proprietary algorithms that could vary across learners. Therefore, hyperparameters may be deemed confidential").

120. *See* Bellovin et al., *supra* note 42, at 48–49 (discussing how adding differential privacy to synthetic data could provide privacy guarantees).

121. Nathan Reitinger, Bruce Wen, Michelle L. Mazurek & Blase Ur, *What Does It Mean to Be Creepy? Responses to Visualizations of Personal Browsing Activity, Online Activity, and Targeted Ads*, PROC. ON PRIV. ENHANCING TECH. [redacted] 1, 3 (forthcoming 2024) (on file with Author) (utilizing a "shadow model" to train a secondary machine learning model on output from a primary machine learning model).

122. Tim Capes, Paul Coles, Alistair Conkie, Ladan Goliour, Abie Hadjitarkhani, Qiong Hu et al., *Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System*, INTERSPEECH 4011, 4014 (2017) (discussing how Apple's Siri used a particular type of neural network, a Recurrent Neural Network).

123. *See* Wolf et al., *supra* note 42.

understanding AI (i.e., machine learning) at its most basic—part machine and part learning. The machine may be thought of as a neural network, and the learning may be thought of as gradient descent. Taken together, AI is a mathematical function fit to data. Equipped with this knowledge, policymakers, law enforcement, and the judiciary will be better prepared to understand and analyze problems arising at the AI-law intersection.