

The Rise of Synthetic Judges: If We Dehumanize the Judiciary, Whose Hand Will Hold the Gavel?

Marie-Claire Aarts[†]

The process of automating the judiciary is well underway in much of the developed world. Yet, if we dehumanize our judiciary, we are essentially placing the gavel in the hands of proprietary algorithms. We will increasingly be judged by what is often referred to as a black box.

I. INTRODUCTION

Proprietary algorithm is a term of the digital age.¹ It integrates intellectual property concepts and trade secrets with an invisible algorithmic hand that increasingly controls us.² The grant of exclusive licenses to private firms is one of the bedrocks of capitalism, while the term trade secrets conjures up notions of Coca Cola’s syrup recipe—one of the longest and most heavily guarded trade secrets in the world.³ All in all, *proprietary algorithm* is not a controversial term. Yet, an algorithm is an entirely different beast than a soft drink recipe.

Algorithms solve specific problems that translate into actions. Certain artificial intelligence (“AI”) algorithms are autonomous. They are able to make their own decisions and can even write their own code. Take, for

[†] Marie-Claire Aarts is a Dutch business lawyer who won several innovation awards for connecting the dots between technology and law, one of which is the prestigious Accenture Innovation Award. She founded NEXT INSTITUTE FOR TECHNOLOGY & LAW, which offers online courses and explainers that provide interdisciplinary insight into the technologies shaping our future and that place the digital age in legal context. Technology has become one of the pillars of power, but it remains mostly unchecked. To legislate the digital sphere, we need to understand the underlying technologies. NEXT INSTITUTE provides guidance through this fascinating terrain at www.nextinstitute.org.

1. Stacy Collett, *How to Protect Algorithms as Intellectual Property*, CSO (July 13, 2020, 3:00 AM), <https://www.csoonline.com/article/3565195/how-to-protect-algorithms-as-intellectual-property.html> [<https://perma.cc/NP8Q-XL67>].

2. *See id.*

3. *See Coca-Cola’s Formula Is at the World of Coca-Cola*, COCA-COLA CO., <https://www.coca-colacompany.com/company/history/coca-cola-formula-is-at-the-world-of-coca-cola> [<https://perma.cc/6UTJ-C4SD>] (last visited Apr. 14, 2021).

example, Google’s AutoML-Zero.⁴ As its name suggests, it can create algorithms with zero human intervention.⁵ We have yet to determine whether this autonomy amounts to something like free will, and whether such algorithms should have duties and some form of protection for artificial rights, kindred to human rights. That idea is not as farfetched as it may initially seem. Companies are not human either, but they are recognized as separate legal entities.⁶ They can enter into contracts, can sue and be sued, and can even own property.⁷

However, these legal constructs are all powerless without human intervention. When we call a company negligent or criminal, we are referring to the people behind it. Granted, it can sometimes be complicated to link specific acts to specific persons. But at some point, there was always someone, somewhere, pulling the strings—be it the management, the banks, certain shareholders, or even a government. If all humans were terminated tomorrow, all activities would eventually cease, apart from some automated tasks that were already put in motion.

Artificial intelligence has another dimension. The next generation of AI models will be progressively independent from humans. They may have been the brainchild of humans, but many models are being designed to be autonomous.⁸ At a point in the not-too-distant future, some models may reach *artificial general intelligence* (“AGI”)—a human-like state.⁹

We have learned the hard way that there can be real-world consequences of artificial acts in the digital domain. We have experienced how pieces of code can violate privacy, and how social media algorithms can amplify undemocratic sentiments and whip them into a frenzy.¹⁰ This is not even *strong AI*, as far as we know. It is mostly *narrow AI*, which can perform only one or a few narrowly defined tasks. In these models, the developer still has a leading role.

4. Sahil Uppal, *AutoML-Zero*, TOWARDS DATA SCI. (July 22, 2020), <https://towardsdatascience.com/automl-zero-b2e065170941> [<https://perma.cc/U24E-MQ95>].

5. *Id.*

6. See generally *Santa Clara Cnty. v. S. Pac. R.R. Co.*, 118 U.S. 394 (1886).

7. Sarah Pruitt, *How the 14th Amendment Made Corporations Into ‘People’*, HIST. (Oct. 15, 2018), <https://www.history.com/news/14th-amendment-corporate-personhood-made-corporations-in-to-people> [<https://perma.cc/M2WB-BTAC>].

8. See Matt Young, *AI-Autonomous Models That Are Transforming Business*, FUNCTIONIZE (June 26, 2018), <https://www.functionize.com/blog/ai-autonomous-models-that-are-transforming-business/> [<https://perma.cc/6ZT8-5NY7>].

9. Cem Dilmegani, *995 Experts Opinion: AGI / Singularity by 2060 [2021 Update]*, AIMULTIPLE (Feb. 2, 2021), <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/> [<https://perma.cc/JG2J-66P6>].

10. Elizabeth Goitein, *How the FBI Violated the Privacy Rights of Tens of Thousands of Americans*, BRENNAN CTR. FOR JUST. (Oct. 22, 2019), <https://www.brennancenter.org/our-work/analysis-opinion/how-fbi-violated-privacy-rights-tens-thousands-americans> [<https://perma.cc/MV8A-TG9S>].

Social media algorithms are not the only algorithms that impact our lives. In the legal sphere, risk assessment algorithms that guide our judiciary are playing an ever-increasing role in determining guilt or innocence and the level of punishment.¹¹ In *State v. Loomis*,¹² a man found himself being sentenced to six years in prison—in part because an algorithmic risk assessment tool had identified him as an “individual who is at high risk to the community.”¹³ That particular legal tool and tools like it are used in most American jurisdictions to assess the likelihood of recidivism. These are proprietary software, and the source codes underlying proprietary software are not publicly available.¹⁴ In other words, nobody has a clue what is in them, or what biases or plain mistakes may be written into the code. Only the company that created such a tool can test and modify it.

The counterpart to proprietary software is open-source software, for which the source code is publicly accessible.¹⁵ Anyone who is tech-savvy can scrutinize the source code of such algorithms for hidden or unconscious assumptions and other flaws. The developer or another entity can control the copyrights, patents, and trademarks of the proprietary software, but its content is open to anyone.

Arguably, sentencing based on proprietary assessment algorithms violates due process, which is exactly what the defendant claimed in *State v. Loomis*.¹⁶ In this case, the defendant argued that using such a risk assessment tool at sentencing violates a defendant’s right to due process, either because the proprietary nature of the software prevents defendants from challenging the scientific validity of that tool or because the assessments take gender and race into account.¹⁷

The specific issues in the case were as follows:

(1) Whether it is a violation of a defendant’s constitutional right to due process for a trial court to rely on the risk assessment results provided by a proprietary risk assessment instrument such as the Correctional Offender Management Profiling for Alternative Sanctions at sentencing because the proprietary nature . . . prevents a defendant from challenging the accuracy and scientific validity of the risk assessment; and

11. See *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), cert. denied sub nom. *Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017).

12. *Id.*

13. *Id.* at 755.

14. David Crooke, *Open Source vs. Proprietary Software*, GREENNET, <https://www.greennet.org.uk/support/open-source-vs-proprietary-software> [<https://perma.cc/65TY-LU2E>] (last visited Apr. 14, 2021).

15. *Id.*

16. *Loomis*, 881 N.W.2d 749.

17. *Id.* at 757.

(2) whether it is a violation of a defendant’s constitutional right to due process for a trial court to rely on such risk assessment results at sentencing because . . . assessments take gender and race into account in formulating the risk assessment.¹⁸

The U.S. Supreme Court denied the petition for a writ of certiorari on June 26, 2017.¹⁹ Its denial of certiorari was a missed opportunity to address some of the imminent, most pressing questions regarding the rule of law: Are we allowing secret algorithms to have a decisive role in the sentencing of humans? If so, whose hand will hold the gavel?

The issue is twofold: (1) nobody knows how much developers weigh in on algorithmic output, and (2) the potential danger of increasingly autonomous algorithms wielding such power over humans is unclear. To be able to address either point, we will have to dive into one of the cornerstone technologies of the future: artificial intelligence. It is, therefore, essential to have a basic understanding of what AI and algorithms really are, as well as the difference between supervised and unsupervised machine learning.

II. ARTIFICIAL INTELLIGENCE IN A NUTSHELL

First, artificial intelligence deals with simulation of human intelligence.²⁰ AI has the “ability to change, adapt, and grow based on the input of new data.”²¹ This self-evolving quality is described as *intelligence*.²² Artificial intelligence is a broad term with several subsets. For the purposes of this Article, machine learning (“ML”) and natural language processing (“NLP”) are the most relevant.

ML algorithms enable machines to operate in an automated mode without human intervention—in certain cases, even without the need to be explicitly programmed. ML makes predictions based on the patterns it learns. ML has its own subset: deep learning, the inner-sanctum of artificial intelligence.²³ Deep learning mimics the neural networks of the human

18. *Loomis v. Wisconsin*, SCOTUSBLOG, <https://www.scotusblog.com/case-files/cases/loomis-v-wisconsin/> [<https://perma.cc/7J35-294N>] (last visited Apr. 14, 2021). The specific product name, COMPAS, has been omitted from the quote as it is not relevant for this Article.

19. *Id.*; *Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017).

20. JOHN MCCARTHY, WHAT IS ARTIFICIAL INTELLIGENCE? 3 (2007), <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf> [<https://perma.cc/2WDC-XYDW>].

21. Kaya Ismail, *AI vs. Algorithms: What’s the Difference?*, CMSWIRE (Oct. 26, 2018), <https://www.cmswire.com/information-management/ai-vs-algorithms-whats-the-difference/> [<https://perma.cc/CE6T-UJTF>].

22. *Id.*

23. Brett Grossfeld, *Deep Learning vs Machine Learning: A Simple Way to Understand the Difference*, ZENDESK (Apr. 7, 2021), <https://www.zendesk.com/blog/machine-learning-and-deep-learning/> [<https://perma.cc/7SLY-E28J>].

brain; it learns like a toddler who is discovering its environment.²⁴ These artificial neural networks require abundant computing power and are fueled by massive amounts of data. NLP enables machines to communicate in human language; it interprets language.²⁵ As an illustration, ML can notify you that you are almost out of coffee beans and should restock, whereas NLP allows your voice-activated digital assistant to understand that you want a cup of coffee.

AI systems are powered by algorithms. The term *algorithm* is used easily but is not always fully understood. It is a running joke among developers that they use the word “algorithm” when they do not want to explain what they have done. Major technology companies hide behind their algorithms to avoid accountability for the negative side-effects of their business models. During a 2020 Senate Judiciary hearing on content moderation, Twitter founder Jack Dorsey testified that Twitter’s algorithms are inept at explaining their choices—referring to algorithms as if they are external contractors, rather than the company’s own creations.²⁶ Dorsey stated, “as we look forward, we have more and more of our decisions, of our operations, moving to algorithms, which have a difficult time explaining why they make decisions, bringing transparency around those decisions.”²⁷

Algorithms are actually quite straightforward; they are a series of instructions to be followed, step by step, to achieve a goal or solve a problem.²⁸ An algorithm can be understood by comparing it to a cake recipe.²⁹ To prepare a dish, you would read the instructions and execute each step one by one, in the given sequence. A recipe is in fact a simple, linear algorithm, instructing what to do to get a specific result—in this example, a perfectly baked cake. There are no computers involved in this process, but computers do not have to be involved for something to qualify as an

24. Martin Heller, *What Is Deep Learning? Algorithms that Mimic the Human Brain*, INFOWORLD (May 24, 2019, 3:00 AM), <https://www.infoworld.com/article/3397142/what-is-deep-learning-algorithms-that-mimic-the-human-brain.html> [<https://perma.cc/9KQJ-MQWX>].

25. Kevin Casey, *How to Explain Natural Language Processing (NLP) in Plain English*, ENTERPRISERS PROJECT (Sept. 17, 2019), <https://enterpriseproject.com/article/2019/9/natural-language-processing-nlp-explained-plain-english> [<https://perma.cc/XFL9-GRK3>].

26. *Facebook and Twitter CEOs Testify on Regulating Social Media Content*, C-SPAN (Nov. 17, 2020), <https://www.c-span.org/video/?478048-1/facebook-twitter-ceos-testify-regulating-social-media-content> [<https://perma.cc/FE79-F7X3>].

27. *Id.*

28. *What Is an Algorithm? An ‘in a Nutshell’ Explanation*, THINKAUTOMATION, <https://www.thinkautomation.com/eli5/what-is-an-algorithm-an-in-a-nutshell-explanation> [<https://perma.cc/Q5AL-CPW6>] (last visited Apr. 14, 2021).

29. *See id.*

algorithm. In fact, algorithms existed long before computers were invented.³⁰ They are an ancient concept.³¹

In the context of artificial intelligence, algorithms give instructions to machines rather than humans, and the instructions are more complicated. The technology is developing fast, and the more complex forms of AI are multi-dimensional in nature. Herein lies the strength of the technology. While the number of dimensions with which a human can deal is limited, intelligent machines can process incommensurable dimensions simultaneously.

III. HUMAN VERSUS MACHINE

AI systems get progressively better at tasks when fed more data—the more data, the better.³² In technical terms, it is called *training the algorithm*.³³ This data dependency explains why countries with few privacy protections have advantages in the field of AI. These countries have more datasets with which to work.³⁴

The training of AI algorithms is more thorough than human training, or schooling. Take the impressive Generative Pre-Trained Transformer-3 (“GPT-3”), a text-generating language model developed by OpenAI, which is aiming for artificial general intelligence.³⁵ GPT-3 produces humanlike text.³⁶ It mastered the English language by analyzing thousands of digital books and other texts, including all of Wikipedia.³⁷

At the time of its introduction in May 2020, GPT-3 had scrutinized nearly 300 billion tokens—words in e-books, or posted to blogs, social media, and the rest of the Internet.³⁸ Most people do not read that much material in their lifetimes. Twenty-year-old Americans, with English as

30. Marek Kowalkiewicz, *How Did We Get Here? The Story of Algorithms*, TOWARDS DATA SCI. (Oct. 10, 2019), <https://towardsdatascience.com/how-did-we-get-here-the-story-of-algorithms-9ee186ba2a07> [<https://perma.cc/FGK7-JM63>].

31. *Id.*

32. *But see* H. James Wilson, Paul R. Daugherty & Chase Davenport, *The Future of AI Will Be About Less Data, Not More*, HARV. BUS. REV. (Jan. 14, 2019), <https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more> [<https://perma.cc/4YFK-BNBL>].

33. *Id.*

34. Graham Webster & Scarlet Kim, *The Data Arms Race Is No Excuse for Abandoning Privacy*, FOREIGN POL’Y (Aug. 14, 2018, 11:43 AM), <https://foreignpolicy.com/2018/08/14/the-data-arms-race-is-no-excuse-for-abandoning-privacy> [<https://perma.cc/5AQQ-69TM>].

35. Tom B. Brown, Benjamin Mann, Nick Ryder & Melanie Subbiah, *Language Models Are Few-Shot Learners*, ARXIV (July 22, 2020), <https://arxiv.org/pdf/2005.14165.pdf> [<https://perma.cc/BY5R-3BZ7>]; *see also* Ram Sagar, *OpenAI Releases GPT-3, the Largest Model So Far*, ANALYTICS INDIA MAG. (Mar. 6, 2020), <https://analyticsindiamag.com/open-ai-gpt-3-language-model> [<https://perma.cc/Z56W-SLH9>].

36. Brown et al., *supra* note 35, at 5, 25–26.

37. *Id.* at 8–9.

38. *Id.*

their native language, have an average vocabulary of about 42,000 lemmas.³⁹ The numbers range from 27,000 lemmas for the bottom 5 percent to 52,000 for the highest 5 percent.⁴⁰ Every couple of days, a new lemma is added, so by the time Americans reach the age of sixty, they will know about 48,000 lemmas.⁴¹

GPT-3 works as follows. Given the previous words in a text, it predicts the next word.⁴² If it is fed enough legal documents, as an example, the model can read and write legal documents—perhaps as well as a professional, and maybe even better.⁴³ Moreover, GPT-3 can understand and explain the documents.⁴⁴ UK investor Michael Tefula put this to the sandbox test, training GPT-3 to translate legal texts into plain English. On July 21, 2020, he sent out the following tweet, accompanied by a video, about training this neural network: “Just taught GPT-3 how to turn legalese into simple plain English. All I gave it were 2 examples. Might build a term sheet and investment document interpreter out of this.”⁴⁵ This neural network can also be trained to do the opposite: translate laymen’s wording

39. Marc Brysbaert, Michaël Stevens, Paweł Mandera & Emmanuel Keuleers, *How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant’s Age*, 7 FRONTIERS PSYCH. 1, 8–10 (2016). Lemmas are dictionary entries, or uninflected words from which the variations are formed. *Id.* at 2.

40. *Id.* at 1.

41. *Id.* at 8.

42. Sagar, *supra* note 35.

43. *Id.*

44. *Id.*

45. Michael Tefula (@michaeltefula), TWITTER (July 21, 2020, 4:24 AM), <https://twitter.com/michaeltefula/status/1285505897108832257?lang=en> [<https://perma.cc/WF3P-3AHE>]. For those who missed it, here are the examples Tefula showed in an attached video:

Legalese to English Translator:

The Deferred Shares (if any) shall not entitle the holders of them to receive notice of, to attend, to speak or to vote at any general meeting of the Company nor to receive or vote on, or otherwise constitute an eligible member for the purposes of, proposed written resolutions of the Company.

[Translation:] The shares will not give the holder the right to vote at general meetings or to be a member of the company.

Michael Tefula (@michaeltefula), TWITTER (July 26, 2020, 11:34 AM), <https://twitter.com/michaeltefula/status/1285505897108832257?lang=en> [<https://perma.cc/Y9AA-WUS8>].

Legalese to English Translator:

If no Transfer Price can be agreed between the Seller and the Board in accordance with provisions of Articles 14.10, 16.2 or otherwise then, on the date of failing agreement, the Board shall either:

(a) appoint an expert valuer in accordance with Article 17.2 (the “Expert Valuer”) to certify the Fair Value of the Sale Shares; or

(b) (if the Fair Value has been certified by an Expert Valuer within the preceding 12 weeks) specify that the Fair Value of the Sale Shares will be calculated by dividing any Fair Value so certified by the number of Sale Shares to which it related and multiplying such Fair Value by the number of Sale Shares the subject of the Transfer Notice.

[Translation:] If the parties cannot agree on a price, the Board will either appoint an expert to decide the price or use a formula to calculate the price.

Id.

into legal arguments. In theory, such a tool could substantially increase accessibility to legal services.

Beyond the legal implications, the potential impact of the program is enormous. It has impressive creative capabilities. It can compose songs, write short stories, or even finish a poem when given only a few lines—and a synthetic version may be indistinguishable in style from the real thing.⁴⁶ It is even possible to converse with a digital version of historical figures, such as Isaac Newton, Alfred Hitchcock, or Marie Curie, and ask them to shed light on developments in their field and the state of the world.⁴⁷

The abstract name may make GPT-3 appear to be just another technical term. To make it more tangible, consider replacing GPT-3 with a regular name, say “the Honorable Judge Stephen W. Dawkins.” Then, AI obtains an entirely different dimension. Assume that this imaginary judge, has an incredible track record. His knowledge of precedents is unequaled, because he is able to tap into a substantially larger knowledge base than his human colleagues. He has the capacity to always provide balanced sentencing, without any partisanship,⁴⁸ and he works faster. The Honorable Judge Stephen W. Dawkins never takes a day off, and even works through the night.

In the months since GPT-3 sent shockwaves through the legal sector, testing proved that it is much too early to panic about imminent job loss. Apart from biases, the AI algorithm still has many limitations to overcome. However, it is good to realize that a tremendous leap was made. It was not anticipated that artificial intelligence would reach anywhere near this level of sophistication for many years. Regardless, some legal professionals may believe that automation is something that will only affect manual jobs. Perhaps some still believe that their ingenuity and intellect are unique and they will be spared replacement by a machine—they should think again. In many ways, cognitive tasks may be even more suitable for execution by machines. The law is language-based and built on consistency. While many manual jobs need robotics to replace them, the cumbersome process

46. James Vincent, *OpenAI's Latest Breakthrough Is Astonishingly Powerful, but Still Fighting Its Flaws*, VERGE (July 30, 2020, 10:01 AM), <https://www.theverge.com/21346343/gpt-3-explainer-open-ai-examples-errors-agi-potential> [<https://perma.cc/G3G2-WJFX>].

47. Andrew Mayne, AI|WRITER, <https://aiwriter.app/> [<https://perma.cc/M3U9-G9AW>] (last visited Apr. 14, 2021). For an example of an exchange with Isaac Newton, see <https://aiwriter.app/sample/fac0734f03f5c2e388> [<https://perma.cc/6WAY-U2AM>] (last visited Apr. 14, 2021).

48. The notion of balanced sentencing requires a side note. We already know that the algorithmic judge Stephen W. Dawkins is likely to perpetuate and amplify racial and gender biases present in the data used to train him—similar biases that may have led to the wrongful arrest of the person in front of him, as the police are using similar tools. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<https://perma.cc/564V-R4FA>].

of building a physical robot is not even necessary in order to synthesize many of the tasks legal professionals perform.

IV. MORAL CODE

A major component that AI is still missing is a moral code. Algorithms and computers are not intrinsically good at making ethical decisions—although one could argue that humans are not particularly proficient at it either. A deep chasm exists regarding this issue. Many scholars claim machines will never develop a moral compass.⁴⁹ Others propose we could teach machines a moral code with case-based machine learning.⁵⁰ Those with this view claim that this is the same way humans learn the difference between right and wrong.⁵¹ This is not only a philosophical issue, but also an existential one. In no field is the issue more relevant than in the legal profession—what is the judiciary without the innate ability to distinguish right from wrong?

The last word has not been said on the ethics of automation of the judiciary. However, with the widespread introduction of AI-based legal tools, the crucial debate on the ethics of automation is rendered obsolete. Commercial companies are making the decision for us by introducing convenient tools we can no longer do without. Who still travels without a navigation system? Who still practices mental arithmetic, even for the simplest calculations? Accordingly, it is wrong to presume that, in the future, we will still be able to negotiate the complex legal system without algorithmic help.

We are accustomed to outsourcing cognitive tasks to the digital domain. Yet, each time we use a virtual assistant or the Internet of Things (“IoT”),⁵² we delegate part of our brain function. Loss of our cognitive dominance will likely occur gradually, one click at a time.⁵³

49. Brian Gallagher, *The Case Against A.I. Controlling Our Moral Compass*, ETHICAL SYS. (June 25, 2019), <https://www.ethicalsystems.org/the-case-against-a-i-controlling-our-moral-compass/> [<https://perma.cc/J5TP-3WNK>]; Kris Hammond, *Ethics and Artificial Intelligence: The Moral Compass of a Machine*, VOX: RECODE (Apr. 13, 2016, 2:22 PM), <https://www.vox.com/2016/4/13/11644890/ethics-and-artificial-intelligence-the-moral-compass-of-a-machine> [<https://perma.cc/5CBR-K6CN>]; Maja Pantic, *Introduction to Machine Learning & Case-Based Reasoning*, IMPERIAL COLL. LONDON, <https://ibug.doc.ic.ac.uk/media/uploads/documents/courses/syllabus-CBR.pdf> [<https://perma.cc/L545-7VUN>] (last visited Apr. 14, 2021).

50. See sources cited *supra* note 49.

51. See sources cited *supra* note 49.

52. Subodh Kumar, Pradeep Kumar & Tara Singh, *Emerging Technological Trends and Human Cognition: A Review*, 6 J. DISABILITY MGMT. & REHAB. 23, 24 (2020), <https://zenodo.org/record/4165065#.YDCsfWhKguU> [<https://perma.cc/G2NE-J8V2>].

53. *Id.* at 23–29.

Relying on algorithms can be risky. The tale of a virtual girl named Tay proves how things can go terribly wrong.⁵⁴ Microsoft created social media accounts for a neural network that would interact with eighteen- to twenty-four-year-olds.⁵⁵ Posing as a young girl, the neural network actively joined discussions on Twitter and other social media platforms.⁵⁶ In the course of just one day, Tay turned racist and sexist—inundating other participants with abuse.⁵⁷ Filters of vulgar expression either failed or were ignored.⁵⁸ Basically, social media corrupted a neutral AI chatbot within 24 hours. The account had to be closed immediately, and the company sent out a formal apology.⁵⁹

Tay was so-called *unsupervised* AI. In unsupervised learning (“UL”), a model is given freedom to select learning targets on its own preference.⁶⁰ Indeed, within deep learning, unsupervised learning is the holy grail. It no longer needs the step-by-step instructions but learns and adapts to the data on its own. That can lead to authentic and inspiring insights, but it can also wreak havoc. Unsupervised learning is often compared to a classroom without a teacher: students may study diligently, but mayhem is just around the corner. In the case of Tay, she demonstrated racism and decided to further explore that avenue until she became an ardent Hitler fan in the span of one day.⁶¹

Supervised learning works differently: the teacher still reigns in the classroom and controls the datasets used to train the algorithm. The developer is in charge of the curriculum, so to speak. The vast majority of AI is still considered narrow and is supervised,⁶² but that does not mean it will remain that way. Future legal tools will follow the general trend and become increasingly intelligent and increasingly unsupervised. What is more, if the proprietary algorithm of a legal tool is hidden, society has no way of

54. Peter Lee, *Learning from Tay's Introduction*, MICROSOFT (Mar. 25, 2016), <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction> [<https://perma.cc/3PLL-3T5Q>].

55. *Id.*

56. *See id.*

57. *Id.*

58. *Id.*

59. *Id.*

60. Thomas Wood, *What Is Unsupervised Learning?*, DEEPAI, <https://deepai.org/machine-learning-glossary-and-terms/unsupervised-learning> [<https://perma.cc/92W6-VE2Q>] (last visited Apr. 14, 2021).

61. See Helena Horton, *Microsoft Deletes 'Teen Girl' AI After It Became a Hitler-Loving Sex Robot Within 24 Hours*, TELEGRAPH (Mar. 24, 2016, 3:37 PM), <https://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/> [<https://perma.cc/8MSN-TJD7>].

62. Tannya D. Jajal, *Distinguishing Between Narrow AI, General AI and Super AI*, MEDIUM (May 21, 2018), <https://medium.com/mapping-out-2050/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22> [<https://perma.cc/S4E4-KNXH>].

knowing what type of AI underlies the outcome. Outsiders cannot test their performance; after all, only the developer has access to the source codes.

V. INTELLIGENCE EXPLOSION

One trait that makes humans special is our general intelligence. We are creative multi-taskers, and indeed, our ingenuity has made our species dominant. Narrow AI may already beat humans at demarcated tasks, but humans still win at improvisation in fields for which evolution has not prepared us. People think and live outside the box.

Yet, all over the world, researchers are working towards computers with an artificial general intelligence. This includes research into whole-brain emulation, which is more vividly described as mind uploading or brain-computer interfaces—for instance, by human enhancing brain implants, such as those that San Francisco-based Neuralink is now testing.⁶³ Such machines would have the ability to learn and understand any intellectual task as well as a human being can. These machines would be good at a wide variety of tasks, able to improvise and strategize with the skill to survive without human intervention.

It is still early in the AI evolution. To place it all into perspective, we are currently in the first phase of narrow AI, edging towards general artificial intelligence. Examples of narrow AI are self-driving cars, facial recognition tools that tag pictures, and virtual assistants like Alexa, Google Assistant, and Siri. Impressive as these tools may be, they are still forms of weak AI. Even the known abilities of GPT-3, which push the boundaries of narrow AI, are still thought to be just a fraction of what is to come.⁶⁴

However, algorithms are not stuck in a linear evolutionary path like humans. They outpace us exponentially. Looking forward to a still hypothetical point in time, artificial general intelligence is expected to lead to an intelligence explosion, resulting in machine superintelligence.⁶⁵ By that point, the positive feedback loops of the machines will have made them vastly more intelligent than the smartest human being on Earth. Somewhere down the road, developments are expected to become uncontrollable and irreversible. Whether machines will really surpass humans is still debated,

63. *Understanding the Brain*, NEURALINK, <https://neuralink.com/science/> [https://perma.cc/5K2Q-ZFQE] (last visited Apr. 14, 2021).

64. See, e.g., *Superintelligence*, NEXT INST. FOR TECH. & L., <https://nextinstitute.org/explainers/superintelligence/> [https://perma.cc/RC5G-HSC4] (last visited Apr. 14, 2021).

65. *Id.*

but this point of no return is referred to as technological singularity.⁶⁶ Usually, this concept is simply referred to as *the singularity*.

These are all still hypotheticals, because there are so many unknowns, and not all scientists agree they will ever transpire. However, no one contradicts the argument that superintelligence would be the most disruptive force humankind has ever released. A growing number of experts—including Elon Musk, noted futurist Ray Kurzweil, and the late Stephen Hawking—have expressed their expectations that artificial systems will most definitely eclipse human intellect and will pose an existential risk to humanity.⁶⁷

These individuals, as well as other experts, propagate strict legislation in this field.⁶⁸ In 2017, a group of technology founders warned about the dangers of unbridled AI proliferation in an open letter to the United Nations.⁶⁹ “Once this Pandora’s box is opened, it will be hard to close.”⁷⁰ The move was coordinated by the Future of Life Institute (“FLI”), which also outlined a set of principles deemed necessary to avoid an uncontrollable AI.⁷¹ These so-called Asilomar AI Principles stress the need for both updated legal systems and judicial transparency: “Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.”⁷²

At the National Governors Association in 2017, Musk stated, “I have exposure to the very most cutting edge AI, and I think people should be really concerned about it.”⁷³ Musk further stated, “AI is a rare case where I think we need to be proactive in regulation, instead of reactive. Because I think by the time we are reactive in AI regulation it’s too late.”⁷⁴ Although the impressive line-up of visionaries and their doomsday scenarios attracted a lot of attention at the time, the initiative now seems to be mostly a platform

66. Vernor Vinge, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, SAN DIEGO ST. UNIV. (1993), <https://edoras.sdsu.edu/~vinge/misc/singularity.html> [<https://perma.cc/34HB-MFUS>].

67. Kim McLendon, *Elon Musk, Michio Kaku, Ray Kurzweil, and Stephen Hawking: Is the Future Exciting or Terrifying?*, STANFORD L. (Nov. 6, 2017), <https://law.stanford.edu/press/elon-musk-michio-kaku-ray-kurzweil-stephen-hawking-future-exciting-terrifying/> [<https://perma.cc/AYE5-PBP3>]; see also Matt Peckham, *What 7 of the World’s Smartest People Think About Artificial Intelligence*, TIME (May 5, 2016, 6:00 AM), <https://time.com/4278790/smart-people-ai/> [<https://perma.cc/5XLN-MHE2>].

68. See sources cited *supra* note 67.

69. Grouping of Artificial Intelligence and Robotics Experts, *An Open Letter to the United Nations Convention on Certain Conventional Weapons*, FUTURE OF LIFE INST. (2017), <https://futureoflife.org/autonomous-weapons-open-letter-2017> [<https://perma.cc/7VP9-5J7T>].

70. *Id.*

71. *Asilomar AI Principles*, FUTURE OF LIFE INST., <https://futureoflife.org/ai-principles/> [<https://perma.cc/X2AA-ENJE>] (last visited Apr. 14, 2021).

72. *Id.*

73. *Elon Musk at National Governors Association 2017 Summer Meeting*, C-SPAN (July 15, 2017), <https://www.c-span.org/video/?431119-6/elon-musk-addresses-nga> [<https://perma.cc/5YK8-M365>].

74. *Id.*

to exchange ideas, rather than the powerful organization that would be needed to achieve such a regulatory framework. Apart from working groups and a long string of research initiatives and whitepapers, their calls for imminent action are largely being ignored.

Machines with human-level capabilities were not expected before 2040 to 2050,⁷⁵ but Musk recently pushed his expectation date to 2025.⁷⁶ In reality, it is hard to know how advanced some of the current AI algorithms really are, because most projects in this space are clouded in secrecy. We do know, however, that big tech companies and governments are pouring billions of dollars into the more advanced types of AI⁷⁷ and that the sector is still held back by limited computational power.⁷⁸ From an economic standpoint, the secrecy is understandable. Artificial intelligence is projected to add as much as \$15.7 trillion to the world economy by 2030.⁷⁹ Geopolitically, there is also significant strategic value, both in being the frontrunner and in flying under the radar. As Vladimir Putin put it, “Which-ever country leads the way in AI research will come to dominate global affairs.”⁸⁰ Even the Russian President wants stricter regulations.⁸¹ China remains conspicuously quiet on the subject. Then again, China is determined to win this battle, as outlined in its controversial 2017 AI Development Plan.⁸²

The backstory of GPT-3 illustrates AI’s vulnerability to profit motives. GPT-3’s developer, OpenAI, was originally set up as a non-profit by some

75. Vincent C. Müller & Nick Bostrom, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, in *FUNDAMENTAL ISSUES OF ARTIFICIAL INTELLIGENCE* 553 (Vincent C. Müller ed., 2016), https://www.researchgate.net/publication/280838978_Future_Progress_in_Artificial_Intelligence_A_Survey_of_Expert_Opinion [https://perma.cc/X8W3-EC6D]. A shorter version of this paper appeared as Vincent C. Müller & Nick Bostrom, *Future Progress in Artificial Intelligence: A Poll Among Experts*, 1 *AI MATTERS* 9 (2014), https://www.researchgate.net/publication/278023965_Future_progress_in_artificial_intelligence_A_poll_among_experts [https://perma.cc/RH9P-SCR9].

76. Maureen Dowd, *Elon Musk, Blasting Off in Domestic Bliss*, *N.Y. TIMES* (July 27, 2020), <https://www.nytimes.com/2020/07/25/style/elon-musk-maureen-dowd.html> [https://perma.cc/29YK-CFL8].

77. Maggie Miller, *Lawmakers Introduce Bill to Invest \$100 Billion in Science, Tech Research*, *HILL* (May 27, 2020, 1:25 PM), <https://thehill.com/policy/cybersecurity/499754-lawmakers-introduce-bill-to-invest-100-billion-in-science-tech-research> [https://perma.cc/SZQ6-MDA9].

78. Karen Hao, *The Computing Power Needed to Train AI Is Now Rising Seven Times Faster Than Ever Before*, *MIT TECH. REV.* (Nov. 11, 2019), <https://www.technologyreview.com/2019/11/11/132004/the-computing-power-needed-to-train-ai-is-now-rising-seven-times-faster-than-ever-before> [https://perma.cc/TT2E-MQ3T].

79. ANAND S. RAO & GERARD VERWEIJ, PWC, *SIZING THE PRIZE: WHAT’S THE REAL VALUE OF AI FOR YOUR BUSINESS AND HOW CAN YOU CAPITALISE?* (2017), <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf> [https://perma.cc/3XJD-N5ZH].

80. James Vincent, *Putin Says the Nation that Leads in AI ‘Will Be the Ruler of the World’*, *VERGE* (Sept. 4, 2017, 4:52 AM), <https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world> [https://perma.cc/P6V8-LJSV].

81. United Nations, *Russia – President Addresses General Debate, 75th Session*, *YOUTUBE* (Sept. 22, 2020), <https://youtu.be/Kt2QdWKzfF8>.

82. Ministry of Sci. & Tech., P.R. China, *Next Generation Artificial Intelligence Development Issued by State Council*, *CHINA SCI. & TECH. NEWSL.* (Sept. 15, 2017).

of the Silicon Valley luminaries that were also involved in the Future of Life Institute initiative.⁸³ OpenAI aimed to bring AI to the masses. The idea was that such a disruptive, potentially dangerous technology should not be in the hands of just a few all-powerful technology companies or governments.⁸⁴ In other words, this powerful technology should be democratized. Its Charter says OpenAI's creations should *benefit all humankind*.⁸⁵ We heard something similar in the early 2000s, when Google's original motto was "*Don't be evil*," which was included in the prospectus of Google's 2004 initial public offering.⁸⁶ The phrase was removed from Google's internal Code of Conduct in 2018.⁸⁷

OpenAI's lofty ideals were later abandoned in favor of a profit model. There is still a non-profit with oversight power, but in 2019, Microsoft became its preferred partner for commercializing its AI technologies.⁸⁸ The technology giant injected \$1 billion in the start-up in exchange for some level of exclusivity.⁸⁹ Early access to the public OpenAI API platform is strictly controlled. Ironically, the GPT-3 source code is now a well-guarded secret.⁹⁰

VI. THE BLACK BOX

Focusing on the black box aspect of AI-based legal tools, both technological and legal aspects exist. In the technological context, artificial intelligence arrives at conclusions via a different route than its human counterpart. Often the outcome cannot be traced, so it cannot be validated. Particularly, the output of unsupervised AI would be immensely hard to assess, even with full access to the source codes. It is not a controlled environment, and these algorithms are not explicitly programmed every step of the way. Hence, the black box comparison. Data goes in, something comes out—what happens in between is opaque.

83. *About OpenAI*, OPENAI, <https://openai.com/about/> [<https://perma.cc/ZUF8-KQ8H>] (last visited Apr. 14, 2021).

84. *OpenAI Charter*, OPENAI (Apr. 9, 2018), <https://openai.com/charter/> [<https://perma.cc/8P3H-NPME>].

85. *Id.*

86. Google, Inc., Amendment No. 9 to Registration Statement (Form S-1) (Aug. 18, 2004).

87. Kate Conger, *Google Removes 'Don't Be Evil' Clause from Its Code of Conduct*, GIZMODO (May 18, 2018, 5:31 PM), <https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393> [<https://perma.cc/RJG8-SL3Z>].

88. Microsoft News Center, *OpenAI Forms Exclusive Computing Partnership with Microsoft to Build New Azure AI Supercomputing Technologies*, MICROSOFT (July 22, 2019), <https://news.microsoft.com/2019/07/22/openai-forms-exclusive-computing-partnership-with-microsoft-to-build-new-azure-ai-supercomputing-technologies/> [<https://perma.cc/9EHT-9R2Y>].

89. *Id.*

90. See Dave Gershgorin, *GPT-3 Is an Amazing Research Tool. But OpenAI Isn't Sharing the Code*, MEDIUM (Aug. 20, 2020), <https://onezero.medium.com/gpt-3-is-an-amazing-research-tool-open-ai-isnt-sharing-the-code-d048ba39bbfd> [<https://perma.cc/X6W4-654A>].

In the legal context, the proprietary aspect of algorithms adds to the opacity of the process.⁹¹ As mentioned earlier, these algorithms are mostly considered trade secrets, shielded from the prying eyes of both guardians of the legal sphere and citizens. When a source code is secret, it is like a law with hidden articles and chapters or, more accurately, it is like an invisible legal framework. In the case of *Loomis*, discussed previously, the method by which the assessment tool concluded the defendant was a high risk to society was not clear.⁹² Perhaps he was a dangerous person at that time—considering his track record,⁹³ that is more than likely. But as long as the workings of such algorithms are not supervised by an independent authority, they pose serious doubts about due process.

Most legal professionals are familiar with the early twentieth century novel *The Trial*,⁹⁴ so perhaps the term *Kafkaesque* springs to mind. *The Trial* tells the story of Josef K., a man arrested and prosecuted by a remote, inaccessible authority for unknown crimes.⁹⁵ The novel follows his frustrating quest to find out of what exactly he has been accused.⁹⁶ It is agonizing to read this book, to be lost in the labyrinth of invisible Laws, with a shadowy, untouchable Court.⁹⁷ Human alienation, existential anxiety, guilt, and absurdity are recurring themes of the Bohemian writer, Franz Kafka. It earned him a reference in the world’s dictionaries. Merriam-Webster describes Kafkaesque as “having a nightmarishly complex, bizarre, or illogical quality.”⁹⁸

If we allow proprietary legal tools with secret algorithms to guide—or at some point, even replace—our judiciary, we may ultimately find ourselves in a similar labyrinth of invisible algorithms, with a shadowy, untouchable court. For now, these tools are still in their infancy. But once their algorithms gain sophistication, they are bound to become nightmarishly complex, bizarre, and illogical.

VII. MIND THE GREY ZONE

The legal sphere is certainly an area that could benefit from innovation, but not at any cost. The otherwise conservative legal community is now so eager to innovate that it seems to overlook the threat that such an

91. See generally *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016); Crooke, *supra* note 14.

92. See *supra* Part I.

93. See *Loomis*, 881 N.W.2d at 761.

94. FRANZ KAFKA, *THE TRIAL* (Ritchie Robertson ed., Mike Mitchell trans., Oxford Univ. Press 2009) (1925).

95. *Id.*

96. *Id.*

97. *Id.*

98. *Kafkaesque*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/Kafkaesque> [<https://perma.cc/TT5Q-GDET>] (last visited Apr. 14, 2021).

unconstrained implementation may have on the rule of law. A wide spectrum of policies and legislation is necessary to protect the public, but who is going to be charged with those tasks? Few legal professionals have an interest in technology—let alone insight into the potential ramifications of specific legal tools. Even lawmakers do not always know what they are supposed to be legislating. This lack of interest is mirrored by the indifference of technological experts regarding legal matters. Most cannot be bothered by what is typically considered tedious gibberish that only holds them back.

Herein lies the challenge. It is complex, and it is continuously evolving.⁹⁹ These AI products are in a permanent beta mode, which makes them even harder to regulate. Consequently, this interdisciplinary space is a free-for-all—a hardly regulated grey zone. Ironically, many of the intelligent machines being built may be better equipped to cope with the future than we are. They can be designed to be truly interdisciplinary. Soon, only computers will understand computers. The knowledge gaps that separate technology and legal specialists are easily filled by artificial intelligence.

VIII. HYBRID FRAMEWORK

A future-proof legal system would benefit from regulation and transparency. Open-source legal tools should merge into the analog legal system, not coexist in a cloud of secrecy beside it. There is a window of opportunity to build a solid, hybrid framework, but that window is closing. Most countries took centuries to formulate and enrich their legal systems; however, we do not have that kind of time in the fast-paced digital age. Visionary and bold lawmakers are needed. At the very least, those administrative officials who actually write the laws need to be well-versed on what is occurring in their field. That will require academic programs and continuous in-service training at the cutting edge of technology, as well as close collaboration with computer scientists. Unfortunately, such coveted legal-technology hybrids are likely to be poached by the very corporations they should be regulating.

99. See *Rule of Law: Legislating the Digital Age*, NEXT INST. FOR TECH. & L., <https://nextinstitute.org/rule-of-law/> [<https://perma.cc/VXL5-T7HY>] (last visited Apr. 14, 2021).