IF YOU THINK AI WON'T ECLIPSE HUMANITY, YOU'RE PROBABLY JUST A HUMAN

Gary D. Brown*

> Day by day, however, the machines are gaining ground upon us;
> day by day we are becoming more subservient to them.
> —Samuel Butler, *Darwin Among the Machines*[1]

Building machines that can replicate human thinking and behavior has fascinated people for hundreds of years. Stories about robots date from ancient history through da Vinci to the present.[2] Whether designed to save labor or lives, to provide companionship or protection, loyal, capable, productive machines are a dream of humanity.

The modern manifestation of this interest in using human-like technology to advance social interests is artificial intelligence (AI). This is a paper about what that interest in AI means and how it might develop in the world of national security.

I. WHAT IS AI?

Over the past few decades, philosophers have provided most of the energy in the AI conversation. That was true because AI was mainly a science fiction staple, neither fully realized nor even clearly conceived. More recently technology has started to close the gap between fictional portrayals of capabilities and reality, and AI is being used in real world applications, including in the defense and intelligence communities.[3]

As AI has appeared on the security stage, lawyers are suddenly relevant. Whenever policy discussions focus on national security, legal issues tend to dominate. That means lawyers are necessary—and seemingly ever-present. One of the (often frustrating) values lawyers bring to strategic discussions is an insistence on precision in terminology, so the first order of business here is to define what is meant by artificial intelligence (AI).[4] There are many definitions of AI. One is "the theory and development

* Professor of Cyber Law, College of Information and Cyberspace, National Defense University. The views expressed in this Article are those of the author and are not an official policy or position of the National Defense University, the Department of Defense, or the U.S. Government.

[1] Samuel Butler, *Darwin Among the Machines*, *in* A FIRST YEAR IN CANTERBURY SETTLEMENT WITH OTHER EARLY ESSAYS 184 (R. A. Streatfeild, 1914).

[2] Michael E. Moran, *Epochs in Endourology: The da Vinci Robot*, 20 J. OF ENDOUROLOGY 986, 989 (2006). *See* Matt Simon, *The WIRED Guide to Robots*, WIRED (Apr. 16, 2020, 9:00 AM), http://www.wired.com/story/wired-guide-to-robots [https://perma.cc/RJ87-SYN3].

[3] *See id.*

[4] *See infra* note 6 (providing a definition will also help avoid the common issue of conflating AI and machine learning (ML), a related but not coterminous concept).

of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."[5] This definition leads neatly into a discussion of the conference often considered the birthplace of the concept of AI.

The term AI was apparently coined in 1956 when a group of researchers gathered for the Dartmouth Summer Research Project on Artificial Intelligence.[6] The participants set out there that they were to act "on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."[7] That may seem like a very reductionist view of human intelligence, to say nothing of personhood, but it was the genesis of a fascinating line of thinking.

To keep things in perspective, it might be useful to go back in time even further than the Dartmouth conference. Douglas Hofstadter suggests that AI began "at the moment when mechanical devices took over any tasks previously performable only by human minds."[8] He sets that beginning as Charles Babbage's calculating engine in 1821.[9] Of course, performing simple arithmetical functions is not the kind of activity we would now consider an example of AI. But it's not just calculators that we no longer consider AI. Other things that seemed marvelous at the time but are utterly ho-hum now include proportional typefaces (like this one) and optical character recognition—for example scanning a hard copy document to an editable PDF. These are functions that used to require human intelligence but now are performed by simple computers. That kind of AI did not seem threatening at its birth, and certainly does not now.[10]

Hofstadter's discussion of AI also referred to what he calls Tesler's theorem, which defines AI as "whatever hasn't been done yet."[11] In 1979, Hofstadter's list of

[5] *Artificial Intelligence*, OXFORD REFERENCE (2021), https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960 [https://perma.cc/KJY2-ABPN].

[6] Bernard Marr, *The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance*, FORBES (Feb. 14, 2018), https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#7e64885d4f5d [https://perma.cc/2L8Y-GL5Z].

[7] J. McCarthy et al., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (Aug. 31, 1955) (unpublished report), http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf [https://perma.cc/8R7F-X478].

[8] DOUGLAS HOFSTADTER, GÖDEL, ESCHER, BACH: AN ETERNAL GOLDEN BRAID 601 (1979).

[9] *Id.* at 600–01.

[10] Unlike "artificially intelligent" elevators, which caused great consternation when driverless elevators were introduced in New York in the 1940s. Steve Henn, *Remember When Driverless Elevators Drew Skepticism*, NPR (July 31, 2015, 5:08 AM), https://www.npr.org/2015/07/31/427990392/remembering-when-driverless-elevators-drew-skepticism [https://perma.cc/8D9R-BTHC].

[11] HOFSTADTER, *supra* note 8, at 601.

goals for AI included playing checkers, understanding spoken words, and facial recognition, all of which a few decades later are quite competently performed by the cell phone we carry in our pockets.[12] Perhaps AI is on a sliding scale that will always be over the horizon, just beyond technology with which we are comfortable. If that is the case, then scary AI will continue to make great movie fodder, but will never actually arrive.

Since that definition would not be especially useful in policy discussions about choosing the best track for development of the technology, this Article will use the following. Artificial intelligence is a computer algorithm capable of adapting to or learning from unique or unanticipated circumstances without human input.

II. FEARS ABOUT AI

> I toast therefore I am.
>
> —Talkie Toaster, *Red Dwarf*[13]

There have been cries about the dangers of AI for decades,[14] but that has not, and likely will not, slow its development. In this century's first decade, industry and governments were aware of growing cyber insecurity and the ease with which hackers were able to penetrate seemingly any system connected to the internet. About the time when cybersecurity professionals were throwing up their hands and talking less about "cyber defense" and more about "cyber resilience"—accepting that defense would fail and recovery would be necessary—the practice of connecting electronic devices to the internet was becoming so widespread that it needed a name. That name turned out to be the Internet of Things.[15] The trend started fast and continued faster. Since 2012, the number of connected devices has increased from under 9 billion to over 50 billion.[16] This figure includes devices that previously were

[12] HOFSTADTER, *supra* note 8, at 601–02.

[13] *Red Dwarf: White Hole* (BBC Studios broadcast Mar. 7, 1991).

[14] *See infra* notes 19–23.

[15] According to Google, the terms "internet of things" and "cyber resilience" essentially appeared in books in 2006 and 2009, respectively. *Cyber Resilience*, GOOGLE BOOKS NGRAM VIEWER, https://books.google.com/ngrams/graph?content=cyber+resilience &year_start=2007 &year_end=2019&corpus=26&smoothing=3 [https://perma.cc/WMW8-MTZN] (tracking use of term "cyber resilience" in books between 2007 and 2019); *Internet of Things*, GOOGLE BOOKS NGRAM VIEWER, https://books.google.com/ngrams/graph?content=internet+of+things&year _start=2006&year_end=2019&corpus=26&smoothing=3 [https://perma.cc/KP5K-TUAD] (tracking use of term "internet of things" in books between 2006 and 2019). The term "cyber defense", on the other hand, made its significant appearance in 1994. *Cyber Defense*, GOOGLE BOOKS NGRAM VIEWER, https://books.google.com/ngrams/graph?content=cyber+defense& year_start=1970&year_end=2019&corpus=26&smoothing=3 [https://perma.cc/DDX2-QXC3] (tracking use of term "cyber defense" between 1970 and 2019).

[16] *The IoT Data Explosion: How Big is the IoT Data Market?*, PRICEONOMICS (Jan. 9,

not connected to the internet, i.e., not things like computers and smartphones.[17] So, about the time when it became apparent that the internet was unsecure (if not unsecurable), the collective decision was made to connect all the things that make modern life possible—refrigerators, utility grids, thermostats, etc.—to the internet! The point is that knowing that pursuing a technology is certain to increase risk of real harm to society seems to have little effect on the speed at which that technology is developed.

In the case of AI, most of the dire scenarios sketched out take the form of either killer robots powered by malevolent AI or AI that becomes so advanced that, while not actively evil, simply determines people are too insignificant to be relevant and acts in ways that crush the human race.[18]

The concern about dangerous AI led experts in the field to release an open letter warning of the risks of using AI in offensive autonomous weapons "beyond meaningful human control."[19] Two of the most famous signatories of the letter were theoretical physicist Stephen Hawking and Elon Musk, tech entrepreneur.[20] Conveniently, they represent the two categories of concern set out above.

In an interview about his position, Stephen Hawking warned that advanced AI might wipe out humanity.[21] Hawking was worried less about actively malicious AI than he was about AI that might kill humans incidentally in pursuit of benevolent (or at least not malevolent) ends.[22]

Musk, on the other hand, seems to fear actively malevolent AI turning weapons or tools it controls on its human creators in a bid to terminate the species.[23] The Muskian concern presupposes that, by design or otherwise, the AI has the ability to set or interpret its own goals.[24] Perhaps the most effective way to address this concern

2019), https://priceonomics.com/the-iot-data-explosion-how-big-is-the-iot-data/ [https://perma.cc/Z835-26KJ].

[17] *Id.*

[18] This leaves aside the significant concern about developing an entity that could eventually become similar enough to humans to earn consideration for human rights. That moral concern is largely beyond the scope of this Article but merits more attention than it currently receives.

[19] *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, FUTURE OF LIFE INST. (July 28, 2015), https://futureoflife.org/open-letter-autonomous-weapons/ [https://perma.cc/6TLA-355N].

[20] *Id.*

[21] Andrew Griffin, *Stephen Hawking: Artificial Intelligence Could Wipe Out Humanity When It Gets Too Clever as Humans Will Be Like Ants*, THE INDEPENDENT (Oct. 8, 2015, 3:28 PM), https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-could-wipe-out-humanity-when-it-gets-too-clever-humans-could-become-ants-being-stepped-a6686496.html [https://perma.cc/74S5-SDSB].

[22] *Id.*

[23] Aatif Sulleyman, *AI Is Highly Likely to Destroy Humans, Elon Musk Warns*, THE INDEPENDENT (Nov. 24, 2017, 8:01 PM), https://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-artificial-intelligence-openai-neuralink-ai-warning-a8074821.html [https://perma.cc/JYL2-4N95].

[24] This leaves aside the possibility of an evil creator, but that person could make an

is to ensure AI has no ability to directly control strategic weapons systems. Self-directed AI would perhaps teach us a lot about learning and motivation, and as long as it lacks an ability to carry out nefarious plans in the physical world, potential hazards are contained.[25]

Hawking's concern presents a greater challenge. Machines are designed to perform specific functions, but humans often misuse them in creative ways, such as using a screwdriver as a chisel or a chair as a step-stool. If we continue to work toward human-like AI, this type of creativity seems likely to appear. AI-powered cars, blenders, furnaces, etc. are all put to productive use when they respond to human desires to drive, slice/dice, or heat to a comfortable temperature. Internet connectivity is becoming standard for the most mundane household appliances, and elements of AI are sure to follow. Why not have a car that learns the driver's schedule and warms itself up in the winter? Or kitchen appliances that monitor comments about meals and respond by changing cooking times or temperatures? If linked to enough processing power to host AI, these devices might decide that blending is the most important thing in the universe and *everything* should be blended, for example. This is often referred to as the Paperclip problem.[26] Added to the earlier point that it may be difficult to prevent AI from jumping over the metaphorical fence, Hawking's concern seems legitimate.

Before dismissing this as pure science fiction, recall the goal is to develop AI that can adapt to unique circumstances, i.e., become more like a human.[27] Even when not malevolent, humans sometimes act dangerously, whether through error or in ill-fated attempts to achieve a legitimate goal. AI designed to reason like a human seems likely to do the same.

Despite the challenges, AI has started to make an appearance in the national security context.[28] Exploring how that might develop is the subject of the remainder of the Article.

III. AI IN NATIONAL SECURITY

> The whole point of the Doomsday Machine is lost if you keep it a secret![29]

appearance even if AI development were to be completely banned, and so it would add nothing of note to this discussion.

[25] NICK BOSTRUM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 158–60 (Oxford Univ. Press 2014).

[26] Nick Bostrum, *Ethical Issues in Advanced Artificial Intelligence*, https://nickbostrom.com/ethics/ai.html [https://perma.cc/6TYD-F6SH] (last visited Dec. 13, 2021).

[27] McCarthy et al., *supra* note 7.

[28] *See infra* notes 32–53 and accompanying text.

[29] DR. STRANGELOVE OR: HOW I STOPPED WORRYING AND LOVE THE BOMB (Columbia Pictures 1964) (Dr. Strangelove said this to the Soviet Ambassador in response to the

It seems fairly obvious that a system capable of destroying the planet without human input would be a bad idea. Thankfully, even those comfortable with the idea of AI shy away from the thought of developing such a capability and then handing the keys to an AI.[30] In the case of more distinct, articulable tasks supporting national security, there may well be a role for AI, however, to replace or supplement human efforts. The challenge is that so many national security functions involve creative thinking or unscripted responses to emerging situations. AI might eventually be quite good in these roles, but these are precisely the situations that concern AI skeptics.

Forms of AI have already been incorporated into national security systems. Among the areas that use AI are kinetic weapons, cyber security, and cyber defense.[31] Lethal autonomous weapons systems (LAWS) continue to stir strong emotions, especially in the humanitarian community, but they are already employed in various iterations.[32] Some have a human in or on the loop, but that standard is a slippery one, and it is likely to erode further as we go forward. The control here is the ability of programmers to set and enforce appropriate goals for LAWS. In the long run, we will see how it turns out.

Less scrutable is the use of AI in cyber warfare. While the general objections to using AI for cyber warfare are no different than those to its use in kinetic military activities, cyber warfare is still a developing and largely unexplored field. AI can certainly be used to inform the decisions of human operators and, if it is, the onus is on the human choosing to follow or not follow the advice. The execution of cyber operations can happen very quickly; however, it might defeat the purpose of incorporating lightning-fast AI into the system if action were to be stayed waiting for human input. The efficient way to use AI in cyber warfare would be to allow the AI to select its own action and then engage the target without human input. Since that engagement might include significant disruptive or destructive actions, caution is warranted. Because of the concern, using AI for cyber offense is improbable in the immediate future. The use of AI in cyber defense is more likely and is, in fact, already being done.[33]

Soviets' failure to tell the United States about the machine). Apparently, the USSR was actually working on developing a doomsday system that would have launched nuclear missiles at the United States, potentially without human input, if it detected nuclear explosions inside the Soviet Union. Eric Schlosser, *Almost Everything in "Dr. Strangelove" Was True*, NEW YORKER (Jan. 17, 2014), https://www.newyorker.com/news/news-desk/almost-everything-in-dr-strangelove-was-true [https://perma.cc/LN2M-FUCF].

[30] It would be comforting to conclude that it is beyond comprehension that the device itself would be developed, but alas that would be overly optimistic.

[31] KELLY SALYER, CONG. RSCH. SERV., R45178 ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY (2020).

[32] *Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control*, HUM. RTS. WATCH (Aug. 10, 2020), https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and [https://perma.cc/6QV9-SMFG].

[33] SALYER, *supra* note 31.

How might AI be used in cyber defense? Functions referred to as AI can analyze huge amounts of data and use the analysis to predict or categorize behavior. For example, it might be used to establish patterns of behavior of state-sponsored hackers or uncover links between individuals.[34] A unique way AI could be used to defend cyber systems is to generate decoy material to entice state-sponsored (and other) hackers and thieves, and prevent them from stealing real files, including intellectual property.[35] AI might be able to use data from connected devices (the IoT) to develop threat prediction models.[36] When limited to cyber defense, AI seems less frightening, but leveraging such capabilities in support of influence campaigns is another matter.

As difficult as the issues surrounding AI can be in the areas of kinetic weapons and more traditional cyber national security capabilities, the challenges are orders of magnitude more difficult when it comes to information. U.S. adversaries—especially Russia—have learned to use AI-driven, cyber-enabled information operations to great effect in the United States.[37] In a democracy as large and diverse as the United States there are a number of differing opinions on a great number of subjects, which makes public opinion an easy target. Russian influence operations often use pre-existing disagreements to increase fractures and distrust in U.S. society.[38]

AI-authored articles can be difficult to distinguish from those humans write. For several years, AI programs have written brief news articles about sports and the weather.[39] Perhaps more impressive, students at MIT used a program they wrote to create academic papers for submission to conferences.[40] Their first manufactured

[34] Jennifer Valentino-DeVries, *How Your Phone Is Used to Track You, and What You Can Do About It*, N.Y. TIMES (Aug. 19, 2020), https://www.nytimes.com/2020/08/19/technology/smartphone-location-tracking-opt-out.html?.?mc=aud_dev&ad-keywords=auddevgate&gclid=Cj0KCQjw7MGJBhD-ARIsAMZ0ees7D7oP9iBZH4fv9YKWMdvld8fFBn8RVFRfaPt0XpgOsVaozvd_P18aAk4sEALw_wcB&gclsrc=aw.ds [https://perma.cc/S6ZJ-ULQX].

[35] Jonathan Voris et al., *Bait and Snitch: Defending Computer Systems with Decoys*, DEP'T COMPUT. SCI. COLUM. UNIV. 1, 2–3, 12–13, 22–23 (2013).

[36] Elie Bursztein, *Inside the Infamous Mirai IoT Botnet: A Retrospective Analysis*, CLOUDFLARE (Dec. 14, 2017), https://blog.cloudflare.com/inside-mirai-the-infamous-iot-botnet-a-retrospective-analysis/ [https://perma.cc/G27S-2SPZ] ("What's remarkable about [Mirai's] record-breaking attacks is they were carried out via small, innocuous IoT devices like home routers, air-quality monitors, and personal surveillance cameras.").

[37] Nicholas Thompson & Issie Lapowsky, *How Russian Trolls Used Meme Warfare to Divide America*, WIRED (Dec. 17, 2018), https://www.wired.com/story/russia-ira-propaganda-senate-report/ [https://perma.cc/C46T-HV3M].

[38] Alina Polyakova, *Lessons from the Mueller Report on Russian Political Warfare*, BROOKINGS (June 20, 2019), https://www.brookings.edu/testimonies/lessons-from-the-mueller-report-on-russian-political-warfare/ [https://perma.cc/A4CD-WLAK].

[39] *See* Jaclyn Peiser, *The Rise of the Robot Reporter*, N.Y. TIMES (Feb. 5, 2019), https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html [https://perma.cc/ARL2-MJ5H].

[40] *See* SCIgen—An Automatic CS Paper Generator, https://pdos.csail.mit.edu/archive/scigen/ [https://perma.cc/5SJ2-7U9T] (last visited Dec. 13, 2021).

paper was accepted to a conference in 2005.[41] Should we be concerned about the increasing role of machines in writing what we read? An op-ed in the *Guardian* suggesting that AI presents no threat to humanity offers scant reassurance—an AI authored it![42]

The ability of AI to manufacture text that appears to have been written by a human, and to do it at unprecedented speed and scale, generates a new national security challenge. It is easy enough, as so often happens in cyber-related fields, for national security professionals to dismiss AI-driven influence operations as "nothing new."[43] As has become clear in other cases, exponential increases in speed and scale are, in fact, enough to indeed create something new. Tried-and-true counter-espionage techniques aimed at preventing humans from stealing sensitive information proved insufficient to prevent the damage to national security caused by the vast export through cyber means of intellectual property, security clearance files, and other sensitive data.[44] It was still espionage, but cyber techniques made it different. Similarly, Cold War era tactics to protect the state from provocative radio broadcasts, newsletters, and other propaganda are, quite simply, inadequate to deal with hundreds of AI-powered bots flooding social media platforms with millions of messages.[45] National security professionals might want to call it the same thing, but the effects are profoundly different, and the required responses are not similar, so it is difficult to see how calling it the same thing is helpful.

Using traditional international relations tools to address the use of AI in these situations is complicated by the paucity of international law that is clearly applicable

[41] *See* Adam Conner-Simons, *How Three MIT Students Fooled the World of Scientific Journals*, MIT NEWS (Apr. 14, 2015), https://news.mit.edu/2015/how-three-mit-students-fooled-scientific-journals-0414 [https://perma.cc/2557-G3GP].

[42] *See* GPT-3, *A Robot Wrote this Entire Article. Are You Scared Yet, Human?*, THE GUARDIAN (Sep. 8, 2020), https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3 [https://perma.cc/9SG7-N63Y].

[43] *E.g.*, Simon Crosby, *Separating Fact from Fiction: The Role of Artificial Intelligence In Cybersecurity*, FORBES (Aug. 21, 2017), https://www.forbes.com/sites/forbestechcouncil/2017/08/21/separating-fact-from-fiction-the-role-of-artificial-intelligence-in-cybersecurity/?sh=103e66971883 [https://perma.cc/C3JP-LZ65].

[44] *See generally* Reuters, *China Theft of Technology Is Biggest Law Enforcement Threat to US, FBI Says*, THE GUARDIAN (Feb. 6, 2020), https://www.theguardian.com/world/2020/feb/06/china-technology-theft-fbi-biggest-threat [https://perma.cc/2N5D-XK6Y]; Josh Fruhlinger, *The OPM Hack Explained: Bad Security Practices Meet China's Captain America*, CSO (Feb. 12, 2020), https://www.csoonline.com/article/3318238/the-opm-hack-explained-bad-security-practices-meet-chinas-captain-america.html [https://perma.cc/8U4F-5RZS]; Ken Dilanian, *Suspected Russia Hack: Was It an Epic Cyber Attack or Spy Operation?*, NBC NEWS (Dec. 18, 2020), https://www.nbcnews.com/news/us-news/suspected-russian-hack-was-it-epic-cyber-attack-or-spy-n1251766 [https://perma.cc/856V-3DMC].

[45] Ashley Deeks, Sabrina McCubbin & Cody M. Poplin, *Addressing Russian Influence: What Can We Learn From U.S. Cold War Counter-Propaganda Efforts?*, LAWFARE BLOG (Oct. 25, 2017), https://www.lawfareblog.com/addressing-russian-influence-what-can-we-learn-us-cold-war-counter-propaganda-efforts [https://perma.cc/9HYU-MGAT].

to influence operations. Because influence campaigns themselves fall below the level that would trigger armed conflict, and often occur in the absence of pre-existing armed conflict, the law of war (international humanitarian law) does not apply.[46] If there is an international law transgression, it is most likely a violation of sovereignty. However, international law can be opaque in the best of circumstances and cyber-enabled information operations are rather new This combination makes it even harder to determine when violations of sovereignty occur than it would be in more traditional circumstances.

The applicable legal standard is the non-intervention principle, which prohibits states from coercively interfering in the affairs of other states.[47] It is unclear what circumstances would push operations to disseminate information across the non-intervention threshold. Complicating matters, AI-driven persuasion operations can include misinformation, disinformation, and other information that is simply targeted for effect.[48] As a result, influence operations driven by AI should be expected to increase.

IV. CONTROLLING AI

> It's alive!
>
> —Dr. Frankenstein, *Frankenstein* (Universal Pictures, 1931)[49]

Given that some forms of AI are already being used in national security systems,[50] and history suggests AI development and use will continue, it is vital to consider ways to ensure AI remains in check and responsive to human goals.

In national security situations, the temptation is to give AI the ability to act as well as think, because humans "in the loop" are too slow to be effective, or in any event too slow to catch up with the AI's reasoning. The worst-case scenario would be to have poorly trained humans forced to intervene at a moment of crisis. If an AI *can* act independently, the time when it *must* act independently is during crises.

[46] INT'L COMM. OF THE RED CROSS, HOW IS THE TERM "ARMED CONFLICT" DEFINED IN INTERNATIONAL HUMANITARIAN LAW? 1–2 (2008).

[47] MICHAEL N. SCHMITT, TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS 312–25 (Michael N. Schmitt ed., 2d ed. 2017). The non-intervention principle is well-established in international law, but not succinctly phrased in official documents. The pithiest statement is *Tallinn Manual 2.0*, Rule 66. The accompanying text explains the source and scope of the rule.

[48] Christina Nemr & William Gangware, *Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age*, PARK ADVISORS (Mar. 2019), https://static1.squarespace.com/static/5714561a01dbae161fa3cad1/t/5c9cb93724a694b834f23878/1553774904750/PA_WMD_Report_2019.pdf [https://perma.cc/3PF6-GMYR].

[49] FRANKENSTEIN (Universal Pictures 1931).

[50] SALYER, *supra* note 31.

Determining how to control AI seems paradoxical. The goal has been to design AI to be very, very smart—ultimately, smarter than the most intelligent human. While maintaining the raw intellectual capacity, we also desire AI to have an elevated sense of morality or, at a minimum, to be free of the prejudices, biases, mood swings, etc. that exist in normal, fallible human beings.[51]

The current approach to controlling AI is to insist on transparent decision-making, ensuring that humans will be able to understand, second-guess, and correct for errors in chains of decisions.[52] While not useless, this seems to be a temporary solution, at best. First, as AI develops, the factors involved in its decision-making will continue to increase exponentially, making it more and more difficult for humans to ingest, analyze, and understand how factors one through ten million led to the decision in question. The ostensible answer to this is that, before the point of human incomprehension is reached, the AI would be dumbed down. As noted, however, limiting AI's capability is not the way this is likely to play out.

Complicating the issue further is the goal of AI developers to create "human-like intelligence."[53] As developers move AI closer to human intelligence, its decisions will be based on a greater number of factors, will be less explainable, and may even appear random or counterintuitive. Just as humans may not be able to articulate the rationale behind every poor decision (which are the ones that cause concern with AI), human-like AI may not be able to provide objective reasoning behind each and every decision. Human reasons for apparently irrational acts of bravery or violence, for example, may include doing something for love, revenge, pride, or for reasons unknown even to them. Of course, humans also prevaricate or even lie about decisions to protect themselves or others.[54] Programmers would have to ensure that human-like AI is incapable of engaging in the very human act of deception. It is perhaps a fundamental inconsistency that we want to create AI that is superhuman in every way: mentally, physically (through control of machines), and even morally—yet still tolerate inferior humans in a position of control over it. That does not sound like a sustainable model for the long term, but for now it may be the best option.

[51] Slava Polonski, *Can We Teach Morality to Machines? Three Perspectives on Ethics for Artificial Intelligence*, MEDIUM (Dec. 19, 2017), https://medium.com/@drpolonski/can-we-teach-morality-to-machines-three-perspectives-on-ethics-for-artificial-intelligence-64fe479e25d3 [https://perma.cc/SEC2-K5PS].

[52] Greg Satell & Josh Sutton, *We Need AI That Is Explainable, Auditable, and Transparent*, HARV. BUS. REV. (Oct. 28, 2019), https://hbr.org/2019/10/we-need-ai-that-is-explainable-auditable-and-transparent [https://perma.cc/RLQ6-RD3S].

[53] Janna Anderson & Lee Raine, *Artificial Intelligence and the Future of Humans*, PEW RSCH. CTR. (Dec. 10, 2018), https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/ [https://perma.cc/7FER-J7XT].

[54] Christian L. Hart et al., *Development of the Lying in Everyday Situations Scale*, 132 AM. J. PSYCH. 343, 344–46 (2019).

CONCLUSION

AI will continue to develop at a rapid pace, and it is almost certain to be used in an increasing number of roles, limited only by human (or its own) imagination. As discussed earlier, there is no indication that knowledge of the risks involved will stop, or even slow, this development and use. Rather than joust this windmill, efforts should be focused on mitigating the inherent risks. Simply put, policy makers should focus on how best to control this new creation.

As noted above, transparency in AI decisions is a step, but unlikely to be the ultimate solution. Perhaps, as with other national security capabilities, only tragic consequences will circumscribe the use of AI. Atomic, biological, and chemical weapons were all used before there was consensus over condemning their use. Cyber capabilities continue to be used without effective restriction, and many experts believe it will take an extraordinary event to generate limits. Unfortunately, in the case of AI, the same might be true. And, as what we have to hope is that the event that triggers universal concern will not be too catastrophic—and that it will occur before we lose the capability to maintain control over our creation.

On a separate note, perhaps it is not too early to think about what it actually means to be alive. What does it mean to have a soul? AI already demonstrates creativity through creating music, art, and poetry.[55] Is the important part of life the biological vessel for the mind, or the mind itself? Without answering this fundamental question, we will not have a satisfactory way to deal with AI. If our machines become advanced enough to reason as humans, it seems likely they will develop a personality. If that personality turns out to be malicious, society must stand ready to punish AI. Will that punishment be more similar to a citizen, with rights, or will it be as a wild animal, "killing" them by wiping their hard drive without the benefit of a trial?

As discussed above, there is no bright line that determines when a sophisticated algorithm becomes AI. Lawyers might well conclude that we have an AI when a court determines that a given algorithm merits recognition as an entity with rights. This may seem farfetched, but less than two-hundred years ago, the U.S. Supreme Court denied some human beings recognition as fully human,[56] and more recently it determined that corporations were, in some ways, entitled to rights.[57] It does not seem a stretch to imagine that the law would continue to progress to extend rights to AI.

[55] James Vincent, *Can You Tell the Difference Between Bach and RoboBach?*, THE VERGE (Dec. 23, 2016, 10:25 AM), https://www.theverge.com/2016/12/23/14069382/ai-music-creativity-bach-deepbach-csl [https://perma.cc/A6RK-PRRJ].

[56] *See* Dred Scott v. Sandford, 60 U.S. 393, 517 (1856) (stating that a free black person could not be an American citizen under the Constitution if their ancestors were brought over as slaves from Africa).

[57] *See* Citizens United v. Fed. Election Comm'n, 130 S. Ct. 876, 882–83 (2010) (holding that the government may not suppress "political speech on the basis of the speaker's corporate identity").

Right now, websites leverage AI's infallibility to prove it is not human, using tricky captchas that assume a user is human if the user errs.[58] Similarly, we might argue that flaws and imperfections are not bugs but rather features of humanity. That does not seem to be the trend, however. Modern society, with its increasingly detailed normative and legal underpinnings, celebrates perfection and punishes errant behavior. Given that perfection appears to be the goal, it is inevitable that AI will eventually surpass human capacity. Perhaps the best we can do in the meantime is to refrain from intentionally building machines designed to cause mayhem and grief, while also embracing the notion that humanity's computer progeny will eventually outclass their creators. We might then hope to ensure that when humanity fades into the celestial dust it will be with a whimper and not a bang.

[58] Josh Dzieza, *Why CAPTCHAs Have Gotten So Difficult*, THE VERGE (Feb. 1, 2019, 11:00 AM), https://www.theverge.com/2019/2/1/18205610/google-captcha-ai-robot-human-difficult-artificial-intelligence [https://perma.cc/CPS6-THFX].