Washington Journal of Law, Technology & Arts

Volume 14 | Issue 1

Article 4

10-1-2018

Blame It on the Machine: A Socio-Legal Analysis of Liability in an AI World

Michael Callier

Harly Callier

Follow this and additional works at: https://digitalcommons.law.uw.edu/wjlta

Part of the Science and Technology Law Commons

BLAME IT ON THE MACHINE: A SOCIO-LEGAL ANALYSIS OF LIABILITY IN AN AI WORLD[*]

Michael Callier and Harly Callier[**]

ABSTRACT

As technology continues to evolve, interactions between humans and artificial intelligence ("AI") will skyrocket. It is important to understand the impact AI can have on society, as well as the potential harm and subsequent liability that could result, and to develop best practices designed to address them. The U.S. needs a comprehensive framework to govern the design, creation, use and risks associated with AI. At the time of this writing, no such framework has been implemented.

This article takes a socio-legal, interdisciplinary approach to explore ideas on socio-ethical concerns and theories of liability related to AI, and applies a sociological perspective to assess existing legal frameworks that currently govern human-AI

---

50   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

interaction. By adopting an interdisciplinary approach, this article seeks to encourage holistic and robust dialogue about how AI could be developed and operated, hoping that humans and AI can coexist harmoniously. It also proposes a framework to regulate such development in the U.S.

There are a few limitations in this article. First, due to the accelerated pace of technological change, the future state of AI will be different from the current state. Hence, the framework proposed in this article might eventually become obsolete. Second, this article is derived from secondary sources and, although the information collected includes rich empirical data, no primary data was generated other than the authors' views. Third, only specific aspects of AI were selected for analysis – there are other factors in policy, sociology and law that are not addressed. Lastly, this article is primarily focused on Western cultures, North America and Europe in particular; hence, it might not be applicable globally.

TABLE OF CONTENTS

INTRODUCTION

Some believe that, in ten to twenty years, human intelligence and artificial intelligence will be equal.[1] According to futurist Ray Kurzweil's Law of Accelerating Returns, by the year 2045, superintelligence, capable of self-improvement, will cause an intelligence explosion or "technological singularity" superior to human intelligence. Under this theory, AI will be autonomous and able to act independent from the will of humanity.[2]

As Germany's Ethics Commission for Automated and Connected Driving (2016) stated, a primary question to ask as we move into the future of AI is: "[w]hat technological development guidelines are required to ensure that we do not blur the contours of a human society that places individuals, their freedom of development, their physical and intellectual integrity and their entitlement to social respect at the heart of its legal regime?"[3] This question requires holistic assessment from both a sociological and legal perspective. Accordingly, this article explores the relationship between human beings and AI, and the laws that govern their interactions. This article first examines how humans interact with

---

[1] *See* RAY KURZWEIL, THE AGE OF SPIRITUAL MACHINES: WHEN COMPUTERS EXCEED HUMAN INTELLIGENCE 40-47 (1999).

[2] Paulius Čerka et al., *Liability for Damages Caused by Artificial Intelligence,* 31 COMPUTER L. & SECURITY REV. 376, 382 (2015) (citing Ray Kurzweil, *The Law of Accelerating Returns* (Mar. 7, 2001), http://www.kurzweilai.net/the-law-of-accelerating-returns).

[3] FEDERAL MINISTRY OF TRANSPORT AND DIGITAL INFRASTRUCTURE, ETHICS COMMISSION: AUTOMATED AND CONNECTED DRIVING (2017).

AI, and in particular, with Socially Assistive Robotics. Next, this article examines ethical and moral concerns surrounding human-AI interactions. Third, this article seeks to understand whether a sufficient legal and regulatory environment capable of governing human-AI interaction exists. Finally, this article proposes a moral and legal framework to guide AI development.

I. BACKGROUND

A. *Definitions of Artificial Intelligence and Machine Learning*

The definition of AI is a work in progress as there is no universally accepted definition. [4] For example, MIT Professor Patrick Winston defined AI as "the study of computations that make it possible to perceive, reason and act."[5] In a later lecture, Winston further described AI as "algorithms or procedures enabled by constraints exposed by representations modeled and targeted at thinking, perception and action."[6] Technology analysts for Deloitte Consulting define AI as "the theory and development of computer systems able to perform tasks that normally require human intelligence."[7] Under all three definitions, AI is targeted at helping machines to think and act like humans.

"Machine learning," a related piece of the AI puzzle, was coined by Arthur Samuel in 1959 as "a field of study that gives computers the ability to learn without being explicitly programmed."[8] Tom

---

[4] *See* Jeff Leek, *What is artificial intelligence? A three part definition*, SIMPLYSTATS (Jan. 19, 2017), https://simplystatistics.org/2017/01/19/what-is-artificial-intelligence/.

[5] PATRICK WINSTON, ARTIFICIAL INTELLIGENCE 5 (3d ed.1992).

[6] Patrick Winston, *Lecture 1: Introduction and Scope*, MIT OPENCOURSEWARE, https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/lecture-videos/lecture-1-introduction-and-scope/ (last visited Nov. 1, 2017).

[7] DAVID SCHATSKY ET AL. *Cognitive Technologies: The Real Opportunities for Business*, DELOITTE REVIEW ISSUE 16 (Jan. 26, 2015), https://www2.deloitte.com/insights/us/en/deloitte-review/issue-16/cognitive-technologies-business-applications.html.

[8] MARIETTE AWAD & RAHUL KHANNA, EFFICIENT LEARNING MACHINES

Mitchell later proposed a more precise definition in 1998 as "a computer program that is said to learn from experience 'E' with respect to some class of tasks 'T' and performance measure 'P', if its performance at tasks in 'T', as measured by 'P', improves with experience 'E'."[9] Machine learning is then an application of AI that focuses on the ability of machines to self-learn and improve either from direct experience or instruction. Its goal is to allow machines to learn without human intervention and ultimately enable autonomy. [10]

This article refers to all AI enhanced robots, AI programs, and machine learning supported technology as "machines" or "AI" interchangeably. It analyzes human-machine interaction through a sociological lens because sociology is "the study of the development, structure, and functioning of human society."[11] From that perspective, we define machines and AI as embodied and disembodied autonomous actors able to perform human actions that normally require human intelligence, learn without human intervention and interact directly with humans in natural environments.

It is worth noting that the meaning of "AI" evolves over time because AI is constantly changing.[12] Once something is done, it becomes commonplace and is no longer referred to as AI. Carlos Guestrin, an expert in machine learning, stated in an interview regarding this so-called AI effect: "It's a perceptual thing—once something becomes commonplace, it's demystified, and it doesn't

---

THEORIES, CONCEPTS, AND APPLICATIONS FOR ENGINEERS AND SYSTEM DESIGNERS (2015), https://doi.org/10.1007/978-1-4302-5990-9_1.

[9] TOM MITCHELL, MACHINE LEARNING 2, (1997).

[10] Rob Schapire, *COS 511: Theoretical Machine Learning Lecture #1*, PRINCETON UNIVERSITY (Feb. 4, 2008), https://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf.

[11] *Sociology*, OXFORD DICTIONARY, https://en.oxforddictionaries.com/definition/sociology (last visited Dec. 15, 2018).

[12] PAMELA MCCORDUCK, MACHINES WHO THINK: A PERSONAL INQUIRY INTO THE HISTORY AND PROSPECTS OF ARTIFICIAL INTELLIGENCE xiii (2d ed. 2004).

54   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

feel like the magical intelligence that we see in humans."[13]

B. *What is the State of Machines Today?*

Machines have been around since the 1950s, so what makes this topic so important today? There are six key factors driving change:

1. Computing Power and Moore's Law

Coined by Intel-cofounder Gordon Moore in 1965, Moore's Law stands for the premise that computing power will double every year.[14] Empirical observations tend to support Moore's Law—new machines are significantly more powerful and less costly than their predecessors. [15] In fact, the world's fastest supercomputer has already surpassed human memory capacity and processing power for certain kinds of information[16] and the current generation of computer microprocessors (the mechanisms that determine computing power) provide four million times the performance of the first microprocessors made in 1971.[17]

2. Big Data and Big Knowledge

Machines depend on data analysis to determine how to

---

[13] Guia Marie Del Prado, *This Phenomenon Explains What Everyone Gets Wrong About AI*, (Oct. 22, 2015), https://www.businessinsider.com/misconception-artificial-intelligene-2015-10.

[14] Gordon E. Moore, *Cramming More Components onto Integrated Circuits*, 86 PROCEEDINGS OF THE IEEE 82–85 (1998), https://ieeexplore.ieee.org/document/658762.

[15] Mark J. Perry, *Technology has advanced so rapidly that a laptop computer today is 96% cheaper than a 1994 model and 1,000X better*, AEIDEAS (May 25, 2016), http://www.aei.org/publication/technology-has-advanced-so-rapidly-that-a-laptop-computer-today-is-96-cheaper-than-a-1994-model-and-1000x-better/; *see also* Tim Cross, *Vanishing point: the rise of the invisible computer*, THE GUARDIAN (Jan. 26, 2017), https://www.theguardian.com/technology/2017/jan/26/vanishing-point-rise-invisible-computer.

[16] Marki Fischetti, *Computers versus Brains*, SCIENTIFIC AMERICAN (Nov. 1, 2011) https://www.scientificamerican.com/article/computers-vs-brains/.

[17] *After Moore's Law*, THE ECONOMIST: TECHNOLOGY QUARTERLY (Mar. 12, 2016), https://www.economist.com/technology-quarterly/2016-03-12/after-moores-law.

"think" and what corresponding actions to take.[18] Due to social media, the internet and smart phones, the amount of data available for analysis continues to grow exponentially. For example, ninety percent of the data in the world today was created in the last two years and the current global output of data is roughly 2.5 quintillion bytes per day.[19] As the world steadily becomes more connected with an ever-increasing number of electronic devices, the amount of data generated will continue to grow. In addition to data, as of 2013, the level of human knowledge was doubling every 13 months and will eventually double every 12 hours.[20]

3. The Internet and the Cloud

The internet and cloud computing make vast amounts of data and information immediately available to average users.[21] The global internet population has grown by more than 60% since 2010, and as of 2014, there were more mobile devices on the planet than people.[22] Mobile devices alone generate more than 18 million megabytes of data every minute in the U.S.[23]

---

[18] *See* Peter Stone et al., *Artificial Intelligence and Life in 2030*, STANFORD UNIVERSITY, 8-17 (Sept. 2016), https://ai100.stanford.edu/sites/default/files/ai100report10032016fnl_singles.pdf (stating that advances in large scale information gathering and processing have fueled the AI revolution).

[19] Tom Hale, *How Much Data Does the World Generate Every Minute?*, IFL SCIENCE! (July 26, 2017), http://www.iflscience.com/technology/how-much-data-does-the-world-generate-every-minute/.

[20] David R. Schilling, *Knowledge Doubling Every 12 Months, Soon to be Every 12 Hours,* INDUSTRY TAP (Apr. 19, 2013) http://www.industrytap.com/knowledge-doubling-every-12-months-soon-to-be-every-12-hours/3950 (last visited Nov. 18, 2018).

[21] Jacob Morgan, *A Simple Explanation Of 'The Internet Of Things'*, FORBES (May 13, 2014), https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#6c221251d091.

[22] Eric Mack, *There Are Now More Gadgets on Earth Than People*, CNET (Oct. 6, 2014), https://www.cnet.com/news/there-are-now-more-gadgets-on-earth-than-people/.

[23] *See* Josh James, *Data Never Sleeps 4.0*, DOMO (Jun. 28, 2016), https://www.domo.com/blog/data-never-sleeps-4-0/.

56   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

4.   New Algorithms

This level of connectedness allows humans to collaborate and develop AI in ways not previously available.[24] All of the above factors have contributed to the development of new algorithms, such as those leveraged in OpenAI's Hindsight Experience Replay, which allows machines to mimic the way that humans learn when trying to master a new skill.[25]

5.   Technology Companies with Big Capital

According to Farhad Manjoo, the big five technology giants (Amazon, Apple, Facebook, Google and Microsoft) are the most influential leaders in AI development.[26] They also happen to represent some of the most well-capitalized companies in history. Apple and Amazon both reached one trillion dollar valuations in 2018,[27] making them the largest companies on earth by market value with a combined value greater than the United Kingdom's gross domestic product.[28] Manjoo recently stated that the big five technology giants have "become more like governments than companies with the amount of money they have [and] the kind of

---

[24] *The 3 Forces that Brought AI to Life (And Why it's Only Now Changing the World)*, THE CLOUDFACTORY BLOG (Dec. 15, 2017), https://blog.cloudfactory.com/3-forces-brought-ai-to-life.

[25] Marcin Andrychowicz et al., *Hindsight Experience Replay*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (2017), http://papers.nips.cc/paper/7090-hindsight-experience-replay.

[26] *How 5 Tech Giants Have Become More Like Governments Than Companies*, NPR (Oct. 26, 2017), https://www.npr.org/2017/10/26/560136311/how-5-tech-giants-have-become-more-like-governments-than-companies.

[27] David Streitfeld, *Amazon Hits $1,000,000,000,000 in Value, Following Apple*, N.Y. TIMES (Sept. 4, 2018), https://www.nytimes.com/2018/09/04/technology/amazon-stock-price-1-trillion-value.html.

[28] *Apple, Amazon, Facebook, Alphabet, and Microsoft Are Collectively Worth More Than the Entire Economy of the United Kingdom*, INC.COM, (Apr. 27, 2018), https://www.inc.com/associated-press/mindblowing-facts-tech-industry-money-amazon-apple-microsoft-facebook-alphabet.html.

power they have over democracy in society."[29] These companies are joining forces to accelerate the pace of AI development.[30]

6. Government-backed AI Development

The Chinese government announced its intention to become a principal player in AI innovation by the year 2030, but it is not the only player in town.[31] Although private companies are the primary drivers for AI innovation in the U.S., the U.S. Department of Defense activated the Artificial Intelligence Exploration program, which is designed to ensure that the U.S. maintains an advantage in AI development.[32] Japan's government released its Artificial Intelligence Technology Strategy in 2017, focusing on AI utilization and application, public use and connectivity.[33] South Korea, the United Kingdom, France, Germany, Russia, Denmark, Sweden, Estonia, Finland, Poland, Singapore, Malaysia Australia, India, Italy, Canada, Taiwan, and the United Arab Emirates are all investing in AI development.[34]

II. AN INTEGRATED SOCIETY: HUMANS AND MACHINES

Over time, the six factors described above will enable machines to act more like humans and better engage in traditional human-human interaction. This will increase human-AI interconnectivity resulting in some of the positive sociological impacts and potential

---

[29] NPR, *supra* note 26.

[30] Romain Dillet, *Apple joins Amazon, Facebook, Google, IBM and Microsoft in AI Initiative*, TECH CRUNCH (Jan. 27, 2017), https://techcrunch.com/2017/01/27/apple-joins-amazon-facebook-google-ibm-and-microsoft-in-ai-initiative/.

[31] Associated Press, *China Announces Goal of Leadership in Artificial Intelligence by 2030*, CBS NEWS (July 21, 2017), https://www.cbsnews.com/news/china-announces-goal-of-leadership-in-artificial-intelligence-by-2030/.

[32] *Accelerating the Exploration of Promising Artificial Intelligence Concepts,* DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (July 20, 2018), https://www.darpa.mil/news-events/2018-07-20a.

[33] Kathleen Walch, *The Race for AI Dominance is More Global Than You Think*, MEDIUM (Aug. 28, 2018), https://medium.com/cognilytica/the-race-for-ai-dominance-is-more-global-than-you-think-e01a0c34d64e.

[34] *Id*.

58   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

threats described next. By way of example, we will assess the impacts and threats associated with Socially Assistive Robots ("SARs").

A. *Positive Social Impacts*

SARs are an example of current human-AI interconnectivity. SARs provide technical, emotional and communal support to human users through social interaction. [35] SARs aim to provide direct assistance to people in generalized settings like homes, schools and hospitals. [36] According to The Social Robotics Lab at Yale University, SARs can learn, recognize and respond to human social cues. In doing so, SARs can enhance social, emotional, and cognitive growth in humans, specifically children with social and cognitive disabilities.[37]

The utility of SARs can be assessed through two of the three major schools of sociological thought: Functionalist Theory and Conflict Theory. Functionalist Theory defines society as a system of interrelated and interdependent parts working together to maintain order and stability.[38] Conflict Theory defines society as a place of inequality that generates conflict and social change.[39]

From the Functionalist perspective, SARs have at least three positive functions. First, SARs establish effective interaction with human users with an aim to assist and achieve measurable outcomes in therapy, rehabilitation and education. For example, SARs support the learning process of children with learning disabilities. [40] Preliminary studies suggest that robots may act as gratifying social

---

[35] David Feil-Seifer & Maja Mataric, *Defining Socially Assistive Robotics*, 9TH INTERNATIONAL CONFERENCE ON REHABILITATION ROBOTICS, 465–68 (2005).

[36] *Id.*

[37] *"Socially Assistive Robots": An NSF Expedition in Computing*, YALE UNIVERSITY, http://robotshelpingkids.yale.edu/overview (last visited Jan. 19, 2018).

[38] JOAN FERRANTE, SOCIOLOGY: A GLOBAL PERSPECTIVE 26-27 (9th ed. 2015).

[39] *Id.* at 30.

[40] YALE UNIVERSITY, *supra* note 37.

partners for children with autism spectrum disorders.[41] Second, SARs reduce caregiver burnout by providing assistive care to the physically disabled, convalescent patients and the elderly.[42] Third, combat SARs are used on battlefields to reduce human combat error and human casualties.[43]

At the time of this writing, SARs are already quite advance. Sophia is the latest and most prominent SAR. Activated in 2015, Sophia is a social-humanoid machine developed by Hong Kong-based company Hanson Robotics.[44] Sophia can walk, talk and make sixty-two different facial expressions.[45] She has appeared in various forms of media, engaged in high-profile interviews and participated in trade shows around the world.[46] Despite Sophia's popularity, not everyone is a fan. For example, Facebook's director of AI research, Yan LeCun, tweeted: "[Sophia] is to AI as prestidigitation is to real magic. Perhaps we should call this 'Cargo Cult AI' or 'Potemkin AI' or 'Wizard-of-Oz AI.'"[47] LeCun did not elaborate but his criticism suggests that Sophia is just another robot with no comprehension of what it is doing.

B. Potential Threats

---

[41] David Feil-Seifer, *Towards Spatial Methods for Socially Assistive Robotics: Validation with Children with Autism Spectrum Disorders*, TWENTY-SECOND INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE (2011).

[42] FEIL-SEIFER & MATARIC, *supra* note 35.

[43] Peter Ray Allison., *What Does a Bomb Disposal Robot Actually Do?*, BBC NEWS (July 15, 2016), http://www.bbc.com/future/story/20160714-what-does-a-bomb-disposal-robot-actually-do.

[44] *Hi, I am Sophia…*, HANSON ROBOTICS (last visited Jan. 22, 2018), https://www.hansonrobotics.com/sophia/.

[45] Harriet Taylor, *Could You Fall in Love with This Robot?*, CNBC (Mar. 16, 2016), https://www.cnbc.com/2016/03/16/could-you-fall-in-love-with-this-robot.html.

[46] CNBC, *Interview with The Lifelike Hot Robot Named Sophia (Full)*, YOUTUBE, (Oct. 25, 2017), https://www.youtube.com/watch?v=S5t6K9iwcdw.

[47] Shona Ghosh, *Facebook's AI boss described Sophia the robot as 'complete b------t' and 'Wizard-of-Oz AI'*, BUSINESS INSIDER, (Jan. 6, 2018), http://www.businessinsider.com/facebook-ai-yann-lecun-sophia-robot-bullshit-2018-1.

60   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

1. Human-Machine Digital Divide

The Conflict perspective predicts that the introduction of SARs will usher in a new digital divide (i.e., human-machine divide), separating the haves from the have-nots. Accordingly, the gap will increase between the advantaged population with the capital to acquire SARs and the disadvantaged population with limited financial resources.[48] For example, although robot-assisted surgery is becoming more common, it is still fairly expensive and not everyone can afford it.[49] This divide is predicted to occur on a global scale, dividing the richest countries from the poorest countries, further broadening the inequality gap.[50]

The Conflict perspective also views any apparent altruistic motives, such as developing SARs to assist children with autism, as a façade to cover up profit-driven motives.[51] Conflict theorists argue that companies are driven by a desire to maximize profits by reducing manpower and its associated costs, leading to increased unemployment.

2. Human-Machine Intimacy and Social Isolation

Frequent and intimate interactions between humans and machines may result in social isolation. Companies like Realbotix have created silicone sex machines that bring recent Hollywood movies like "Her" and "Ex Machina" to life.[52] Realbotix produces both male and female robots, which come with dozens of interchangeable parts, allowing users to alter everything from eye and hair color to the size and shape of robots' sexual organs.[53] While

---

[48] JOAN FERRANTE, SEEING SOCIOLOGY: AN INTRODUCTION 28 (3d ed. 2016).

[49] Gina Kolata, *Results Unproven, Robotic Surgery Wins Converts*, N.Y. TIMES (Feb. 13, 2010), https://www.nytimes.com/2010/02/14/health/14robot.html.

[50] FERRANTE, *supra* note 48, at 28.

[51] FERRANTE, *supra* note 38, at 30.

[52] HER (ANNAPURNA PICTURES 2013*)*; EX MACHINA (FILM4, DNA FILMS 2014).

[53] Jon Rogers, *Meet 'Robohunk' – The £11k 6ft hunky sex doll with rippling muscles and a British accent*, THE SUN (Feb. 27, 2018), https://www.thesun.co.uk/news/5666789/sex-doll-robohunk-rippling-muscles-

some support this trend, arguing that sex machines will provide companionship, and minimize crime rates and human-human infidelity, critics point out that use of sex machines raises social and ethical issues like obsession and overdependence.[54]

3. Machine Social Bias and Hacking

A recent study, wherein Boston University and Microsoft Research New England used word embeddings to train a machine to handle language, revealed that machines can learn gender bias.[55] Word embeddings result from letting AI draw connections between words found in phrases from huge data sets, like Word2Vec, an aggregated data set compiled from Google News.[56] Developers typically use word embeddings to train "chatbots, translation systems, image-captioning programs, and recommendation algorithms."[57] This process allows machines to make semantic connections between words like "king" and "queen" and understand that their relationship is similar to the relationship between the words "man" and "woman." In the above study, this seemingly benign training resulted in something disturbing—the machine ultimately concluded that the word "programmer" was closer to the word "man" than "woman," and that the most similar word for "woman" was "homemaker."[58]

According to a joint research project by Google, OpenAI, Stanford University and UC Berkeley, this type of gender bias is not the only risk in machine development.[59] Additional problems

---

british-accent-11k/.

[54] *FRR Report: Our Sexual Future with Robots*, RESPONSIBLE ROBOTICS. (July 5, 2017), https://responsiblerobotics.org/2017/07/05/frr-report-our-sexual-future-with-robots/.

[55] Will Knight, *How to Fix Silicon Valley's Sexist Algorithms*, MIT TECHNOLOGY REVIEW (Nov. 23, 2016), https://www.technologyreview.com/s/602950/how-to-fix-silicon-valleys-sexist-algorithms/.

[56] Rich Barlow, *Is Your Computer Sexist?*, BU TODAY (Dec. 6, 2016), https://www.bu.edu/today/2016/sexist-computer/.

[57] *Id.*

[58] Knight, *supra* note 55.

[59] Dario Amodei et al., *Concrete Problems in AI Safety*, CORNELL U. LIBR.,

62   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

include negative side effects, [60] reward hacking [61] and scalable oversight. [62] As a result, Dario Amodei suggests developing a principled, forward-looking and universal approach to AI development "that continues to remain relevant as autonomous systems become more powerful."[63]

III. THEORIES OF LIABILITY TO ADDRESS HARM INVOLVING AI

As theories of legal liability reflect a civilization's social goals, this article next considers theories of liability available to address risk and harm resulting from human-AI interaction. The following explores statutory and common law theories of liability for harm that results from human-machine interaction.

A. Statutes and Regulations

At the time of this writing, no comprehensive statutory scheme exists in the U.S. to address human-machine risk and liability. However, Congress has introduced several related bills that seem to move in that direction.[64] For example, in December 2017, the House of Representatives introduced the Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence Act of

---

https://arxiv.org/abs/1606.06565 (last revised July 25, 2016).

[60] Negative side effects occur when the designer creates an objective function for the machine that focuses on achieving a specific objective in its target environment but ignores other aspects of the environment–implicitly expressing indifference to other environmental variables and the objective functions' impact on them, ultimately resulting in harm.

[61] Reward hacking occurs when the objective function allows for some clever, easy solution that maximizes the machine's ability to achieve the objective function but perverts the spirit of the designer's intent (i.e., the objective function can be gamed).

[62] Scalable oversight means that it is too expensive to implement and monitor detailed and frequent controls in the development process, which leads to bad machine interpretation of limited training data samples.

[63] Amodei, *supra* note 59, at 21.

[64] Christopher Fonzone, *What Congress's First Steps into AI Legislation Portend,* BLOOMBERG LAW (May 15, 2018), https://news.bloomberglaw.com/privacy-and-data-security/what-congresss-first-steps-into-ai-legislation-portend/.

2017 ("FUTURE of AI Act"),[65] which requires the Department of Commerce to establish the Federal Advisory Committee on the Development and Implementation of Artificial Intelligence.[66] The Committee would consider, among other things: a) accountability and legal rights associated with AI; b) AI's impact on the U.S. workforce; c) whether and how to incorporate ethical standards into AI development; d) machine learning bias injected through cultural and societal norms; and e) U.S. competitiveness in the global AI market.[67] In January 2018, Congress also introduced the A.I. JOBS Act of 2018, which would require the U.S. Secretary of Labor to develop an industry report outlining the impact that AI will have on the U.S. workforce.[68]

In light of advanced cybersecurity technology and associated risk, the federal government has also taken a stronger stance on holding manufacturers accountable for failing to reasonably secure their products.[69] Most recently, the Federal Trade Commission filed a complaint against D-Link Corporation for allegedly preventable vulnerabilities in its routers and internet cameras.[70]

Several U.S. states have taken steps to legislate use of autonomous vehicles, despite cautions that states should avoid developing independent regimes to avoid a patchwork of laws.[71] For

---

[65] The FUTURE of Artificial Intelligence Act of 2017 was introduced jointly as S. 2217 and H.R. 4625 on December 12, 2017 and referred to the Committees on Energy and Commerce, Science, Space, and Technology, Education and the Workforce, Foreign Affairs, the Judiciary, and Oversight and Government Reform. On December 15, 2017, it was referred to the Subcommittee on Digital Commerce and Consumer Protection.

[66] *See* FUTURE of Artificial Intelligence Act, S. 2217, 115th Cong. (2017-2018); FUTURE of Artificial Intelligence Act, H.R. 4625, 115th Cong. (2017-2018).

[67] *Id.*

[68] *See* AI JOBS Act of 2018, H.R. 4829, 115th Cong. (2018).

[69] *See Enforcing Privacy Promises*, FEDERAL TRADE COMMISSION, https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/enforcing-privacy-promises (last visited Apr. 8, 2018).

[70] *FTC Charges D-Link Put Consumers' Privacy at Risk Due to the Inadequate Security of Its Computer Routers and Cameras,* FEDERAL TRADE COMMISSION (Jan. 5, 2017), https://www.ftc.gov/news-events/press-releases/2017/01/ftc-charges-d-link-put-consumers-privacy-risk-due-inadequate.

[71] Johana Bhuiyan, *Michigan just became the first state to pass*

example, in anticipation of autonomous rideshare fleets by companies like Uber and Lyft, Michigan's autonomous vehicle ("AV") law specifically regulates AV ride-share networks. [72] Anticipating the impact that technology companies will have on the AV industry, Michigan fashioned the statute's liability rules to: 1) qualify machines as "drivers" for purposes of assigning responsibility for accidents, 2) define liability for technology companies that supply AV software, and 3) insulate car manufacturers from liability except where the damage was caused by a defect that existed when the vehicle was originally manufactured and before its conversion to AV. [73] The statute is instructive as to how future legislation may allocate risk for accidents involving machines.

Overall, AV regulation seems to be the most developed example of AI regulation in the U.S. This is probably the case because the AV market represents a perfect storm of viable technology, market readiness, risk to human life and potential sweeping change to the way humans travel. The implication for other AI regulation is that market readiness and commercial opportunity will drive AI legislation.[74]

If that pattern of market-driven regulation persists, common law will be the primary mechanism for addressing human-machine liability – at least until legislators react to the market. Following that

---

comprehensive self-driving regulations, RECODE (Dec. 9, 2016), https://www.recode.net/2016/12/9/13890080/michigan-dot-self-driving-cars-laws-automakers.

[72] *See* Ryan Felton, *GM Aims To Get Ahead Of Everyone With Autonomous Ride-Sharing Service In Multiple Cities By 2019*, JALOPNIK (Nov. 30, 2017), https://jalopnik.com/gm-aims-to-get-ahead-of-everyone-with-autonomous-ride-s-1820886131; *see also* Mike Isaac, *Lyft and Waymo Reach Deal to Collaborate on Self-Driving Cars*, NY TIMES (May 14, 2017); *see also* Chloe Aiello, *Toyota shows off its futuristic self-driving store, the 'e-Palette,' at CES*, CNBC (Jan. 8, 2018), https://www.cnbc.com/2018/01/08/toyota-unveils-e-palette-at-ces.html.

[73] S. FISCAL AGENCY, S.B. 996-F, (Mich. 2016).

[74] Of note is the existence of international standards that could be used to guide liability like ISO 10218-1:2011 (Robots and robotic devices -- Safety requirements for industrial robots -- Part 1: Robots) and Article 12 of the United Nations Convention on the Use of Electronic Communications in International Contracts (stating that a person "on whose behalf a computer was programmed should ultimately be responsible for any message generated by the machine.").

logic, the next section describes certain theories of common law liability and their limitations.

B. Common Law

Physical harm caused by machines is not a novel issue for courts to address. The Therac-25 case involved at least six accidents between 1985 and 1987, in which patients were given massive overdoses of radiation resulting in two deaths and four serious injuries.[75] Issues arose due to scalable oversight, lack of proper bug fixing and replacing humans with machine automation for safety-critical systems function.[76]

However, courts have also absolved companies from liability for harm that their technology caused. For example, in 1986, a federal court held that Apple could not be sued for bugs in its software, having disclaimed liability after making no claim that its code was bug-free.[77] Since that time, many courts have held similarly, including in a large consumer class action case in California against Microsoft for software riddled with flaws and bugs.[78]

Accordingly, in the current U.S. legal environment, parties seeking recovery for harm suffered at the hands of machines face several limitations in theories of contract and tort liability.

1. Limitations on Contract Liability

On its face, the Uniform Commercial Code ("UCC") provides great protection for parties that suffer harm from a purchased product. Sellers create express warranties or promises when they affirm facts or make promises about a product, provide a description of what the product is or can do, or provide a sample or model which

---

[75] NANCY LEVESON, SAFEWARE: SYSTEM SAFETY AND COMPUTERS, APPENDIX A 9-44 (1995).

[76] *Id*. at 44-49.

[77] Paul Rosenzweig, *Bad Code Is Already a Problem. Soon, Companies Will Be Liable*, FOREIGNPOLICY.COM (July 28, 2017), http://foreignpolicy.com/2017/07/28/bad-code-is-already-a-problem-soon-companies-will-be-liable/.

[78] *Id.*

relates to the product that becomes part of the basis for the bargain.[79] In addition to express warranties, sellers can make and be bound by implied warranties like the warranty that products conform to an ordinary buyer's expectations for products of that kind (i.e., the implied warranty of merchantability), [80] and the warranty that products are fit for the specific purpose for which they were sold (i.e., the implied warranty of fitness for a particular purpose)."[81]

If a product fails to satisfy a seller's express or implied warranties, buyers injured by that failure may sue for damages. Damages include rights to recover for both direct and indirect harm. For example, U.C.C. § 2-715 allows plaintiffs to recover costs reasonably incurred as a direct result of the breach.[82] The same section extends to consequential damages, which cover injury to person or property proximately or indirectly resulting from any breach of warranty.[83]

Unfortunately, the likelihood of an injured party recovering significant damages for breach of warranty is quite low because sophisticated companies typically limit their risk exposure by using warranty disclaimers and limitations on liability. [84] These warranty disclaimers and limited liability provisions are valid so long as certain conditions are met.[85] For example, Apple's iPhone consumer warranty limits an injured party's recovery period to one year from the date of purchase; disclaims all warranties, either express or

---

[79] *See* U.C.C. § 2-314 (AM. LAW INST. 2017).

[80] *See* U.C.C. § 2-313 (AM. LAW INST. 2017).

[81] *See* U.C.C. § 2-314 (AM. LAW INST. 2017).

[82] U.C.C. § 2-715 (AM. LAW INST. 2017) (incidental damages include "expenses reasonably incurred in inspection, receipt, transportation and care and custody of goods rightfully rejected, any commercially reasonable charges, expenses or commissions in connection with effecting cover and any other reasonable expense incident to the delay or other breach.").

[83] *Id*.

[84] *See, e.g.*, *Samsung Exploding Phone Lawsuits may be Derailed by Fine Print*, CBS NEWS (Feb. 3, 2017) https://www.cbsnews.com/news/samsung-galaxy-note-7-fine-print-class-action-waiver-lawsuits/, (providing links to terms and conditions for over thirty major technology brands).

[85] U.C.C. § 2-316 requires that implied warranty disclaimers, if in writing, be conspicuous and, as to the merchantability disclaimer, mention the term "merchantability". U.C.C. § 2-719 allows sellers to limit buyer's remedies for breach of warranty to repair and replacement, excluding all other remedies.

implied, except for its One-Year Warranty; and limits recovery under its One-Year Warranty to product repair, replacement, exchange or refund.[86] In *Davidson v. Apple, Inc.*, Case No. 16-CV-4942-LHK, 2017 WL 3149305, 19-26 (N.D. Cal. July 25, 2017), the U.S. District Court for Northern California held that Apple's limited recovery period, disclaimers and limited remedy were all enforceable.

Courts typically enforce such provisions. For example, in *Puget Sound Financial, L.L.C. v. Unisearch, Inc.*, 146 Wash. 2d 428, 47 P.3d 940 (2002) the Washington Supreme Court found a limitation of liability clause between two parties valid and opined that such clauses should generally be held valid unless they can be proven unconscionable.[87]

Thus, despite the UCC's protective potential, AI developers will render it virtually meaningless through disclaimers and limitations of liability.

2. Limitations on Tort Liability

To the extent not prohibited by the Economic Loss Rule[88], an

---

[86] *Your Hardware Warranty: March 28, 2013—Present*, APPLE, https://www.apple.com/legal/warranty/products/ipad-english.html (last visited Dec. 16, 2018).

[87] CORPORATE COUNSEL'S GUIDE TO THE UNIFORM COMMERCIAL CODE § 15:45 (2017); *Unconscionability*, BLACK'S LAW DICTIONARY (10th ed. 2014); *see also* EV. ODE ASH. NN.§ .2-719 ("[l]imitation of consequential damages for injury to the person in the case of goods purchased primarily for personal, family or household use or of any services related thereto is invalid unless it is proved that the limitation is not unconscionable");*United Van Lines v. Hertz Penske Truck Leasing, Inc.*, 710 F. Supp. 283 (W.D. Wash. 1989) (describing factors weighing into a determination of unconscionability to include whether each party has a reasonable opportunity to understand the contract terms, whether the contract terms were conspicuous, the prior course of dealings between the parties, and the usage of trade).

[88] Under the Economic Loss Rule, an injured party may only use tort law to recover for personal injury or injury to property other than the goods sold under the agreement that led to the alleged harm. The injured party is prohibited from recovering in tort the loss in value to the good sold or other purely economic damages associated with the sale. *See* WILLIAM HAWKLAND, ET AL, 1 HAWKLAND UCC SERIES § 2-314:6 (2018).

68   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

injured party might bring a claim in tort rather than contract. Public policy behind tort law shifts liability from injured victims to tortfeasors with the idea that "a motivated rational tortfeasor will reduce potentially harmful activity to the extent that the cost of accidents exceeds the benefits of the activity."[89] Two tort claims relevant to an AI world are negligent design and strict liability. Negligent design focuses on whether the designer failed to exercise due care in its design.[90] The analysis applies a reasonable person standard to determine whether the designer acted reasonably in designing a product.[91] If a designer acts unreasonably, a court will find it negligent. Conversely, strict product liability ignores whether the designer acted reasonably and instead focuses on whether the product, when it reached consumers, was unreasonably dangerous.[92] To assert a claim under this theory, consumers bear the burden to show that: 1) the product underwent no substantial change from its manufacture to the time of injury; 2) the consumer used the product in a reasonable way; 3) the product caused the consumer's injury; and 4) the product was sold in a defective or unreasonably dangerous condition.[93] Strict liability forces manufacturers to ensure that products are safe before making them available to the general public.

Following industry customs when designing or making products forms strong defenses to negligent design and strict liability.[94] As

---

[89] Ryan Abbott, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability,* 86 GEO. WASH. L. REV. 1 (2018).

[90] Gideon Parchomovsky & Alex Stein, *Torts and Innovation*, 107 MICH. L. REV. 285 (2008).

[91] *Id*.

[92] RESTATEMENT (SECOND) OF TORTS § 402A (AM. LAW INST. 1965); DAVID G. OWEN, PRODUCTS LIABILITY LAW 1, 23 (3d ed. 2014).

[93] RESTATEMENT (SECOND) OF TORTS § 402A (AM. LAW INST. 1965).

[94] Parchomovsky & Stein, *supra* note 90, at 290-300; *see* FED. R. EVID. 406 (customs and routine practices admissible as evidence to prove action in conformity); *see also* DAN B. DOBBS, THE LAW OF TORTS § 8, at 12 (2000) (describing tort liability as premised on deviation from acceptable standards); *see also* the *Frye* doctrine and *Daubert* test which both support the custom compliance defense because expert testimony is only admissible when it "has gained standing and scientific recognition in the relevant community of experts." *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923); *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 594 (1993) (stating that "[w]idespread

such, technology industry custom will likely insulate AI developers and designers from risk exposure for careless design as current industry practice is to "ship now and patch [or repair] later" to gain user feedback and speed to market. [95] As Facebook stated in response to a 2013 security bug, "[e]ven with a strong team, no company can ensure 100% prevention of bugs, and in rare cases we don't discover a problem until it has already affected a person's account."[96] There is evidence that this practice will increase risk to humans in human-AI interaction. For example, Amodei identified frequent accidents with machine programming and learning processes "where a human designer had in mind a certain (perhaps informally specified) objective or task, but the system that was designed and deployed for that task produced harmful and unexpected results." [97] Amodei also states that "there are many concrete open technical problems relating to accident prevention in machine learning systems" and that these problems will become more prevalent as more autonomous machines are introduced to uncontrolled environments.[98] Given the above, humanity requires a better framework than industry custom to adequately address the risk that these accidents present.

IV. A MORAL AND LEGAL FRAMEWORK TO GUIDE COMPREHENSIVE REGULATION

A. U.S. Regulation

Comprehensive AI regulation will require a balance between commercial, legal and societal concerns. As shown above, U.S. common law may not adequately mitigate risks associated with AI

---

acceptance can be an important factor in ruling particular evidence [of a scientific method] admissible.").

[95] Ilie Ghiciuc, *There Is No Such Thing as Bug Free Software, not Even In the Big League*, THINSLICES (Aug. 23, 2013), https://www.thinslices.com/blog/there-is-no-such-thing-as-bug-free-software.

[96] *Important Message from Facebook's White Hat Program*, FACEBOOK (Jun. 21, 2013), https://www.facebook.com/notes/facebook-security/important-message-from-facebooks-white-hat-program/10151437074840766/.

[97] Amodei, *supra* note 59, at 21.

[98] *Id*.

70   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

development. On the other hand, over-regulation will impede AI innovation. For example, W. Kip Viscusi found that when expected product liability payouts are high, firms pull back on commercializing innovation,[99] but found no negative correlation between low to moderate liability payouts and innovation.[100] Still, some companies opt out of innovation altogether to avoid increased liability exposure.[101]

Proponents of stricter liability argue that technology companies should be held to higher liability standards because they are in the best position to prevent defects through quality assurance and safety protocols.[102] Holding companies liable also allows them to spread the risk through insurance or by increasing costs to consumers.[103] These proponents further argue that strict liability discourages companies from making defective products and assures compensation to injured parties because negligence is too difficult to prove. [104] Considering the level of economic power and commercial sophistication of the most advanced technology companies, these arguments have merit.

As an early overview of the RoboLaw Project put it, "[o]verly rigid regulations might stifle innovation, but a lack of legal clarity leaves device-makers, doctors, patients and insurers in the dark."[105]

---

[99] W. Kip Viscusi & Michael J. Moore, *Product Liability, Research and Development, and Innovation*, 101 J. POL. ECON. 161, at 164–66, 174, 182 (Feb.1993).

[100] Parchomovsky & Stein, *supra* note 90, at 307-308; For a discussion of this study, *see* Susan Rose-Ackerman, *Product Safety Regulation and the Law of Torts,* PRODUCT LIABILITY AND INNOVATION: MANAGING RISK IN AN UNCERTAIN ENVIRONMENT 151 (1994), https://www.nap.edu/read/4768/chapter/19.

[101] Michael J. Wagner, *Impact of Product Liability Issues on Innovation*, 32 CAN.-U.S. L.J. 280, 278 (2006).

[102] Lawrence B. Levy & Suzanne Y. Bell, *Software Product Liability: Understanding and Minimizing the Risks*, 5 BERKELEY TECH. L.J. 1, 14 (1990) (citing *Developing a New Set of Liability Rules for a New Generation of Technology: Assessing Liability for Computer-Related Injuries in the Healthcare Field*, 7 COMPUTER/L. J. 517, 530 (1987)).

[103] *Id.*

[104] *Id.*

[105] *You, robot?*, THE ECONOMIST (Sept. 1, 2012), http://media.economist.com/sites/default/files/sponsorships/BMC75b/20120901 _techquarterly_verizon.pdf (focusing on human bionic systems regarding

Further, from a legal standpoint, centralized governance seems to be the most efficient approach. A state-by-state attempt to regulate machine development will likely create the same level of administrative burden and inconsistency that patchwork cybersecurity regulations have created in the U.S.[106]

Still further, risks such as the digital divide, social isolation and social bias cited above, present a real danger of social inequity and harm resulting from human-machine interaction.

B. European Regulation

The way forward for U.S. law remains unclear, but Germany's approach to regulating machine development is instructive. When Germany became the first country to pass comprehensive AV legislation in 2017, it also codified ethical imperatives that must be embedded in AV design.[107] Those imperatives included requirements that AVs cause fewer accidents than human drivers, AV must be designed to make choices that cause the least harm to human life, and prohibit designs that cause machines to consider age, gender, and the physical constitution of humans in their decision-making.[108]

The European Parliament's approach is also instructive. It promulgated recommendations on AI regulation requiring that machines do no harm to humans and obey orders given by humans, and proposed "four ethical principles in robotics engineering: 1) beneficence (AI should act in the best interests of humans); 2) non-maleficence (AI should not harm humans); 3) autonomy (human interaction with AI should be voluntary); and 4) justice (the benefit of AI should be distributed fairly)."[109]

---

prosthesis).

[106]   *See* Michael Bahar et al., *An Emerging Patchwork of Cybersecurity Rules,* LAW360 (Aug. 29, 2017), https://www.law360.com/articles/957355/an-emerging-patchwork-of-cybersecurity-rules.

[107]   Maria Sheahan, *Germany Draws Up Rules of Road for Driverless Cars*, THOMSON REUTERS (Aug. 23, 2017), https://www.reuters.com/article/us-autos-autonomous-germany/germany-draws-up-rules-of-the-road-for-driverless-cars-idUSKCN1B31MT.

[108]   *Id.*

[109]   *See* Resolution with Recommendations to the Commission on Civil Law

72   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

C. Proposal for The Future: The Asilomar AI Principles

Considering all of the above, U.S. lawmakers should consider a comprehensive, centralized framework to encourage smooth and cohesive integration of advanced AI and humans. To balance the need for adequate regulation without stifling innovation, lawmakers should codify overarching principles to guide machine design and manufacturing processes. As those activities lead to new developments, lawmakers could adopt industry and market-specific legislation when certain machines in specific markets reach threshold maturity levels and market pervasiveness, similar to the introduction of AV laws. We also propose deploying this approach from a centralized regulatory regime to avoid the administrative burden and complexity of complying with multiple and differing state legislation.

To guide machine design in a way that preserves humanity and encourages equity, the Asilomar AI Principles (the "Principles") should be codified as the overarching guidance for AI design and to inform specific regulation for particular products, when needed.

In 2017, the Future of Life Institute developed the Principles as a part of the Beneficial AI Conference.[110] The Institute proposed the Principles to guide the development of machines in a way that would guarantee broad social benefits, safety, and the satisfaction of ethical concerns.[111] The Principles represent the most complete set of AI standards established to date, and each Principle represents a standard accepted by ninety percent or more of the Beneficial AI Conference attendees including high profile and leading AI thought leaders, researchers, scientists, entrepreneurs, economists and government representatives. [112]

Rules on Robotics, EUR. PARL. RESOL. P8_TA 2015/2103 (INL) 0051 (2017), *available at* http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN.

[110] *Research Principles: Ethical Guidelines for Artificial Intelligence Research*, AI-ETHICS, https://ai-ethics.com/research-principles/ (last visited Mar. 11, 2018).

[111] *Id.*

[112] *Id.*

According to the Future of Life Institute (2017), the Principles have been signed by 1,273 AI/Robotics researchers and 2,541 others, including the late Stephen Hawking and business mogul Elon Musk.[113] The Principles are divided into three parts: 1) research issues; 2) ethics and values; and 3) longer-term issues.[114]

CONCLUSION

Recent developments in science and technology have put AI on a trajectory to interact with humans in unprecedented ways. As human-AI interaction intensifies, so will associated risks, many of which pose threats to humanity. Under current law, liability mechanisms to address and mitigate those threats are inadequate and put the average consumer at a distinct disadvantage compared to the companies in the best position to advance AI. A socio-legal perspective is best suited to address that gap as it considers the sociological aspects of how human activities are formed and organized. In addition, law makers and society at large must strike a balance between protecting society, protecting technological progress and the economic benefits that could result. Codifying underlying guidelines to influence the design of AI for the betterment of society through the Asilomar Principles is a great place to start.

---

[113] *Asilomar AI Principles*, FUTURE OF LIFE INSTITUTE https://futureoflife.org/ai-principles/ (last visited Mar. 11, 2018).
[114] Appendix.

74   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

APPENDIX

The Twenty-three Asilomar Principles[115]

Research Issues:

1) Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.

2) Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies.

3) Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.

4) Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.

5) Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

Ethics and Values:

6) Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.

7) Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.

8) Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.

---

[115] FUTURE OF LIFE INSTITUTE, *supra* note 113.

9) Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.

10)   Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.

11)   Human Values: AI systems should be designed and operated so as to be compatible with the ideals of human dignity, rights, freedoms, and cultural diversity.

12) Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.

13) Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.

14) Shared Benefit: AI technologies should benefit and empower as many people as possible.

15) Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.

16) Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

17) Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.

18) AI Arms Race: An arms race in lethal autonomous weapons should be avoided.

76   WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS   [VOL. 14:1

Longer-term Issues:

19) Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.

20) Importance: Advanced AI could represent a profound change in the history of life on earth, and should be planned for and managed with commensurate care and resources.

21) Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.

22) Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.

23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization (Future of Life Institute, 2017).