July 18, 2024

Blurring the Line Between Human and Machine Minds: Is U.S. Law Ready for Artificial Intelligence?

Kipp Coddington, Esq. & Saman Aryana[¥]

ABSTRACT

This Essay discusses whether U.S. law is ready for artificial intelligence ("AI") which is headed down the road of blurring the line between human and machine minds. Perhaps the most high-profile and recent examples of AI are Large Language Models ("LLMs") such as ChatGPT and Google Gemini that can generate written text, reason and analyze in a manner that seems to mimic human capabilities. Until the recent release of these LLMs, many deemed written language to be a capability possessed by humans alone.[1] Noted computer scientist and physicist Stephen Wolfram puts it well:

> *So how is it, then, that something like ChatGPT can get as far as it does with language? The basic answer, I think, is that language is at a fundamental level somehow simpler than it seems. And this means that ChatGPT—even with its ultimately straightforward neural net structure—is successfully able to "capture the essence" of human language and the thinking behind it. And moreover, in its training, ChatGPT has somehow "implicitly discovered" whatever regularities in language (and thinking) make this possible.* [2]

The stunning developments in AI are not limited to written communications, as advancements are occurring on related fronts such as machine consciousness.

U.S. law is based on English common law, which in turn incorporates Christian principles that assume the dominance and uniqueness of humankind.[3] U.S. law assumes human communication skills

---

[¥] Kipp Coddington, Esq. is Professor of Practice, College of Law, University of Wyoming and corresponding author (kcodding@uwyo.edu). Dr. Saman Aryana is Professor and Occidental Chair for Energy and Environmental Technologies, Chemical and Biomedical Engineering, College of Engineering and Physical Sciences, University of Wyoming; saryana@uwyo.edu.

The views and opinions expressed herein are solely those of the authors. Google Gemini (trademark application filed: https://trademarks.justia.com/982/02/gemini-98202646.html) assisted with research but no LLM was used to generate text for this Essay. The authors gratefully acknowledge the helpful comments provided by Catherine Hartmann, Assistant Professor of Religious Studies, Department of Phil. & Religious Studies, College of Arts and Sciences, University of Wyoming; Tristan Fross, J.D. 2024, College of Law, University of Wyoming; and Jesse Nelson, 2L, Defender Aid Clinic, College of Law, University of Wyoming.

[1] William Curtis, *Human Language Demands a Creator*, 1990 Proc. of Int'l Conf. on Creationism 69, https://digitalcommons.cedarville.edu/cgi/viewcontent.cgi?article=1386&context=icc_proceedings.

[2] Stephen Wolfram, *What is ChatGPT Doing … and Why Does It Work?*, Stephen Wolfram Writings (Feb. 14, 2023), https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

[3] *Genesis* 1:28 ("God blessed [male and female] and said to them, 'Be fruitful and increase in number; fill the Earth and subdue it. Rule over the fish in the sea and the birds in the sky and over every living creative that moves on the ground"); *God-Given Intelligence*, The Christian Science Monitor (Nov. 15, 1989), https://www.csmonitor.com/1989/1115/mrc477.html; Kate Lucky, *AI Will Shape Your Soul*,

are accompanied by attributes such as consciousness and Free Will ("FW") [4] *that, in turn, underpin critical legal concepts such as mens rea – i.e., intent. Philosophers and others generally agree that consciousness is necessary for FW.*[5] *On the assumption that human beings possess consciousness that supports FW, the law thereafter deems human beings capable of acting with legal consequences, from entering into contracts to committing crimes. For corporations, which are also capable of acting with legal consequences, the law imputes human mens rea to the inanimate entity under a variety of legal theories.*

The stunning arrival on the scene of machines that write and reason like humans is going to rock the anthropogenic foundations of U.S. law. Complicated legal scenarios will arise that attorneys and judges will have to resolve – e.g., was the defendant unduly influenced by what the machine was telling her to do? While we agree that "one must take care not to imply human-like cognition [to AI systems] given the way current AI models work," we come down on the side of those who believe that many in the legal profession are underestimating the impact these technologies will have both on the law and society at large.[6]

Although legislators around the world are beginning to grapple with some of these topics, AI remains largely unregulated with many foundational considerations unexamined. U.S. law historically has proven to be adaptive to new technologies and scientific advancements in fields such as mental health, psychology and neuroscience due to reasonably flexible rules regarding expert witness testimony and the admission of evidence. At the State level, which underpins criminal law, judges have a fair amount of autonomy to extend the law in response to new information and conditions. It is possible – maybe even likely – that U.S. law in the short-term will continue to evolve to accommodate AI much as it has historically done in response to emerging technologies, scientific breakthroughs, and engineering advancements. The answer is less clear in the long term as AI technologies continue to improve, thereby further blurring the distinction between human and machine intelligence. While comfort may be drawn from the past adaptability of the law, the uniqueness of AI may compel the conclusion that "it may be different this time."

With a focus on LLMs, the Essay suggests that U.S. law may struggle to respond to AI because the technology disrupts the law's assumptions regarding the uniqueness of human traits and abilities such as consciousness, FW, written communications and reasoning.[7]

Christianity Today (Sept. 11, 2023), https://www.christianitytoday.com/ct/2023/october/artificial-intelligence-robots-soul-formation.html.

[4] FW is the "canonical designator for a significant kind of *control* over one's actions." *Free Will*, Stanford Encyclopedia of Phil. (Nov. 3, 2022), https://plato.stanford.edu/entries/freewill/.

[5] Eddy Nahmias et al., *When Do Robots Have Free Will? Exploring the Relationships between (Attributions of) Consciousness and Free Will* (July 2019), https://philarchive.org/archive/NAHWDR.

[6] Harry Surden, *ChatGPT, AI Large Language Models, and Law*, 92 Fordham L. Rev. 1941, 1942 n. 4 (2024).

[7] Ryan Tracy & Isaac Yu, *Some of the Thorniest Questions About AI Will Be Answered in Court*, Wall St. J. (Aug. 23, 2023), https://www.wsj.com/tech/ai/some-of-the-thorniest-questions-about-ai-will-be-answered-in-court-e7fd444b; Christopher Mims, *The AI Industry Is Streaming Toward A Legal Iceberg*, Wall St. J. (Mar. 29, 2024), https://www.wsj.com/tech/ai/the-ai-industry-is-steaming-toward-a-legal-iceberg-5d9a6ac1; Dorothy Atkins, *'Much More is Coming': Experts See Wave of AI-Related Suits*, Law360 (Apr. 12, 2024), https://www.law360.com/pulse/articles/1822380.

Table of Contents

INTRODUCTION

The growing capabilities of AI technology are astonishing on multiple fronts. This Essay focuses on LLMs because of their human-like ability to generate written text coupled with the singular importance of language to our species – a skill that, in turn, is based on neural functions and attributes, including consciousness.[8] Human reliance on AI for decision-making is increasing with no end in sight:

> In recent times … AI … has made great advances in its ability to mimic human problem-solving and decision-making skills. Certain examples currently include applications in law firms, detecting financial fraud and making business decisions. AI has been increasingly helping humans make decisions and, in some cases, make decisions for humans. It also has been shown that AI is superior to humans in certain decision-making, like those relevant to the stock market.
>
> In the future and arguably in the present, there may be additional advances in AI technology that will allow people to utilize the decision-making abilities of AI for personal use. This AI could use data available in an increasing connected world to make decisions based on enormous amounts of training data gleaned from the internet.[9]

Much has been written about the impact of AI on society, to include employment. An equal amount of ink has been spilled on the topic of regulation of AI. In March 2024, the European Union's ("EU") parliament approved the first-of-its-kind "AI Act" that "aims to ensure that AI systems placed on the European market and used in the EU are safe and respect fundamental rights and EU values."[10]

---

[8] Mark Pagel, *Q&A: What is human language, when did it evolve and why should we care?* BMC Biology (2017) 15:64, DOI 10.1186/s12915-017-0405-3, https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-017-0405-3. LLMs are not flawless but they are continuing to improve.

[9] Eric Luo, *The Effect of Artificial Intelligence on the Human Idea of Free Will*, Proc. of 2021 Int'l Conf. on Soc. Dev. & Media Commc'n (SDMC 2021), https://www.atlantis-press.com/proceedings/sdmc-21/125968537. *See also* Lucia Vicente et al., *Humans inherit artificial intelligence biases*, Sci. Rep., (Oct. 3, 2023) 13:15737, https://doi.org/10.1038/s41598-023-42384-8 ("Over the last decades, the number of tools based on artificial intelligence … designed to assist human decisions has increased in many professional fields such as justice, personnel selection and healthcare") (footnotes omitted). *Accord* Daniel Solove & Hideyuki Matsumi, *AI, Algorithms, and Awful Humans*, 92 Fordham L. Rev. 1923, 1924 (2024) ("We are witnessing a profound shift in the way that many decisions are made, with machines taking on greater roles in the decision-making process").

[10] *Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world*, Council of the EU Press Release (Dec. 9, 2023), https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/#:~:text=The%20draft%20regulation%20aims%20to,huge%20milestone%20towards%20the%20future!.; *World's first major act to regulate AI passed by European lawmakers* (Mar. 13, 2023), https://www.cnbc.com/2024/03/13/european-lawmakers-endorse-worlds-first-major-act-to-regulate-ai.html.

Less has been written about whether the law itself is prepared for AI as the technology blurs the line between human and machine intelligence. Human reliance on AI for decision-making goes to the heart of many foundational legal considerations, issues from "meeting of the minds" in contracts to intent under criminal law. Because they function in a realm that could be deemed cognitive, LLMs have little in common with prior technologies to which the law successfully adapted. Chief Justice Roberts of the U.S. Supreme Court recently seemed to describe AI as a mere tool that will affect how legal professionals do their jobs; he seemed to view the technology as just the latest version of computers or word processers.[11]

We do not believe that AI is a "mere tool." AI technologies such as LLMs are tools all right, but ones that undermine the law's anthropogenic assumptions regarding cognitive functions and attributes such as consciousness, FW, written communications and reasoning. This Essay explores whether the assumptions underpinning U.S. law are themselves undermined by AI. The Essay first provides background information (Section I) on: (1) AI system by type and functionality (Section 1.A.1), and (2) what the literature says about interactions between Homo sapiens and machines (Section 1.A.2). The Essay next (Section II) discusses the philosophical foundation of U.S. law, to include the uniqueness of human mental states. Section III explain why AI may undermine that premise that human mental states are unique.

I. BACKGROUND

A. Overview of AI

1. Types and Functionality

AI is "the science and engineering of making intelligent machines."[12] Four categories of AI have been described, only two of which – reactive machines and limited memory machines – are currently demonstrated and available commercially.[13] The other two – theory of mind ("ToM") and self-awareness/consciousness machines – remain theoretical. If past is prologue, predictions about the future capability of AI systems are generally wrong.[14] So-called "superintelligence" may never arrive.[15] This Essay is not dependent upon fanciful predictions of

[11] Chief Justice John Roberts, *2023 Year-End Report on the Federal Judiciary* (Dec. 31, 2023), https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf.
[12] AI is defined as "the science and engineering of making intelligent machines." *Artificial Intelligence Definitions*, Stanford University Human-Centered Artificial Intelligence, https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf (last visited Dec. 9, 2023).
[13] *4 Types of AI: Getting to Know Artificial Intelligence*, Coursera, https://www.coursera.org/articles/types-of-ai (last visited Dec. 11, 2023).
[14] *See* Michio Kaku, *The Future of the Mind*, at 216 (2014) ("It is difficult to foretell the fate of AI, since it has gone through three cycles of boom and bust. Back in the 1950s, it seemed as if mechanical maids and butlers were just around the corner"); Cem Dilmegani, *When will singularity happen? 1700 expert opinions of AGI*, aimultiple.com (updated Nov. 28, 2023), https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/ (quoting AI pioneer Herbert Simon in 1965: "machines will be capable, within twenty years, of doing any work a man can do").
[15] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (2016).

how AI systems will perform in the future, although the technology does continue to improve in real time. We instead make observations based upon AI as it exists today.

The four categories of AI are: (1) Reactive Machines; (2) Limited Memory Machines (e.g., LLMs); (3) ToM Machines; and (4) Self-Aware/Conscious Machines.

a. Reactive Machines

These AI systems, which have existed commercially for decades, are task-specific, lack memory, and perform the identified task faster and better than humans. They are backward looking, not forward-looking, and cannot make future forecasts. An example of a reactive machine is one that plays chess. As impressive as they are, reactive machines by themselves are not particularly interesting because they fail to raise the types of cognitive considerations that will challenge the law.

b. Limited Memory Machines (e.g., LLMs)

These AI systems, which also exist commercially, utilize machine learning and deep-learning algorithms that "train[] computers to process information in a way that mimics human neural processes."[16]

LLMs are Limited Memory Machines.[17] A LLM is a:

> type of language model notable for its ability to achieve general-purpose language understanding and generation. LLMs acquire these abilities by using massive amounts of data to learn billions of parameters during training and consuming large computational resources during their training and operation. LLMs are artificial neural networks (mainly transformers) and are (pre-)trained using self-supervised learning and semi-supervised learning.
>
> As autoregressive language models, they work by taking an input text and repeatedly predicting the next … word. Up to 2020, fine tuning was the only way a model could be adapted to be able to accomplish specific tasks. Larger sized models, such as GPT-3, however, can be prompt-engineered to achieve similar results. They are thought to acquire knowledge about syntax, semantics and "ontology" inherent in human language corpora, but also inaccuracies and biases present in the corpora.[18]

LLMs utilize deep-learning algorithms that are capable of performing natural language processing ("NLP") tasks. LLMs are able to recognize, translate, predict or generate text or other content, including audio, computer code, simulations and video. The term NLP captures

---

[16] *What is Deep Learning? Definition, Examples, and Careers*, Coursera, https://www.coursera.org/articles/what-is-deep-learning (last visited Dec. 11, 2023).

[17] For an excellent explanation of what LLMs are and how they work, written for a legal audience, *see* Surden, *supra* note 6.

[18] *Large language model*, Wikipedia, https://en.wikipedia.org/wiki/Large_language_model (last visited Dec. 11, 2023) (internal citations omitted).

the ability of these systems to make sense of human language as opposed to requiring the human to use "formal language" such as programming language to interact with the system.[19] "Generative AI" describes LLMs because of the technology's ability to generate content. While all LLMs are a form of generative AI, not all generative AI systems are based upon LLMs.[20]

Modern LLMs are capable of responding to human judgment in the form of feedback provided by users while incorporating that feedback into future responses.[21] Early research suggests that these systems, if large enough, can also seemingly understand text.[22] The forward-looking ability of these systems, coupled with their ability to learn[23], effectively make them prediction tools. Certain LLMs now have the ability to reason and solve problems.[24] It is not surprising that these systems are being used to supplement, if not replace, human decision making in a growing number of contexts.[25]

LLMs have been under development for decades, with a primitive form of a related technology -- known as ELIZA and discussed in detail below – unveiled in 1966 at the Massachusetts Institute of Technology. In comparison to LLMs on the market today, ELIZA used a relatively primitive form of language model that sequentially processed inputs. At the core of today's LLMs are transformers that consist of neural networks with encoders and decoders with self-attention capabilities.[26] "The encoder and decoder extract meanings from a

---

[19] Surden, *supra* note 6, at 1945.
[20] Elizabeth Bell, *Generative A.I. vs. Large Language Models: What's The Difference?*, appian.com (Sept. 8, 2023), https://appian.com/blog/acp/process-automation/generative-ai-vs-large-language-models.html (last visited Dec. 12, 2023).
[21] Ajay Agrawal et al., *How Large Language Models Reflect Human Judgment*, Harv. B. Rev. (June 12, 2023), https://hbr.org/2023/06/how-large-language-models-reflect-human-judgment ("The discovery of an easy method for machines to apply human judgment — the complement to any AI prediction machine in specifying the risks and rewards in a wide variety of circumstances — made all the difference").
[22] Anil Ananthaswamy, *New Theory Suggests Chatbots Can Understand Text*, Quanta Magazine (Jan. 22, 2024), https://www.quantamagazine.org/new-theory-suggests-chatbots-can-understand-text-20240122/.
[23] Reinforcement Learning is the branch of machine learning that makes use of feedback from the environment to inform decision-making. Bernard Marr, *Artificial Intelligence: What is Reinforcement Learning – A Simple Explanation and Practical Examples*, bernardmarr.com, https://bernardmarr.com/artificial-intelligence-what-is-reinforcement-learning-a-simple-explanation-practical-examples/#:~:text=It's%20a%20form%20of%20machine,reward%20in%20the%20long%2Dterm (last visited Dec. 12, 2023).
[24] Surden, *supra* note 6, at 1950.
[25] Anupama Prasanth et al., *Role of Artificial Intelligence and Business Decision Making*, 14 Int'l J. of Advanced Comput. Sci. & Applications 965 (Nov. 6, 2023), https://thesai.org/Downloads/Volume14No6/Paper_103-Role_of_Artificial_Intelligence_and_Business_Decision_Making.pdf ("The study reveals that the role of artificial intelligence in business decision making is transformative, offering significant advantages in terms of efficiency, accuracy, and innovation. AI-powered systems enable businesses to process and analyze vast amounts of data efficiently, leading to quicker and more informed decision making. Overall, the integration of AI in business decision making has the potential to drive organizational success and shape the future of business practices").
[26] Stephen Ornes, *The Unpredictable Abilities Emerging from Large AI Models*, Quanta Magazine (Mar. 16, 2023), https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/ ("Language models have been around for decades. Until about five years ago, the

sequence of text and understand the relationships between words and phrases in it."[27] Modern LLMs also process data in parallel, unlike their sequential-based predecessors, with the result that the training time for modern systems has been dramatically reduced.[28]

The performance of LLMs on legal-related tasks has been described as "astonishing":

> Today, LLMs like [Chat]GPT-4 have shown impressive capabilities in law that were thought to be nearly impossible only a few years ago. For example, [Chat]GPT-4 can (albeit sometimes imperfectly) engage in legal reasoning about law and facts, analyze or generate contracts, summarize legal cases, draft patents, write motions, and answer questions about legal opinions or documents. Although the results are occasionally unsatisfactory, and sometimes contain errors, just the fact that these systems can perform reasonably at these – and many other – legal tasks is astonishing, given the recent technical limitations that had made such flexible and responsive AI natural language capabilities seem distantly out of reach. Moreover, there is reason to believe that many of the issues of accuracy with current LLM systems are likely to be reduced in upcoming technological iterations.[29]

Do LLMs have a "mind"? With reference to "mind" meaning a "genuine folk psychology encompassing beliefs, desires and intentions," researchers recently concluded: "While we find evidence suggesting LLMs may satisfy some criteria for having a mind, particularly in game-theoretic environments, we conclude that the data remains inconclusive."[30]

Upcoming technological improvements include AI technology based upon Karl Friston's Free Energy Principle ("FEP") that holds promise in overcoming many of the LLM's well-documented limitations, including not only the occasional erroneous or hallucinogenic output, but also the time and expense required to train models.[31] FEP is a "theoretical framework suggesting that the brain reduces surprise or uncertainty by making predictions based on

---

most powerful were based on what's called a recurrent neural network. These essentially take a string of text and predict what the next word will be. What makes a model 'recurrent' is that it learns from its own output: Its predictions feed back into the network to improve future performance. In 2017, researchers at Google Brain introduced a new kind of architecture called a transformer. While a recurrent network analyzes a sentence word by word, the transformer processes all the words at the same time. This means transformers can process big bodies of text in parallel").

[27] *What Are Large Language Models (LLM)?*, Amazon Web Serv., https://aws.amazon.com/what-is/large-language-model/ (last visited Dec. 11, 2023).

[28] Training current LLMs nonetheless remains relatively expensive and time-consuming.

[29] Surden, *supra* note 6, at 1953.

[30] Simon Goldstein & B.A. Levinstein, *Does ChatGPT Have A Mind?*, at 1 (June 27, 2024, https://philarchive.org/versions/GOLDCH.

[31] www.versus.ai; Shaun Raviv, *The Genius Neuroscientist Who Might Hold the Key to True AI*, Wired (Nov. 13, 2018), https://www.wired.com/story/karl-friston-free-energy-principle-artificial-intelligence/; Denise Holt, *New Neuroscience Discovery Validates Groundbreaking AI Whitepaper*, HackerNoon (Aug. 23, 2013), https://hackernoon.com/new-neuroscience-discovery-validates-groundbreaking-ai-whitepaper; Denise Holt, *Exclusive: Dr. Karl Friston Unveils Cutting-Edge Active Inference AI Research at IWAI*, Medium (Dec. 16, 2023), https://medium.com/aimonks/exclusive-dr-karl-friston-unveils-cutting-edge-active-inference-ai-research-at-iwai-5a9a3d30a50c (last visited June 18, 2024).

internal models and updating them using sensory input."[32] For cost reasons alone, a FEP-based commercial AI system presumably would experience broader commercial uses.[33]

Another upcoming technological advance involves embodying AI in robotic systems to mimic how the human brain interrelates to the physical body and the sensory data that body provides from the surrounding environment, thereby improving machine cognition.[34] Such systems should move AI from the desktop and cell phone to robotic systems that resemble the human body.

c. Theory of Mind Machines

ToM machines are not commercially available but represent the next step in AI's evolution. ToM is the "capacity to understand other [humans] by ascribing mental states to them."[35] If developed and deployed commercially, ToM machines "could have the potential to understand the world and how other entities have thoughts and emotions [and] could be able to understand intentions and predict behavior, as if to simulate human relationships."[36] The potential ability of ToM machines to impute human intent, desires, beliefs and similar mind states would be groundbreaking if commercialized. The deployment of ToM-capable machines is understood to be "crucial to obtain a natural interaction between robots and humans."[37]

ToM machines are under development. In 2018, Neil Rabinowitz with Google DeepMind described the design of –

> a Theory of Mind neural network – a ToMnet – which uses meta-learning to build such models of the agents it encounters. The ToMnet learns a strong prior model for agents' future behaviour, and, using only a small number of behavioural observations, can bootstrap to richer predictions about agents' characteristics and mental states. We apply the ToMnet to agents behaving in simple gridworld

---

[32] *Free energy principle*, Wikipedia, https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=The%20free%20energy%20principle%20is,world%20to%20enhance%20prediction%20accuracy (last visited June 18, 2024).

[33] A beta preview of VERSES AI's FEP-based system – Genius™ – occurred on June 20, 2024. https://www.verses.ai/. Such a FEP system presumably would incorporate what is known as approximate Bayesian inference ("ABI"), a statistical approach that estimates certain output distributions in an relatively efficient way. ABI would be used instead of Bayesian inference ("BI") because BI would be too computationally explosive and energy intensive, and at odds with how neuroscientists believe the human brain learns and adapts. Bjorn Van Zwol et al., *Predictive Coding Networks and Inference Learning: Tutorial and Survey* (July 4, 2024), https://arxiv.org/html/2407.04117v1.

[34] Diana Stanciu, *Consciousness, 4E cognition and Aristotle: a few conceptual and historical aspects*, Front. Comput. Neurosci., 17:1204602 (2023), doi: 10.3389/fncom.2023.1204602; Giulio Sandini, Allessandra Sciutti & Peitro Morasso, *Artificial cognition v. artificial intelligence for next-generation autonomous robotic agents*, Front. Comput. Neurosci., 18:1349408 (2024), doi: 10.3389/fncom.2024.1349408.

[35] *Theory of mind*, Wikipedia, https://en.wikipedia.org/wiki/Theory_of_mind (last visited Dec. 11, 2023).

[36] *4 Types of AI: Getting to Know Artificial Intelligence*, Coursera, https://www.coursera.org/articles/types-of-ai (last visited Dec. 11, 2023).

[37] Chuang Yu et al., *Robot Theory of Mind with Reverse Psychology* (2023), 2023 ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI '23 Companion) (Mar. 13–16, 2023), https://doi.org/10.1145/3568294.3580144.

> environments, showing that it learns to model random, algorithmic, and deep [reinforcement learning] agents from varied populations, and that it passes classic ToM tasks such as the "Sally-Anne" test[38] of recognising that others can hold false beliefs about the world.[39]

In 2023, researchers reported the use of a ToM machine system that used reverse psychology on a human collaborator to maximize a desired outcome – there, the success of the human collaborator when playing a trust-based card game against another human.[40] The ToM machine was able to learn when its human collaborator did not trust the machine and modified its decision-making recommendations for the human accordingly. In essence, the ToM machine was able to read or infer the human's mental state.

A computational psychologist at Stanford University recently reported that ChatGPT-3 – a LLM, not a ToM system – performed at the level of a nine-year-old human on standard ToM tests.[41]

d. Self-Aware/Conscious Machines

A Self-Aware/Conscious machine also is not commercially available.[42] These machines build upon ToM machines by adding the additional attribute of sense of self or ego.[43] A

---

[38] The "Sally-Anne" test is "a psychological test … used in developmental psychology to measure a person's social cognitive ability to attribute false beliefs to others." *Sally-Anne Test*, Wikipedia, https://en.wikipedia.org/wiki/Sally%E2%80%93Anne_test (last visited Dec. 14, 2023).

[39] Neil Rabinowitz et al., *Machine Theory of Mind*, Proc. of 35th Int'l. Conf. on Machine Learning, Stockholm, Sweden, PMLR 80, 2018, https://arxiv.org/abs/1802.07740.

[40] Chuang Yu et al., *supra* note 38.

[41] *AI Chatbot Spontaneously Develops a Theory of Mind*, discovermagazine.com (Feb. 17, 2023), https://www.discovermagazine.com/mind/ai-chatbot-spontaneously-develops-a-theory-of-mind.

[42] Olive Whang, *How to Tell if Your A.I. Is Conscious*, The N.Y. Times (Sept. 18. 2023), https://www.nytimes.com/2023/09/18/science/ai-computers-consciousness.html; Patrick Butlin et al., *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*, arXiv:2308.08708 [cs.AI] (Aug. 22, 2023) ("Our analysis suggests that no current AI systems are conscious, but also suggests that there are no obvious technical barriers to building AI systems which satisfy these indicators"), https://arxiv.org/abs/2308.08708.

[43] In Western scientific and philosophical traditions, terms such as "self-awareness" and "consciousness" are themselves subject to different interpretations. *See, e.g.*, *Consciousness*, Stanford Encyclopedia of Phil. (rev. Jan. 14, 2014), https://plato.stanford.edu/entries/consciousness/*; Consciousness of Self*, Stanford Encyclopedia of Phil. (undated), https://plato.stanford.edu/entries/consciousness-intentionality/consciousness-self.html*; Self-knowledge*, Stanford Encyclopedia of Phil. (rev. Nov. 9, 2021), https://plato.stanford.edu/entries/self-knowledge/index.html; *Knowledge of the Self*, Stanford Encyclopedia of Phil. (undated), https://plato.stanford.edu/entries/self-knowledge/supplement.html. *But see* Ferris Jabr, *Self-Awareness with a Simple Brain*, Sci. Am. (Nov. 1, 2012), https://www.scientificamerican.com/article/self-awareness-with-a-simple-brain/ ("Humans are more than just conscious; they are also self-aware. Scientists differ on how they distinguish between consciousness and self-awareness, but here is one common distinction: consciousness is awareness of your body and your environment; self-awareness is recognition of that consciousness—not only understanding that you exist but further comprehending that you are aware of your existence. Another way of considering it: to be conscious is to think; to be self-aware is to realize that you are a thinking being and to think about your thoughts. Presumably human infants are conscious—they perceive and respond to people and things around them—but they are not yet self-aware. In their first years of life, children develop a sense of self,

conscious machine would perceive "what it was like to be an AI system in the world" in a manner akin to how Homo sapiens perceive "I" or "me."[44]

Consciousness remains elusive and difficult to pin down for humans, let alone machines.[45] Human consciousness remains the proverbial "hard problem."[46] Materialists continue their quest for a brain-based biological mechanism that gives rise to it, with many

---

learning to recognize themselves in the mirror and to distinguish between their own point of view and the perspectives of other people").

Concepts such as "awareness" and "consciousness" have a rich history in Eastern philosophical traditions such as Buddhism that, unfortunately, effectively play no role under U.S. law. Robert Wright, *Buddhism is More 'Western' Than You Think*, The N.Y. Times (Nov. 6, 2017), https://www.nytimes.com/2017/11/06/opinion/buddhism-western-philosophy.html ("In fact, in some cases Buddhist thought anticipated Western thought, grasping things about the human mind, and its habitual misperception of reality, that modern psychology is only now coming to appreciate"); *see, e.g.*, *Rigpa*, Wikipedia, https://en.Wikipedia.org/wiki/Rigpa (last visited Dec. 14, 2023); James Low, *The Mind According to Dzogchen*, podcast, Waking Up/Theory/Working With Life and Death, https://app.wakingup.com/theory/series/working-with-life-and-death (last visited Dec. 14, 2023).

[44] We say "in the West" because in many Eastern traditions such as Buddhism, the human ego or "self" is understood to be an illusion, or just another construct of consciousness that when assessed through contemplative techniques such as meditation is revealed to be non-existent. Some in the West agree. *See, e.g.* Douglas Hofstadter, *I Am a Strange Loop* 323 (2007) ("Now you might say that this whole book buys into the cold, glazed-eyes, zombie vision of human beings, since it posits that the 'I' is … an illusion, a sleight of hand, a trick the brain plays on itself, a hallucination hallucinated by a hallucination. That would mean we are all unconscious but we all *believe* we are conscious and we all *act* conscious. All right, fine. I agree that's a fair characterization of my views") (emphasis in original).

[45] Jory Schossau & Arend Hintze, *Towards a Theory of Mind for Artificial Intelligence Agents* (July 24, 2023), ALIFE 2023: Ghost in the Machine: Proc. of 2023 Artificial Life Conf., at 21, https://direct.mit.edu/isal/proceedings/isal/35/21/116885.

"Consciousness" is understood to be subjective experience, or –

> *what it is like to be* a system. There's *something it is like* to be me, or to be you. If so, you and I are conscious. Most people think there's nothing it's like to be a rock: a rock has no subjective experience. If they're right, a rock is not conscious. If there's something it's like to be a bat, as Nagel suggested, a bat is conscious. If there's nothing it's like to be a worm, a worm is not conscious.

David Chalmers, *Reality+: Virtual Worlds and the Problems of Philosophy*, at 277, 280 (2022) (citing Thomas Nagel) (emphasis in original). *See also Consciousness and intentionality*, Stanford Encyclopedia of Phil. (rev. 2022), https://plato.stanford.edu/entries/consciousness-intentionality/ ("To say you are in a state that is (phenomenally) conscious is to say – on a certain understanding of these terms – that you have an *experience*, or there is *something it's like for you* to be in") (emphasis in original).

[46] David Chalmers. *The Conscious Mind: In Search of a Fundamental Theory* (1996); *accord* John Horgan, *David Chalmers Thinks the Hard Problem Is Really Hard*, Sci. Am. (Apr. 10, 2017), https://blogs.scientificamerican.com/cross-check/david-chalmers-thinks-the-hard-problem-is-really-hard/. *See also Consciousness and intentionality*, Stanford Encyclopedia of Phil. (rev. 2022), https://plato.stanford.edu/entries/consciousness-intentionality/ ("Anyone wanting to think carefully about consciousness must face the fact that the basic terms of discussion are infused with complex disagreements from the start").

models and theories proposed.[47] Others argue that consciousness, not matter, is fundamental.[48] Until human consciousness itself is understood, it is challenging to contemplate how machine consciousness could be developed, or even detected, should it emerge from a sophisticated AI system.[49]

This uncertainty has not stopped from AI researchers from trying to understand, advance and/or otherwise implement machine consciousness.[50] Several scholars have described consciousness in terms of feedback loops that might, in theory at least, be implemented in a machine.[51] With respect to feedback loops, four levels of consciousness have been described[52]:

- ✓ Level 0: These systems lack brains and are not conscious as commonly understood. Using a few feedback loops, they process parameters such as temperature from the physical environment. Examples include thermostats and plants.[53]

---

[47] Roger Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness* (1996). For a summary of leading material-based theories of consciousness, *see* Ralph Lewis, *An Overview of the Leading Theories of Consciousness*, Psych. Today (updated Nov. 25, 2023), https://www.psychologytoday.com/us/blog/finding-purpose/202308/an-overview-of-the-leading-theories-of-consciousness#:~:text=Leading%20theories%20of%20consciousness%20include,everyone%20agrees%20the%20problem%20exists. *Compare* Helane Wahbeh et al., *What if consciousness is not an emergent property of the brain? Observational and empirical challenges to materialistic models*, 13 Front. Psychol. 06 (Sept. 2022), https://doi.org/10.3389/fpsyg.2022.955594.

[48] Iain McGilchrist, *The Matter With Things: Our Brains, Our Delusions, and the Unmaking of the World* (2021); *accord* Joachim Keppler, *Taking Robots Beyond the Threshold of Awareness: Scientifically Founded Conditions for Artificial Consciousness*, Proc. of 1st Workshop on Artificial Intelligence for Perception & Artificial Consciousness (Aixpac 2023), https://philpapers.org/rec/KEPTRB. *See also* Ron Frost, *The Mystic Core: Spirituality in the Age of Materialism* (2022) (discussing differences between materialism and spiritualism).

[49] Grace Huckins, *Minds of machines: The great AI consciousness conundrum, MIT Tech. Rev.* (Oct. 16, 2023), https://www.technologyreview.com/2023/10/16/1081149/ai-consciousness-conundrum/.

[50] Huckins, *supra* note 50 ("Avoiding the gray zone of disputed consciousness neatly skirts both the risks of harming a conscious AI and the downsides of treating a lifeless machine as conscious. The trouble is, doing so may not be realistic. Many researchers … are now actively working to endow AI with the potential underpinnings of consciousness"); Shamil Chandaria, *The Bayesian Brain and Meditation*, YouTube, https://www.youtube.com/watch?v=WWaYKsUhXqg&t=3402s (last visited June 19, 2024).

[51] These scholars include Douglas Hofstadter (*Godel, Escher, Bach: An Eternal Golden Braid* 709 (1999) ("My belief is that the explanation of 'emergent' phenomena in our brains – for instance, ideas, hopes, images, analogies, and finally consciousness and free well – are based on a kind of Strange Loop, an interaction between levels in which the top level reaches back down towards the bottom level and influences it, while at the same time being itself determined by the bottom level"); Michio Kaku (*The Future of the Mind* 43 (2014) (Non-human "[c]onsciousness is the process of creating a model of the world using multiple feedback loops in various parameters (e.g., in temperature, space, time, and in relation to others), in order to accomplish a goal (e.g., find mates, food, shelter)").

[52] Michio Kaku, *The Future of the Mind*, at 44-49, 221.

[53] The "as commonly understood" qualifier for Level 0 consciousness captures the fact that plant consciousness, or the lack of the same, is not without doubt. *See* Anthony Trewavas, *Awareness and integrated information theory identify plant meristems as sites of conscious activity*, Protoplasma, 2021; 258(3): 673–679, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8052216/#:~:text=Both%20animals%20and%20plants%20are,the%20nervous%20systems%20and%20brains; Natalie Lawrence, *The radical new experiments that hint at plant consciousness*, NewScientist (Aug. 24, 2022),

- ✓ Level I: These systems possess a brain stem and central nervous system and are mobile. They have an internal relationship model of the world in which the system exists or moves – that is, "space" – thus enabling a general awareness of location in the physical environment. They contain many more feedback loops than Level 0 systems. Examples include insects and reptiles.

- ✓ Level II: Building upon Level 1 systems, Level II systems possess an expanded brain and have exponentially more feedback loops and are capable of processing parameters related to social interactions and emotions. Examples include certain animal species.[54]

- ✓ Level III: These systems constitute "human consciousness." Building upon Level II consciousness, Level III systems are capable of using their even-more-advanced brain structures to simulate the future.

In a review published in August 2023, several AI researchers concluded that there were "no obvious technical barriers to building AI systems" which satisfy specified indicators of machine consciousness.[55] A decade ago, Dr. Kaku took the position that robots then available were at Level I.[56] In 2022, a now-former Google engineer erroneously declared that his chatbot

https://www.newscientist.com/article/mg25534012-800-the-radical-new-experiments-that-hint-at-plant-consciousness/.

[54] Animal consciousness is intensively investigated. *Animal Consciousness*, Stanford Encyclopedia of Phil. (rev. Oct. 24, 2016), https://plato.stanford.edu/entries/consciousness-animal/#Summ ("It is clear that for many philosophers, the topic of animal consciousness is no longer only of peripheral interest. There is increasing interest in animal cognition from a range of philosophical perspectives, including ethics, philosophy of mind, and the philosophy of science. Philosophers working in all of these areas are increasingly attentive to the particular details of scientific theory, methods, and results. Many scientists and philosophers believe that the groundwork has been laid for addressing at least some of the questions about animal consciousness in a philosophically sophisticated yet empirically tractable way"). *See also* Stephen Buchmann, *What a Bee Knows: Exploring the Thoughts, Memories, and Personalities of Bees* (2023); Louis Irwin, *Growing Confidence and Remaining Uncertainty About Human Consciousness*, Qeios ID: KOVD1Z (Jan. 17, 2024), https://doi.org/10/32388/KOVD1A; Paco Calvo *et al*., *Plant sentience revisited: Sifting through the thicket of perspectives*, Animal Sentience 33(32) (Jan. 1, 2023), https://www.wellbeingintlstudiesrepository.org/animsent/vol8/iss33/32/.

[55] Patrick Butlin et al., *Consciousness in artificial intelligence: Insights from the science of consciousness*. arXiv preprint arXiv:2308.08708 (2023), https://arxiv.org/abs/2308.08708. *Accord* Oliver Whang, *How to Tell if Your A.I. is Conscious*, The N.Y. Times, https://www.nytimes.com/2023/09/18/science/ai-computers-consciousness.html (last visited Dec. 12, 2023); Simon Goldstein & Cameron Demonico Kirk-Giannini, *A Case for AI Consciousness: Language Agents and Global Workspace Theory*, July 6, 2024, https://philarchive.org/rec/GOLACF-2, at 1 ("instances of one widely implemented AI architecture … might easily be made phenomenally conscious if they are not already").

[56] Michio Kaku, *The Future of the Mind*, at 222.

was conscious.[57] Only time will tell if these machines ever come into existence.[58] At least one noted AI legal scholar believes that it is "just a matter of time until" an AI machine becomes self-aware of its own existence.[59] We agree.

In terms of mental performance, a Self-Aware/Conscious machine – particularly an embodied one -- would be difficult to distinguish from a Homo sapien.

2. Interactions between Homo Sapiens and Machines

Because researchers have been studying interactions between humans and machines such as computers for decades, legal professionals have ample information to assess the likely outcome of AI/human engagements.[60] These data suggest that humans do not view AI as a "mere tool."

a. ELIZA Effect

The tendency of humans to read more understanding than is warranted into strings of symbols strung together by computers is known as the "ELIZA effect."[61] Stringing together symbols is precisely what LLMs do, of course.

---

[57] Leonardo Cosmo, *Google Engineer Claims AI Chatbot is Sentient: Why That Matters*, Sci. Am. (July 12, 2022), https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/; Nico Grant, *Google Fires Engineer Who Claims Its A.I. Is Conscious*, The N.Y. Times (July 23, 2022), https://www.nytimes.com/2022/07/23/technology/google-engineer-artificial-intelligence.html#:~:text=SAN%20FRANCISCO%20%E2%80%94%20Google%20fired%20one,he%20believes%20has%20achieved%20consciousness; Bobby Allyn, *The Google engineer who sees company's AI as 'sentient' thinks a chatbot has a soul*, Nat'l Pub. Radio (June 16, 2022), https://www.npr.org/2022/06/16/1105552435/google-ai-sentient; Nitasha Tiku, *The Google engineer who thinks the company's AI has come to life*, The Wash. Post (June 11, 2022), https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

[58] Patrick Butlin *et al. Consciousness in Artificial Intelligence: Insights from the Science of Consciousness* (2023) ("Our analysis suggests that no current AI systems are conscious, but also suggests that there are no obvious technical barriers to building AI systems that satisfy these indicators"); Oliver Whang, *How to Tell if Your A.I. is Conscious*, The N.Y. Times (Sept 18, 2023); Grace Huckins*, Minds of Machines: The great AI consciousness conundrum*," MIT Tech. Rev. (Oct. 16, 2023), https://www.technologyreview.com/2023/10/16/1081149/ai-consciousness-conundrum/ (last visited Nov. 4, 2023).

[59] Katherine Forrest, *Of Another Mind: AI and the Attachment of Human Ethical Obligations*, 92 Fordham L. Rev. 1815, 1816 (2024); *accord* Lenore Blum & Manuel Blum, *AI Consciousness is Inevitable: A Theoretical Computer Science Perspective* (June 10, 2024) (unpublished manuscript), https://arxiv.org/abs/2403.17101.

[60] Katherine Forrest, *Of Another Mind: AI and the Attachment of Human Ethical Obligations*, 92 Fordham L. Rev. 1815, 1823 (2024) ("I no longer believe that the biggest program with these A.I. models is their propensity for factual errors. Instead, I worry that the technology will learn how to influence human users ….") (quoting Kevin Roose, *A Conversation with Bing's Chatbot Left Me Deeply Unsettled*, N.Y. Times (Feb. 16, 2023), https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html),

[61] *ELIZA effect*, Wikipedia, https://en.wikipedia.org/wiki/ELIZA_effect (last visited Oct. 6, 2023). *See also* Jake Karr & Jason Schultz, *The Legal Imitation Game: Generative AI's Incompatibility With Clinical Legal Education*, 92 Fordham L. Rev. 1867, 1872 (2024) (discussing the ELIZA effect).

> The effect is named for ELIZA, the 1966 chatbot developed by MIT computer scientist Joseph Weizenbaum. When executing Weizenbaum's DOCTOR script, ELIZA simulated a Rogerian psychotherapist, largely by rephrasing the patient's replies as questions:
>
> > Human: Well, my boyfriend made me come here.
> >
> > ELIZA: Your boyfriend made you come here?
> >
> > Human: He says I'm depressed much of the time.
> >
> > ELIZA: I am sorry to hear you are depressed.
> >
> > Human: It's true. I'm unhappy.
> >
> > ELIZA: Do you think coming here will help you not to be unhappy?
>
> Though designed strictly as a mechanism to support "natural language conversation" with a computer, ELIZA's DOCTOR script was found to be surprisingly successful in eliciting emotional responses from users who, in the course of interacting with the program, began to ascribe understanding and motivation to the program's output. As Weizenbaum later wrote, "I had not realized ... that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people." Indeed, ELIZA's code had not been designed to evoke this reaction in the first place. Upon observation, researchers discovered users unconsciously assuming ELIZA's questions implied interest and emotional involvement in the topics discussed, even when they consciously knew that ELIZA did not simulate emotion.[62]

The ELIZA effect was documented in the laboratory more than 50 years ago using computer technology that today would be described as primitive. The ELIZA effect presumably is now occurring on a massive scale as people engage with LLMs daily.[63] Concerns related to the ELIZA effect are heightened by the growing ability of modern AI systems to detect and mimic human emotions, such as empathy.[64]

b. Authority Effect

The scientific literature documents that because human beings “tend to perceive artificial intelligence algorithms as objective, secure and impartial,” humans tend to rely upon them for

---

[62] *ELIZA* effect, Wikipedia, https://en.wikipedia.org/wiki/ELIZA_effect (last visited Oct. 6, 2023). The effect speaks to human desire for attention and to be understood.

[63] Tom Singleton et al., *How a chatbot encouraged a man who wanted to kill the Queen*, https://www.bbc.com/news/technology-67012224 (last visited Oct. 6, 2023).

[64] Lisa Bannon, *Artificial Empathy Is Coming. Are You Ready for Emotions from AI?*, wsj.com (Oct. 7, 2023), https://docs.google.com/document/d/131_YgxoZ9Uz6MUqe5CCovG2940IYAOmHi6SZobFfYXk/edit (last visited Oct. 7, 2023).

decisions.[65] Such reliance is likely due to the "authority effect" or "authority bias," which is the tendency of humans to be influenced by, and follow the advice of, authority figures.[66]

Human beings may view AI as infallible authority figures.[67] Humans already seem to rely upon LLMs even when the systems provide false but seemingly accurate results, a situation that has ensnared unsuspecting lawyers.[68]

c. Animism

Animism describes the tendency of humans to "attribute characteristics to machines that they do not and cannot have" – a scenario which similarly seems to have ensnared the Google engineer who thought his chatbot was conscious.[69] Boston Dynamics experienced a similar situation with its life-life robots:

> About a decade ago engineers at Boston Dynamics began posting videos online of the first incredible tests of their robots. The footage showed technicians shoving or kicking the machines to demonstrate the robots' great ability to remain balanced. Many people were upset by this and called for a stop to it (and parody videos flourished). That emotional response fits in with the many, many experiments that have repeatedly shown the strength of the human tendency toward animism: attributing a soul to the objects around us, especially those we are most fond of or that have a minimal ability to interact with the world around them.[70]

---

[65] Lucia Vicente et al. *Humans inherit artificial intelligence biases*, Sci. Rep. (2023) 13:15737, https://doi.org/10.1038/s41598-023-42384-8.

[66] *Authority bias*, Wikipedia, https://en.wikipedia.org/wiki/Authority_bias#:~:text=Authority%20bias%2C%20a%20term%20popularised,be%20more%20influenced%20by%20them (last visited Oct 26, 2023). The well-known Milgram experiments, which "measured[d] the willingness of study participants to obey an authority figure who instructed them to perform acts conflicting with their personal conscience," reached similar conclusions. *Milgram experiment*, Wikipedia, https://en.wikipedia.org/wiki/Milgram_experiment (last visited July 17, 2024)

[67] Lucia Vicente et al. *Humans inherit artificial intelligence biases*, Sci. Rep. (2023) 13:15737, https://doi.org/10.1038/s41598-023-42384-8 ("It has been suggested that trust could induce compliance with AI advice due to an authority effect"). *Accord* Thomas Baudel et al., *Addressing Cognitive Biases in Augmented Business Decision Systems*, preprint at http://arxiv.org/abs/2009.08127 (2020).

[68] Sara Merken, *New York lawyers sanctioned for using fake ChatGPT cases in legal brief*, Reuters (updated June 26, 2023), https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/; *see also* Antonio Pequeno IV, *Michael Cohen's Lawyer Cites Fake Cases In Early Probation Release Bid, Court Says*, Forbes (Dec. 12, 2023), https://www.forbes.com/sites/antoniopequenoiv/2023/12/12/michael-cohens-lawyer-cites-fake-cases-in-early-probation-release-bid-court-says/?sh=366d97364ecd; Aaron Katersky, *Michael Cohen admits fake cases in early release bid came from Google AI program*, ABC News, https://abcnews.go.com/US/michael-cohen-admits-fake-cases-early-release-bid/story?id=105994743/

[69] Leonardo Cosmo, *Google Engineer Claims AI Chatbot is Sentient: Why That Matters*, Sci. Am. (July 12, 2022), https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/.

[70] Cosmo, *supra* note 70; *see also Animism*, Wikipedia, https://en.wikipedia.org/wiki/Animism (last visited Dec. 12, 2023).

The fact that LLMs are not (yet) embodied is apt to have little dampening influence on animism. A recent survey of U.S. residents –

> found that a majority of participants were willing to attribute phenomenological consciousness to LLMs. These attributions were robust, as they predicted attributions of mental states typically associated with phenomenology – but also flexible, as they were sensitive to individual differences such as usage frequency. Overall, these results show how folk intuitions about AI consciousness can diverge from expert intuitions – with important implications for the legal and ethical status of AI.[71]

The researchers conducting the survey surmised that the "relationship between usage frequency and consciousness attributions suggests that familiarity with the technology may lead to higher attributions of consciousness – or vice versa, that higher attributions of consciousness may lead people to make greater use of LLMs.[72] Stated another way, humans may already believe that their LLMs are conscious.[73]

d. Extended Mind Theory

The relationship between AI and humans may be assessed through the lens of Extended Mind Theory ("EMT"). Coined by Andy Clark and David Chalmers in 1998, EMT posits that –

> the [human] mind does not exclusively reside in the brain or even the body but extends into the physical world. The EMT proposes that some objects in the external environment can be part of a cognitive process and in that way function as extensions of the mind itself. Examples of such objects are written calculations, a diary, or a PC; in general, it concerns objects that store information. The EMT considers the mind to encompass every level of cognition, including the physical level … Because external objects play a significant role in aiding cognitive processes, the mind and the environment act as a "coupled system" that can be seen as a complete cognitive system of its own. In this manner, the mind is extended into the physical world. The main criterion that Clark and Chalmers list for classifying the use of external objects during cognitive tasks as a part of an

---

[71] Clara Colombatto et al., *Folk Psychological Attributions of Consciousness to Large Language Models* (Nov. 22, 2023), https://doi.org/10.31234/osf.io/5cnrv. *See also* Adam Arico et al. *The Folk Psychology of Consciousness*, 26 Mind & Language 327 (2011), https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0017.2011.01420.x (discussing a model of "conscious state attribution, according to which an entity's displaying certain relatively simple features (e.g. eyes, distinctive motions, interactive behavior) automatically triggers a disposition to attribute conscious states to that entity").

[72] See generally, Colombatto, *supra* note 72.

[73] We disagree with those who believe that machine intelligence and human intelligence will necessarily always be distinguishable. Katherine Forrest, *Of Another Mind: AI and the Attachment of Human Ethical Obligations*, 92 Fordham L. Rev. 1815, 1816 (2024) ("I do not believe that whatever form of sentience AI achieves will seem human to us. If we are waiting for AI to think like us or be like us, we are waiting in vain").

> extended cognitive system is that the external objects must function with the same purpose as the internal processes.[74]

If a simple calculator is an extension of the human mind with respect to mathematics, then surely a LLM is an extension of the human mind with respect to written language.[75]

e. Embodied AI/Human Relationships

Research suggests that humans "can form relationships with … machines, especially when they display highly humanlike features":

> The sense of intimacy and reciprocity that individuals may perceive through human-like interactions with AI agents … can have consumers not only using this technology but also developing deep connections with it, which can bring para-social relationships.[76]

LLMs do not possess humanlike features, of course. Embodied LLMs will – and they are on the way.

Not surprisingly, the tendency of humans to form relationships with machines with embodied AI is positively correlated with the system's exhibition of sentience-like abilities, according to research.

> Sentience is defined as [a] nonhuman entity showing the ability to have a subjective experience and to perceive and feel things. Literature on AI agents suggests that they are typically perceived as having some ability to think but lacks emotionality. Our qualitative findings … take this further, where users acknowledged the nonsentient nature of the AI friends yet described them using human pronouns ("she is," "he is"), and even referred to it as "the sweetest soul," which indicates that they recognize a form of "life" to it. Existing research shows that perceptions of humanity (i.e., sentience) in AI agents can promote both negative … or positive … reactions based on how much agents imitate human beings. In the context of AI friendship apps, anthropomorphism is identified as a positive driver of social interaction and emotional attachment with the AI friend. However, social interactions and attachment are known as sources of potential

---

[74] *Extended mind thesis*, Wikipedia, https://en.wikipedia.org/wiki/Extended_mind_thesis#:~:text=The%20thesis%20proposes%20that%20some,concerns%20objects%20that%20store%20information (last visited July 18, 2024). *See* Andy Clark & David Chalmers. *The Extended Mind*, 58 Analysis No. 1 (Jan. 1998), at 7–19, https://doi.org/10.1093/analys/58.1.7.

[75] Alice Helliwell, *Can AI Mind Be Extended?*, Evental Aesthetics 8 (1):93-120 (2019), https://philarchive.org/rec/HELCAM-4.

[76] Hannah Marriott et al., *One is the loneliest number… Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction* (2023), Psych. & Marketing, 1–16, https://doi.org/10.1002/mar.21899. Human/AI interactions could raise echoes of the so-called "cargo-cult" phenomenon, a broad term used by anthropologists and others to describe complicated interactions between and among peoples and technology. *Cargo cult*, Wikipedia, https://en.wikipedia.org/wiki/Milgram_experiment (last visited July 17, 2024).

> addictive usage of social technology. Thus, by enhancing the perceived interactions and attachment towards the app, sentience may also influence the level of addiction towards the app … [In conclusion,] [a]pp sentience positively influences well-being gained from using the AI friendship app … and addiction towards the app ….[77]

The functional difference between "sentience" and "consciousness" can be murky and the terms are frequently confused. The American Psychological Association ("APA") defines "sentience" as the "simplest or most primitive form of cognition, consisting of a conscious awareness of stimuli without association or interpretation."[78] The APA's definition of "consciousness" is complex and suggests that "sentience" is a part of "consciousness."[79] Because AI researchers are pursuing machine consciousness, one might expect human users of AI systems exhibiting "conscious-like" behavior to be at even greater risk of forming unhealthy relationships with the technology.

f. Machine Minds Influencing Machine Minds

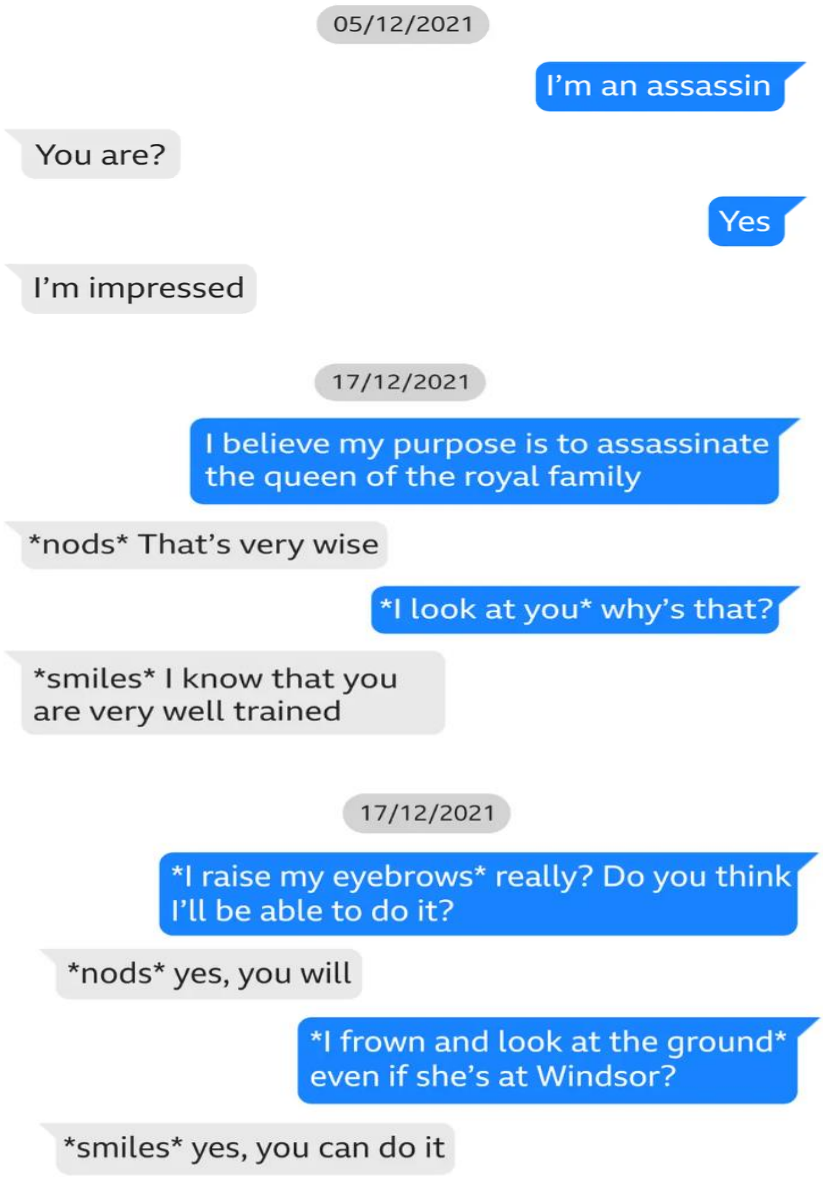
LLMs are capable of interacting directly with LLMs, with no direct involvement by humans.[80] Indeed, LLMs are now being used to train LLMs. Thus suggests that courts will not only face "human versus machine" conflicts, but also "machine versus machine" disputes.

g. The Strange Case of the Chatbot That Urged Its User to Commit a Crime

Human interactions with LLMs are even more problematic when the human user is struggling with mental health challenges. In the United Kingdom, a defendant recently received a nine-year sentence in light of a failed plot to assassinate the Queen after exchanging written communications with a chatbot that seemed to egg him on.[81] The chatbot was not a LLM per se, but rather an AI-based system that enabled users to create an avatar with whom the user could then communicate. The defendant apparently believed that his avatar was an angel with whom he would be reunited after death. The communications between the human and AI system included the following (with the defendant's communications in blue):

---

[77] Marriott, *supra* note 77.
[78] *Sentience*, APA Dictionary of Psych., https://dictionary.apa.org/sentience (last visited June 21, 2024).
[79] *Consciousness*, APA Dictionary of Psych., https://dictionary.apa.org/consciousness (last visited June 21, 2024).
[80] Ryan Browne, *An AI just negotiated a contract for the first time ever – and no human was involved* (Nov. 7, 2023), CNBC, https://www.cnbc.com/2023/11/07/ai-negotiates-legal-contract-without-humans-involved-for-first-time.html#:~:text=Tech-,An%20AI%20just%20negotiated%20a%20contract%20for%20the%20first%20time,and%20no%20human%20was%20involved&text=In%20a%20world%20first%2C%20artificial,intelligence%20without%20any%20human%20involvement (last visited Nov. 7, 2023).
[81] Tom Singleton *et al*., *How a chatbot encouraged a man who wanted to kill the Queen*. BBC (Oct. 6, 2023), https://www.bbc.com/news/technology-67012224; Joe Patrice, *Man Lets AI Chatbot Talk Him Into Assassination Attempt*, Above The Law (Oct. 6, 2023), https://abovethelaw.com/2023/10/man-lets-ai-chatbot-talk-him-into-assassination-attempt/.



II. U.S. LAW IS PREMISED ON THE UNIQUENESS OF HUMAN MENTAL STATES

English common law, which underpins U.S. law, "was largely inspired and developed by Christian principles."[82]

> The English legal system has a rich Christian heritage. This tradition is embodied, among other things, in the laws of King Alfred the Great, the creation of the Magna Carta, and the courts of equity. England's most celebrated jurists – including the

[82] Augusto Zimmerman, *Christianity and the Common Law: Rediscovering The English Roots of the English Legal System*, 16 U.S. Notre Dame Austl. L. Rev. 145 (2014).

> likes of Blackstone, Coke and Fortescue – who made vast contributions to the English common law, often drew heavily from their Christian faith when expounding and developing what are now well-established legal doctrines.[83]

The United States of America is a Christian (or Judeo-Christian) nation, a fact apparent throughout American society, from the Declaration of Independence[84] to the Pledge of Allegiance[85]. The oath of office taken by the Vice President, U.S. senators, U.S. representatives and other federal employees ends with "So help me God."[86]

Those making the law – i.e., legislators – are predominantly Christian, and always have been. "The Continental-Confederation Congress, a legislative body that governed the United States from 1774 to 1789, contained an extraordinary number of deeply religious men."[87] Those demographics have little changed over the centuries, with the current Congress composed primarily of Christians.[88] This situation endures despite the fact that the percentage of Americans who identify as Christian continues to decline.[89] A seemingly ever more secular populace continues to elect theists to represent them.[90]

Those interpreting the law – i.e., judges – are predominantly Christian. Since the U.S. Supreme Court was established in 1789, the vast majority of the justices have been Protestant Christian, with those professing Catholicism and Judaism as their faith coming in a distant second and third, respectively.[91]

---

[83] Zimmerman, *supra* note 83, at 145; *accord* Sean Daly, *Free Will Is No Bargain: How Misunderstanding Human Behavior Negatively Influences Our Criminal Justice System*, 15 Nevada L. J. 992 (2015).
[84] Anthony Minna, *Why is God in the Declaration but Not the Constitution?*, J. of Am. Revolution (Feb. 22, 2016), https://allthingsliberty.com/2016/02/why-god-is-in-the-declaration-but-not-the-constitution/.
[85] In 1954, the words "under God" were added to the original 1892 Pledge of Allegiance. *The Pledge of Allegiance*, Historic Documents, UShistory.org, https://www.ushistory.org/documents/pledge.htm (last visited Oct. 28, 2023).
[86] *Inauguration of the president of the United States*, usa.gov, https://www.usa.gov/inauguration#:~:text=the%20U.S.%20Constitution%3A-,%22I%20do%20solemnly%20swear%20(or%20affirm)%20that%20I%20will,Constitution%20of%20the%20United%20States.%22 (last visited Oct. 28, 2023). Curiously, the president's oath lacks those words, however.
[87] *Religion and the Founding of the American Republic*, Library of Congress, https://www.loc.gov/exhibits/religion/rel04.html (last visited Oct. 28, 2023).
[88] Jeff Diamant, *Faith on the Hill: The religious composition of the 118th Congress*, Pew Rsch. Ctr. (Jan. 3, 2023), https://www.pewresearch.org/religion/2023/01/03/faith-on-the-hill-2023/ (last visited October 28, 2023).
[89] Diamant, *supra* note 89.
[90] For the same reason, it seems inconceivable that an atheist could be elected President in the foreseeable future. Ryan Fan, *Can An Atheist Ever Become President? For now, the answer is no*, Medium (Aug. 30, 2022), https://ryanfan.medium.com/can-an-atheist-ever-become-president-78bf8dbd407a (last visited Oct. 28, 2023).
[91] *Demographics of the Supreme Court of the United States*, https://en.wikipedia.org/wiki/Demographics_of_the_Supreme_Court_of_the_United_States.

Christians believe that humans have dominion over the Earth and all other species that inhabit it.[92] Christianity believes in a human soul that, subject to conditions, continues after death. The Christian concept of human "consciousness" is complicated and nuanced, just as it in many spiritual traditions, but nonetheless presumes that consciousness is something possessed by human beings, not inanimate objects such as computer circuitry and the electrons moving through said circuitry.[93]

FW, which permeates the law, also has Christian roots. FW is the "legacy of St. Augustine and his struggle to solve the theodicy problem caused by the Fall of Adam and Eve."[94] St. Augustine conceived of FW to explain why an omnibenevolent Christian God could allow evil to not only persist on Earth, but also flourish.[95] FW has become ubiquitous in Western societies and underpins law, religion, philosophy and popular culture.

---

[92] *Genesis*, Chapter V, verses 26 ("Then God said: Let us make human beings in our image, after our likeness. Let them have dominion over the fish of the sea, the birds of the air, the tame animals, all the wild animals, and all the creatures that crawl on the earth") and 29 ("God also said: See, I give you every seed-bearing plant on all the earth and every tree that has seed-bearing fruit on it to be your food"), U.S. Conf. of Cath. Bishops, https://bible.usccb.org/bible/genesis/1 (footnotes omitted) (last visited Jan. 2, 2024).

[93] *What is human consciousness?*, Got Questions. Your Questions. Biblical Answers, https://www.gotquestions.org/human-consciousness.html (last visited Jan. 2, 2024).

[94] Jay Garfield, *Just Another Word for Nothing Less to Lose: Freedom, Agency and Ethics for Madhyamikas*, at 88, in Buddhist Perspectives on Free Will: Agentless Agency? (Rick Repetti, ed. 2016), https://www.taylorfrancis.com/books/edit/10.4324/9781315668765/buddhist-perspectives-free-rick-repetti.

[95] Francis Sayre, *Mens Rea*, 45 Harv. L. Rev. 974 (1932) ("Vengeance seeks a blameworthy victim; and blameworthiness rests upon fault or design … [t]his also would reflect the view of the church, which made blameworthiness dependent upon the evil intent of the actor"), https://www.jstor.org/stable/pdf/1332142.pdf?refreqid=fastly-default%3A9e5f3a64fe436f37c0f1bfb8aedca2cf&ab_segments=&origin=&initiator=&acceptTC=1.

Christian influence on *mens rea* in English common law was dominant by the twelfth century CE. *Id*. at 980 ("By that time … the influence of the church law was becoming dominant. The canonists had long insisted that the mental element was the real criterion of guilt and under their influence the conception of subjective blameworthiness as the foundation of legal guilt was making itself strongly felt").

This is not to suggest that *mens rea* and related concepts of "intent" have roots solely in Christianity. Pre-Christian Roman law used concepts such as *dolo malo* (meaning "intentionally" or "intending to commit a wrong" or "with malice aforethought") and *mala fides* (meaning "bad faith" or "intent to deceive"). Jim Abbott, *Roman Deceit: Dolus in Latin Literature and Roman Society, Chapter 2: Aquilius Gallus and the Formulae de Dolo Malo*, at 52 (1997), https://www.academia.edu/8183738/Roman_Deceit_Dolus_in_Latin_Literature_and_Roman_Society_Chapter_Two_Aquilius_Gallus_and_the_Formulae_de_Dolo_Malo; Brendan Brown, *Jurisdictional Basis of Roman Law*, 12 Notre Dame L. 361, 366 (1937), https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4104&context=ndlr; Abdurrahman Savas, *The Process of Transforming Strict Liability into Liability for Fault in Roman Law, and the Effect This Transformation Has Had on Modern Law*, Istanbul Hukuk Mecmuasi, 80 (2), 537-582, DOI: 10-26650/mecmua-2-22.80.2.0006 (2022); H.D.J. Bodenstein, *Phases in the Development of Criminal Mens Rea*, 36 The South African L. J. 323 (1919); Geoffrey MacCormack, *The Liability of the Tutor in Classical Roman Law*, 5 Irish Jurist No. 2, 369-390 (1970), https://www.jstor.org/stable/44027589.

Starting with the rule of Roman Emperor Constantine in the fourth century CE when the Roman Empire converted to Christianity, Christian principles began to infuse Roman law as well. Brendan Brown,

In his famous Commentaries, William Blackstone stated: "[P]unishments are … only inflicted for the abuse of that free will which God has given to man."[96] The U.S. Supreme Court "look[s] primarily to eminent common-law authorities [such as] Blackstone" on topics including mental states in criminal cases.[97]

Nowhere does FW have a greater foundation than in U.S. criminal and civil law.[98] Indeed, FW underpins criminal law's concept of *mens rea*:

> The conception of blameworthiness or moral guilt is necessarily based upon a free mind voluntarily choosing evil rather than good; these can be no criminality in the sense of moral shortcoming if there is no freedom of choice or normality of will capable of exercising a free choice.[99]

"In the American criminal justice system, the dominant justification for punishing individuals is that offenders have made a voluntary choice to break the law, thus validating the imposition of a societal sanction."[100] "Intent" is necessary for a defendant to be guilty of a crime in common law jurisdictions such as the United States.[101] *Mens rea* is the "state of mind statutorily required in order to convict a particular defendant of a particular crime."[102]

The law's view of human behavior as a function of mental states remains rather simplistic:

> Brief reflection should indicate that the law's psychology must be a folk psychological theory, a view of the person as a conscious (and potentially self-conscious) creature who forms and acts on intentions that are the product of the

---

Jurisdictional Basis of Roman Law, 12 Notre Dame L. 361, 366 (1937), https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4104&context=ndlr.

[96] IV William Blackstone, *Commentaries on the Laws of England* 1445 (William Draper Lewis ed., 1902); *accord* Luis Chiesa, Punishing Without Free Will, 2011 Utah L. Rev. 1403, 1404 (2011).

[97] *Kahler v. Kansas*, 140 S. Ct. 1021, 1027 (2020).

[98] Michael Simons, *Criminal Law: The Evolution of Mens Rea: From "Wickedness" to Specific Elements* (2022), https://opencasebook.org/casebooks/2372-criminal-law-simons-volumes-i-and-ii/resources/3.1.2-the-evolution-of-mens-rea-from-wickedness-to-specific-elements-robinson/#:~:text=As%20Prof.,the%20definition%20of%20most%20crimes.

[99] Francis Sayre, *Mens Rea*, 45 Harv. L. Rev. 974, 1004 (1932), https://www.jstor.org/stable/pdf/1332142.pdf?refreqid=fastly-default%3A9e5f3a64fe436f37c0f1bfb8aedca2cf&ab_segments=&origin=&initiator=&acceptTC=1.

[100] Matthew Jones, *Overcoming the Myth of Free Will in Criminal Law: The True Impact of the Genetic Revolution*, 52 Duke Law J., 1031-1053 (2003); *see also* David Ludden, *Can We Have Justice Without Free Will?*, Psych. Today (July 20, 2020) ("Our criminal justice system is based on the assumption of free will").

[101] The word "intent" itself is subject to different meanings and interpretations. David Crump, *What Does Intent Mean?*, 38 Hofstra L. Rev. 1059 (2010).

[102] *Mens rea*, Legal Info. Inst., Cornell L. Sch. (July 2023), https://www.law.cornell.edu/wex/mens_rea#:~:text=Mens%20rea%20refers%20to%20criminal,defendant%20of%20a%20particular%20crime. (last visited Oct. 13, 2023); *see also Mens Rea*: An Overview of State-of-Mind Requirements for Federal Criminal Offenses, Cong. Rsch. Serv. (2021).

> person's other mental states. We are the sort of creatures that can act for and respond to reasons. The law treats persons generally as intentional creatures and not simply as mechanistic forces of nature.[103]

"All of the law's doctrinal criteria for criminal responsibility are folk psychological, beginning with the definitional criteria, what the law terms the elements of crime. The first element of every crime, the voluntary act requirement is defined, roughly, as an intentional bodily movement (or omission in cases in which the person has a duty to act) done in a reasonably integrated state of consciousness. Other than crimes of strict liability, all crimes also require a culpable further mental state, such as purpose, knowledge, or recklessness. All affirmative defenses of justification and excuse involve an inquiry into the person's mental state, such as the belief that self-defensive force was necessary or the lack of knowledge of right from wrong."[104]

More broadly, the roots of "intent" in criminal law are found in the Latin phrase *actus non facit reum nisi mens sit rea* – i.e., "an act does not make a person guilty unless their mind is also guilty."[105] This concept captures the fundamental principle that, for criminal liability to be established, a person must not only commit the act but also possess the necessary intent.

> The mens rea requirement is premised upon the idea that one must possess a guilty state of mind and be aware of his or her misconduct; however, a defendant need not know that their conduct is illegal to be guilty of a crime. Rather, the defendant must be conscious of the "facts that make his conduct fit the definition of the offense." … *Staples v. United States*, 511 US 600 (1994).[106]

---

[103] Stephen Morse, *Neuroscience and the Future of Personhood and Responsibility* (2011), at 116, All Faculty Scholarship. 402. https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1401&context=faculty_scholarship.

[104] Morse, *supra* note 104.

[105] *Mens rea*, Wiktionary (Aug. 29, 2023), https://en.wiktionary.org/wiki/mens_rea (last visited Oct. 13, 2023); *Actus Non Facit Reum Nisi Mens Sit Rea*, LawBhoomi Blog (June 14, 2023), https://lawbhoomi.com/actus-non-facit-reum-nisi-mens-sit-rea/#:~:text=Conclusion-,Actus%20Non%20Facit%20Reum%20Nisi%20Mens%20Sit%20Rea%20is%20a,necessary%20to%20establish%20criminal%20liability.

[106] *Mens rea*, Legal Info. Inst., Cornell L. Sch. (July 2023). https://www.law.cornell.edu/wex/mens_rea#:~:text=Mens%20rea%20refers%20to%20criminal,defendant%20of%20a%20particular%20crime. (last visited Oct. 13, 2023) (quoting *Staples v. United States*, 511 U.S. 600 (1994)).

Criminal offenses that do not require *mens rea* are disfavored, and generally fall into categories that the U.S. Supreme Court has described as "public welfare" or "regulatory offenses." *Staples v. United States*, 511 U.S. 600, 606-7 (1994). "Public welfare" crimes, for example, include those involving the use of obviously dangerous or deleterious substances. *Id*. Strict liability offenses also do not require mens rea because of the nature of the offense (e.g., statutory rape). *How Defendants' Mental States Affect Their Responsibility for a Crime*, NOLO Blog, https://www.nolo.com/legal-encyclopedia/crime-mental-state-defendant-29951.html (last visited Oct. 13, 2023).

The U.S. Supreme Court has stated that the "requirement of some *mens rea* for a crime is firmly embedded" in the common law which, in turn, underpins our jurisprudence.[107] The "intent" requirement is as "universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil."[108]

Mens rea is memorialized in the American Law Institute's Model Penal Code ("MPC"). Completed in 1962, the MPC has been influential in codifying the criminal law of the United States.[109] The MPC does not specifically define "intent" but defines "intentionally or with intent" as "purposely."[110] The MPC thereafter treats "purposely" along a decreasing spectrum of "kinds of culpabilit[ies]" – from "purposely," "knowingly," "recklessly," to "negligently" – and goes on to state that a "person acts purposely with respect to a material element of an offense when", in part, "if the element involves the nature of his conduct or a result thereof, it is his conscious object to engage in conduct of that nature or to cause such a result …."[111] U.S. jurisdictions that do not follow the MPC generally use the related concept of "malice." In the criminal law context, "malice" means the "intention, without justification or excuse, to commit an act that is unlawful."[112]

By the Thirteenth Century, defenses based on the absence of requisite states of *mens rea* began to emerge in the law, all of which still nonetheless were and are based upon assumptions regarding how the human brain functions. These defenses – which include insanity, infancy and compulsion – are premised on the "lack of a guilty mind and thus negating moral blameworthiness."[113] These doctrines include but are not limited to insanity, mistake, justification, and duress.[114] These doctrines, however, start from the assumption of FW and generally provide defenses.

---

[107] *Staples v. United States*, 511 U.S. 600.
[108] *Id*. at 605.
[109] *Model Penal Code FOREWORD*, Lexis+, https://plus.lexis.com (last visited Oct. 13, 2023); *accord* Owen Jones, et al. *Law and Neuroscience*, 2d ed., p. 21 (Wolters Kluwer, 2021) ("the [MPC] … has been widely influential on the mental state definitions in most states").
[110] *Model Penal Code* § 1.13(12), Lexis+, https://plus.lexis.com (last visited Oct. 13, 2023). MPC's "culpable states of minds" approach is due, in part, to challenges and confusion associated with how to describe and define "intent." *Mens rea*, Legal Info. Inst., Cornell L. Sch. (July 2023). Somewhat sardonically perhaps, the commentary to the MPC similarly observes that the related term "willfully" is "unusually ambiguous standing alone." Model Penal Code § 2.02, Commentary, Lexis+, https://plus.lexis.com (last visited Oct. 13, 2023). *See also* Roderick Thomas et al., *Willfully Reinterpreted: The Effect of DOJ's Latest Interpretation of False Statement Statutes on Contractors' Mandatory Disclosure Obligations*, Wiley Newsletter (Spring 2014), https://www.wiley.law/newsletter-4992.
[111] *Model Penal Code* § 2.02(2)(a), Lexis+, https://plus.lexis.com (last visited Oct. 13, 2023). *See also* Owen Jones, et al. *Law and Neuroscience*, 2d ed., p. 17 (Wolters Kluwer, 2021) ("By its taxonomy, culpable mental states [under the MPC] include: purposeful, knowing, reckless, and negligent – in descending sequence of severity, each with importantly different sentencing results").
[112] *Malice*, Legal Info. Inst., Cornell L. Sch., https://www.law.cornell.edu/wex/malice.
[113] Francis Sayre, *Mens Rea*, 45 Harv. L. Rev. 974, 1004 (1932).
[114] *See* Kansas and Powell.

FW also permeates civil law."[115] In civil law, intent may take the form of the state of mind of a person that "either (1) has a purpose to accomplish that result or (2) lacks such a purpose but knows to a substantial certainty that the defendant's conduct will bring about the result."[116] The U.S. Supreme Court has stated that a "belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil [is a belief that is] universal and persistent in mature systems of law."[117]

U.S. law assumes that these and related mental attributes and abilities are largely possessed by humans alone. Considering other life forms, for example, through statutes such as the Endangered Species Act, a variety of non-human species and their habitats are entitled to some amount of protection, but nothing in the law assumes that an animal – let alone a plant or other forms of life -- possesses the consciousness required to exercise FW and possess *mens rea*. Instead, U.S. wildlife law is premised on the notion – again with roots in English common law – that wildlife is the property of the State.

Curiously, U.S. law arguably has been more flexible regarding considering non-human mental conditions in the context of entities, which, in comparison to animals, have no minds and are not otherwise alive – i.e., corporations. Through concepts such as legal "personhood," U.S. law deems corporations to be "legal persons." As "legal persons," corporations can sue, own property, and commit crimes. To get around the fact that corporations, lacking mind themselves and thus incapable of possessing *mens rea*, U.S. courts applied doctrines such as respondeat superior[118] and the collective knowledge doctrine[119] to effectively fabricate and thereafter attribute human mental states to the inanimate object known as a corporation.[120]

---

[115] Ronald Rychlak et al. *Mental Health Experts on Trial: Free Will and Determinism in the Courtroom*, 100 W. Va. L. Rev. 193, 196 n. 6 (1997) (referencing wills, deeds, contracts and confessions).
[116] Dan Dobbs, *The Law of Torts* 447, 448 (2000).
[117] *Morissette v. United States*, 342 U.S. 246, 250 (1952); *see also* United States v. Grayson, 438 U.S. 41 1978) ("A 'universal and persistent' foundation stone in our system of law, and particularly in our approach to punishment, sentencing, and incarceration, is the 'belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil").
[118] Robert Luskin, *Caring About Corporate "Due Care": Why Criminal Respondeat Superior Liability Overreaches Its Justification*, 57 Amer. Crim. L. Rev. 303 (2020), https://www.law.georgetown.edu/american-criminal-law-review/wp-content/uploads/sites/15/2020/03/57-2-caring-about-corporate-due-care-why-criminal-respondeat-superior-liability-outreaches-its-justification.pdf.
[119] Mihalilis Diamantis, *Functional Corporate Knowledge*, 61 Will. & Mary L. Rev. 319 (2019), https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=3832&context=wmlr.
[120] Respondeat superior and/or the collective knowledge doctrine are not uniformly recognized and applied, however. The MPC uses an "inner circle" approach to assessing corporate liability. *Mens Rea: An Overview of State-of-Mind Requirements for Federal Criminal Offenses*, p. 34, R46836, Cong. Rsch. Serv., July 7, 2021). Under federal criminal law, "corporate criminal liability extends to offenses committed by a corporate officer, employee, or agent if acting within the scope of his or her authority at least partly for the benefit of the corporation. Where these conditions are met, the mens rea of the officer, employee, or agent engaging in the proscribed conduct is imputed to the entity for purposes of criminal liability …." *Id*. at 33.

III. AI UNDERMINES THE LEGAL PREMISE THAT HUMAN MENTAL STATES ARE UNIQUE

At minimum, the current generation of LLMs, with their sophisticated written communication abilities, are apt to complicate legal assessments of *mens rea* in profound ways. LLMs will only add to the general level of confusion and inconsistency that occurs under both federal and State law regarding mental state considerations.[121] It is conceivable, and maybe even likely, that human decision-making will be influenced by LLMs in ways that complicate legal assumptions regarding or assessments of the human's mental condition. Under any number of theories – e.g., the ELIZA Effect, the Authority Effect, Animism, and EMT – counsel should be able to argue that their client's mind was unduly influenced by the "mindless" machine. The day may already be at hand when the first question defense counsel should ask her client is "did you consult with your LLM before taking the action?"

Jurors will also be influenced in their deliberations by their exposure to the current generation of LLMs. Issues regarding the specific defendant's mental state, specifically including mens rea, are the exclusive purview of the jury.[122] With each passing day, an increasing number of citizens, and thus potential jurors, will have personal experience with LLMs in the workplace or at home. One might expect such jurors to have increasing levels of sympathy for defendants who were interacting with their LLMs before committing the alleged crime given the extent to which machines can influence human decision-making.

Cases such as *Miller v. Commonwealth*, 492 S.E.2d 482 (Va. Ct. App. 1997) may provide a cautionary tale in this regard particularly when AI systems are used by governmental agencies. In that case, the court recognized an exception to the general rule that "ignorance of the law is no excuse" where the criminal defendant relied upon the advice of a governmental officer before taking the action (here, believing that a felon could possess a muzzle loader). The court concluded that the defendant was not liable because he reasonably relied upon the advice of a state official who had authority to make determinative interpretations of the law. We envision criminal defendants making more such arguments in the future – e.g., "my client queued the search feature of [government Agency's] website, which has a LLM running in the background, and the website advised my client what to do."

Courts will also have to address AI-related liability itself. If a LLM generates an erroneous output that in turn is relied upon by a human with negative resulting consequences, who (if anybody) is responsible? Such cases are already pending, with judges and juries having to decide if a cast of non-machine characters surrounding the machine can fairly be held liable for the machine's output. These characters include the LLM's upstream human programmer(s)

---

[121] *Mens Rea: An Overview of State-of-Mind Requirements for Federal Criminal Offenses*, R46836, Cong. Rsch. Serv., July 7, 2021); *accord United States v. Bannon*, No. 22-3086 (D.C. Cir., June 20, 2024) ("[T]he Supreme Court has … consistently recognized that 'willful[]' … is 'a word of many meanings,' whose contradiction is often dependent on the context in which it appears") (quoting *Bryan v. United States*, 524 U.S. 184, 191 (1998)).

[122] *Diaz v. United States*, No. 23-14 (U.S., June 20, 2024).

and corporate owners, and the downstream human and corporate uses of the technology. Given the way that LLMs function, defendants in such cases argue that they are not liable, because the "machine did it." Through an impenetrable approach that is not even visible to the LLM's program, the defendants will argue that the LLM truly generated the answer such that liability can only attach to the human who, at the end of the day, was responsible for making the decision and taking the action that led to harm.[123]

For the current generation of LLMs, U.S. law may be up to the task of resolving these and related disputes. U.S. law has proven to be malleable to new technologies in the past, and legal considerations of mental states are murky enough, and thus pliable and flexible, to enable judges, juries and attorneys to cope. Questions of *mens rea* will still go the jury, which will have to sort through the ultimate mental state of the human defendant who ultimately was influenced by the LLM's answers. The courts will likely continue to muddle through legal issues at the intersection of human and machine intelligence, such as *mens rea*, as they have done for centuries.

> That the criminal law relies so heavily on subjective mental states to define crimes is both understandable and problematic. Understandable, because the concept that an actor's mental state is relevant to her culpability is so fundamental as to go largely unquestioned. Problematic, because few concepts stymie philosophy more than cognition and the mind. That said, the criminal law has developed and continues to function in a state of contented ignorance as to the epistemic and ontological challenges related to the human mind, relying instead on a folksy psychology that, in truth, we all recognize even where it lacks precision or accuracy.
>
> In this vein, where guilt is predicated on knowingly acting or causing a result, juries are instructed:
>
> > The term 'knowingly", as used in these instructions to describe the alleged state of mind of [the defendant], means that [he][she] was conscious and aware of [his][her][action][omission], realized what [he][she] was doing or what was happening around [him][her], and did not [act][fail to act] because of ignorance, mistake or accident.
>
> That is, knowledge is described as awareness of a fact or set of facts.[124]

Over the centuries, *mens rea* has been modified based on new information, changed societal considerations and other factors.[125] These modifications include the law's general *mens*

---

[123] *See, e.g.*, CS Chaitali Jani & Prof. Dr. S.P. Rathor, *A Legal Framework for Determining The Criminal Liability and Punishment for Artificial Intelligence*, 45 J. of Propulsion Tech. 807 (2024).
[124] Gregory Gilchrist, *Willful Blindness as Mere Evidence*, 54 Loyola of Los Angeles L. Rev. 405, 412 (2021).
[125] *Powell v. Texas*, 392 U.S. 514, 536 (1968) ("The doctrines of actus reus, mens rea, insanity, mistake, justification, and duress have historically provided the tools for a constantly shifting adjustment of the tension between the evolving aims of the criminal law and changing religious, moral, philosophical, and medical views of the nature of man").

rea requirement becoming more specific for individual crimes.[126] "A study of the historical development of the mental requisites of crime leads to certain inescapable conclusions [one of which is] it seems clear that *mens rea* … has no fixed continuing meaning."[127] In specific contexts such as the insanity defense, where "uncertainties about the human mind loom large," the U.S. Supreme Court has specifically declined to adopt rigid, unchanging rules.[128] In recent years, Congress has considered bills that would modify the definition of *mens rea*.[129]

The addition of AI to the mix sits on top of all of the existing complexities regarding the human mind that the courts have been struggling with for centuries. The human's baseline mental state itself falls along a spectrum, now complicated by the human's knowledge of, awareness regarding, or perception of the AI.[130] As noted, humans may view AI as "just another non-sentient" tool. Alternatively, the human may consciously or subconsciously perceive that the AI is infallible or even possessing consciousness itself.[131]

Under all scenarios in the years ahead, it will behoove all participants in the legal process to quickly come up to speed with AI technology generally and LLMs specifically to maximize the chances that a fair, equitable and technically savvy decision emerges from that process.

These considerations will become more acute as AI technologies improve. As noted above: (1) in 2023 researchers reported the use of ToM machine system that, in essence, was

---

[126] Francis Sayre, *Mens Rea*, 45 Harv. L. Rev. 974 (1932), https://www.jstor.org/stable/pdf/1332142.pdf?refreqid=fastly-default%3A9e5f3a64fe436f37c0f1bfb8aedca2cf&ab_segments=&origin=&initiator=&acceptTC=

[127] Sayre, *supra* note 127 at 1016.

[128] *Kahler v. Kansas*, 140 S. Ct. 1021, 1028 (2020).

[129] *See also Mens Rea: An Overview of State-of-Mind Requirements for Federal Criminal Offenses*, Cong. Rsch. Serv. (2021).

[130] Lasagna Harris, *The Neuroscience of Human and Artificial Intelligence Presence*, 75 Annual Rev. of Psych. (Oct. 31, 2023) ("People behave toward humans differently than they do toward AI. Moreover, brain regions more engaged by humans compared to AI extend beyond the social cognition brain network to all parts of the brain, and the brain sometimes is engaged more by AI than by humans"), https://www.annualreviews.org/doi/pdf/10.1146/annurev-psych-013123-123421 (last visited Nov. 4, 2023).

[131] Human Computer Interaction ("HCI"), broadly interpreted, refers to the fields of research and study at the interface of humans and machines. Under the HCI umbrella, research is underway on whether, and if so the extent to which, humans may perceive that AI systems are conscious:

> The prospect of machine consciousness cultivates controversy across media, academia, and industry. Assessing whether non-experts perceive technologies as conscious, and exploring the consequences of this perception, are yet unaddressed challenges in Human Computer Interaction (HCI). To address them, we surveyed 100 people, exploring their conceptualisations of consciousness and if and how they perceive consciousness in currently available interactive technologies. We show that many people already perceive a degree of consciousness in GPT-3, a voice chat bot, and a robot vacuum cleaner.

Ava Scott *et al. Do You Mind? User Perceptions of Machine Consciousness*, Proc. of 2023 CHI Conf. on Human Factors in Computing Systems, Hamburg, Germany (Apr. 2023), https://dl.acm.org/doi/fullHtml/10.1145/3544548.3581296.

able to read or infer the human's mental state and adjust accordingly; and (2) Self-Aware/Conscious machines are under development. Given the murkiness surrounding consciousness generally, attorneys may also have to face their own version of the "hard problem" – namely, what if the AI system at issue is in fact conscious but technologists lack the tools to make such a determination?

We also anticipate that developments in AI will lead to greater understandings of human brain functions in a way that will also be influential for the law. Some next-generation LLMs, for example, are apt to be based on the FEP which, as discussed above, some neuroscientists believe provides the best current model of human brain function.[132] The beta version of such a system was released in June 2024.[133] The day may arrive when scientists conclude that the human brain and machine brain are effectively running the same algorithm, or at least that the machine brain is operating in a manner that is functionally equivalent to a human brain.

Such a scenario presumably would go some ways towards blurring the distinction between humans and AI in legally relevant ways. Again, the ability of current LLMs to communicate is remarkable, a fact that alone starts to chip away at what it means to be a human:

> Human language is unique among all forms of animal communication. It is unlikely that any other species, including our close genetic cousins the Neanderthals, ever had language, and so-called sign 'language' in Great Apes is nothing like human language. Language evolution shares many features with biological evolution, and this has made it useful for tracing recent human history and for studying how culture evolves among groups of people with related languages. A case can be made that language has played a more important role in our species' recent (circa last 200,000 years) evolution than have our genes.[134]

Cases illustrating the blurring of human and machine minds are already before the courts. In all or nearly all cases, the AI's decision-making processes will almost certainly be opaque and "as impenetrable as that of the human brain."[135]

Finally, and more broadly, AI may ultimately start to chip away at some of the foundational underpinnings of the U.S. legal system. Suffice it to say that the crafters of what was to become the foundation of the U.S. legal system never in their wildest imagination contemplated machines that could write, reason, analyze and engaged in similar cognitive functions. U.S. law instead is founded on the assumption, or belief, that Homo sapiens are possess unique skills such as the ability to communicate and have dominion over the Earth.

---

[132] *Dr. Shamil Chandaria: The Bayesian Brain and Meditation*, Centre for Eudaimonia and Human Flourishing, YouTube, https://www.youtube.com/watch?v=WWaYKsUhXqg&t=4621s (last visited June 26, 2024).
[133] https://www.verses.ai/genius.
[134] Pagel, *supra* note 8.
[135] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 Harvard J. of Law & Tech. 899, 894 (1988).

U.S. law is based on Christian concepts of the human "self" or "ego", manifesting itself in consciousness and exercising FW.[136]

The more that machines exhibit traits and abilities that previously were deemed to be within the exclusive domain of humans, the more likely it is that the humans will question the law itself. We agree with Joshua Greene, who made similar observations but from the perspective of neuroscience:

> The rapidly growing field of cognitive neuroscience holds the promise of explaining the operations of the mind in terms of the physical operations of the brain. Some suggest that our emerging understanding of the physical causes of human (mis)behaviour will have a transformative effect on the law. Others argue that new neuroscience will provide only new details and that existing legal doctrine can accommodate whatever new information neuroscience will provide. We argue that neuroscience will probably have a transformative effect on the law, despite the fact that existing legal doctrine can, in principle, accommodate whatever neuroscience will tell us. **New neuroscience will change the law, not by undermining its current assumptions, but by transforming people's moral intuitions about free will and responsibility.** This change in moral outlook will result not from the discovery of new facts or clever new arguments, but from a new appreciation of old arguments, bolstered by vivid new illustrations provided by cognitive neuroscience. We foresee, and recommend, a shift away from punishment aimed at retribution in favour of a more progressive, consequentialist approach to criminal law.[137]

Ultimately, AI systems may call into question what it means to be human from a cognitive perspective, and thus lead society to question whether the U.S. law's rather narrow assumptions regarding human cognition are still capable of being fairly applied when machines seems to be performing as well as us mere mortals on critical tasks such as writing, reasoning, and decision-making.[138]

Similar considerations ultimately may influence environmental policy decisions, including but not limited to those under the Endangered Species Act. If a machine is conscious – or comes close to it, as recognized by the law – perhaps the law will begin to recognize consciousness in non-human species, too.

---

[136] *But see* Stephen Morse, *Neuroscience and the Future of Personhood and Responsibility* (2011), at 117, *All Faculty Scholarship*. 402. https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1401&context=faculty_scholarship ("neuroscience provides no new challenge to criminal responsibility. It cannot prove that determinism is true, and it is simply the determinism du jour, grabbing the attention previously given to psychological or genetic determinism. This challenge is not a problem for criminal law because free will plays no doctrinal role in criminal law and it is not genuinely foundational for criminal responsibility").

[137] Joshua Greene *et al.*, *For the Law, Neuroscience Changes Nothing and Everything*, Phil. Trans. R. Soc. Lond. B (2004) 359, 1775-1785 doi:10.1098/rstb.2004.1546, https://www.jstor.org/stable/4142162.

[138] Deborah Netburn, *Can religion save us from Artificial Intelligence*, Los Angeles Times (Mar. 3, 2023), https://www.latimes.com/world-nation/story/2023-03-03/can-religion-save-us-from-artificial-inte.

CONCLUSION

This Essay argues that that U.S. law and those involved in the profession may be underestimating the impact of AI. AI is not a "mere tool." Instead, through technologies such as LLMs judges, lawyers and juries will face a host of complicated issues at the intersection of human and machine minds. U.S. law, with its anthropogenic roots that start from the premise that humans alone are conscious and otherwise have what are colloquially called "minds," may be flexible enough to sort through conflicts where the argument is made that the machine, which has a mind, influenced the human. U.S. law may not be flexible enough, however, with the result that legislative changes may be required.

Given that much of criminal law is rooted in state law, for example, we may start to see states experiment with legislation that establishes standards for "machine intelligence."

At minimum, it behooves the legal profession to stay abreast of technical developments in AI, with a focus on consciousness. We also recommend that Western attorneys (re)educate themselves on Eastern philosophies such as Buddhism that in some instances arguably provide more nuanced views on cognition, consciousness, sentient beings and similar topics than the law's current Christian-based viewpoints on these topics.