UNITED STATES REPORT ON TRADITIONAL CRIMINAL LAW CATEGORIES AND AI

Alice Giannini[*,†]

1 Definition and legal qualification of Artificial Intelligence system (AI system)

The US has not adopted a legal definition of AI system nor there are court decisions which provide such definition. Furthermore, the US has embraced an agency-by-agency approach to AI regulation rather than a comprehensive one (such as the European approach). The main focus of American efforts has been defense and infrastructure investment, rather than broad regulation. As a consequence, different definitions have been put forward by different state actors ("an alphabet soup of U.S. government agencies").[1]

Generally speaking, US AI-lawmaking activities have been divided into five macro areas:

- *Policies*, including "documents like executive orders, resolutions, and plans that reflect the U.S. government's policies on AI regulation";
- *Accountability*, including "legislative instruments directed to algorithmic accountability, likely reflecting the governments' response to … publicized concerns of algorithmic bias and discrimination";
- *Facial Recognition Technology*, including "the rapidly growing body of law that governs the use of facial recognition technology and associated data";
- *Transparency*, including the body of laws that "are primarily directed to promoting transparency when it comes to the use of AI in different contexts";
- *Other*, including "pending federal bills on general governance or research issues for AI, among other things". [2]

1.1 Federal level and national actors

In October 2016 the National Science and Technology Council (NSTC) and the Executive Office of the President released the "Preparing for the Future of Artificial Intelligence"

*Assistant professor, Department of Criminal law and criminology, Faculty of Law, Maastricht University: a.giannini@maastrichtuniversity.nl.

† The report is updated until April 2022.

[1] Maneesha Mithal, 'Legal Requirements for Mitigating Bias in AI Systems' (*JDSupra*, 14 February 2022) <https://www.wsgrdataadvisor.com/2022/02/legal-requirements-for-mitigating-bias-in-ai-systems/> accessed 3 March 2022.

[2] Yoon Chae, 'U.S. AI Regulation Guide: Legislative Overview and Practical Considerations' (2020) 3 (1) RAIL, 18.

Report, which directly references Russell and Norvig's[3] classification of AI systems, that is:

a. systems that think like humans (e.g., cognitive architectures and neural networks);
b. systems that act like humans (e.g., pass the Turing test via natural language processing; knowledge representation, automated reasoning, and learning),
c. systems that think rationally (e.g., logic solvers, inference, and optimization); and
d. systems that act rationally (e.g., intelligent software agents and embodied robots that achieve goals via perception, planning, reasoning, learning, communicating, decision-making, and acting).[4]

The John S. McCain National Defense Authorization Act for Fiscal Year 2019 ("the Defense Authorization Act of 2019") adopted by the Senate and the house of Representatives on August 13, 2018, provides at sec. 238 (g)) the following definition of AI:

> (g) ARTIFICIAL INTELLIGENCE DEFINED – In this section, the term "artificial intelligence" includes the following:
>
> (1) Any artificial system that performs tasks under varying unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.
>
> (2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.
>
> (3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks.
>
> (4) A set of techniques, including machine learning that is designed to approximate a cognitive task.
>
> (5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision-making, and acting.

The Defense Authorization Act of 2019 also determined that the Secretary of Defense "shall establish a set of activities within the Department of Defense to coordinate the efforts of the Department to develop, mature, and transition artificial intelligence

[3] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, 3rd Edition, 2009).

[4] Executive Office of the President National Science and Technology Council Committee on Technology, *Preparing for the Future of Artificial Intelligence* (2016).

technologies into operational use" (sec. 238, (a)(1)). Amongst the tasks, the Defense Authorization Act of 2019 provides that

> (f) DELINEATION OF DEFINITION OF ARTIFICIAL INTELLIGENCE.—Not later than one year after the date of the enactment of this Act, the Secretary shall delineate a definition of the term ''artificial intelligence'' for use within the Department.

Furthermore, the Defense Authorization Act of 2019 (sec. 1051) also established the National Security Commission on Artificial Intelligence (NSCAI) to "review advances in artificial intelligence, related machine learning developments, and associated technologies". Its purpose was to develop recommendations to the President and Congress on AI. The NSCAI released its final report in March 2021.[5] In the report, the NSCAI urges the US government to get AI-ready by 2025 also by creating a review system for "high-risk" AI systems.

President Trump on February 11, 2019, in its Executive Order (EO) "Maintaining American Leadership in Artificial Intelligence" directed all US Federal agencies to take measures to ensure the American leadership position in AI to be coordinated through the NSTC Select Committee on Artificial Intelligence. The EO also directed the Secretary of Commerce to develop "a plan for Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies" through the National Institute of Standards and Technology (NIST).[6]

In response to the EO, the NIST released "A Plan for Federal Engagement in Developing Technical Standards and Related Tools" on August 9, 2019, where it noted

> While definitions of AI vary, for purposes of this plan AI technologies and systems are considered to comprise software and/or hardware that can learn to solve complex problems, make predictions or undertake tasks that require human-like sensing (such as vision, speech, and touch), perception, cognition, planning, learning, communication, or physical action. Examples are wide-ranging and expanding rapidly. They include, but are not limited to, AI assistants, computer vision systems, biomedical research, unmanned vehicle systems, advanced game-playing software, and facial recognition systems as well as application of AI in both Information Technology (IT) and Operational Technology (OT).

On September 15, 2020, the US Congress adopted the AI in Government Act of 2020, which implements the definition provided by the Defense Authorization Act of 2019. The

[5]The National Security Commission on Artificial Intelligence, *Final Report* (2021).

[6] The NIST is non-regulatory federal agency within the U.S. Department of Commerce. Amongst its functions, it develops standards/guidelines which support federal agencies to comply with the Federal information Security Management Act (FISMA).

act establishes the "AI Center of Excellence" (AI CoE) which shall facilitate the adoption of AI technologies in the Federal Governments, improve cohesion and competency in the adoption and use of AI within the federal government, and carry out such activities for the purposes of benefitting the public and enhancing the productivity and efficiency of federal government operations.

The U.S. congress on January 1, 2021, adopted the National Artificial Intelligence Initiative Act of 2020 (H.R.6216), which followed the National Artificial Intelligence Initiative Act of 2019 (S. 1558). With regards to AI, it establishes significant step towards the development of an American AI policy at federal level, together with the creation of a number of research institutes focused on AI and led by different departments (such as the National Science Foundation, the Department of Energy, the Department of Commerce, NASA and the Department of Defense).

The National Artificial Intelligence Initiative Act of 2021 defines AI at art. 3 (3)

> (3) ARTIFICIAL INTELLIGENCE.—The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to—
> (A) perceive real and virtual environments;
> (B) abstract such perceptions into models through analysis in an automated manner; and
> (C) use model inference to formulate options for information or action.

This definition partly contains the one provided in the OECD Recommendation of the Council on Artificial Intelligence,[7] which, as affirmed recently in the Trade and Technology Council Inaugural Joint Statement of September 29, 2021, has been endorsed by the US.

Following the National AI Initiative Act of 2020, in January 2021 the Office of Science and Technology Policy (OSTP) at the White House established the National AI Initiative Office with the purpose of overseeing and implementing a national AI strategy, and to work as a hub for coordination and collaboration by federal agencies and outside stakeholders across government, industry, and academia in AI research and policymaking.[8] The same act established the National AI Research Task Force which

[7] "AI system: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy". OECD, *Recommendation of the Council on Artificial Intelligence* (2019) OECD/LEGAL/0449.

[8] White House Office of Science and Technology Policy, *The White House Launches the National Artificial Intelligence Initiative Office* (12 January 2021) < https://trumpwhitehouse.archives.gov/briefings-statements/white-house-launches-national-artificial-intelligence-initiative-office/ > accessed January 2022.

should develop a plan to establish a National AI Research Resource (NAIRR). The National AI Initiative Act of 2020 also urged the NIST to develop an "AI Risk Management Framework".

On January 1, 2021, the 116th Congress adopted the Fiscal Year 2021 National Defense Authorization Act, a $731.6 billion defense bill, which draws upon the National Artificial Intelligence Initiative Act of 2020, as well as the 2019 Artificial Intelligence Initiative Act (S. 1558).

On June 8, 2021, the U.S. Senate edited the Innovation and Competition Act (S. 1260) to include the so-called "Advancing American AI Act"[9] provisions. At sec. 4203 of the Advancing American Act it is stated that AI has the meaning given to the term in section 238 (g) of the Defense Authorization Act of 2019. It then further defines "Artificial intelligence systems" as

> (4) Artificial intelligence system.--The term "artificial intelligence system"--
> (A) means any data system, software, application, tool, or utility that operates in whole or in part using dynamic or static machine learning algorithms or other forms of artificial intelligence, whether--
> (i) the data system, software, application, tool, or utility is established primarily for the purpose of researching, developing, or implementing artificial intelligence technology; or
> (ii) artificial intelligence capability is integrated into another system or agency business process, operational activity, or technology system; and
> (B) does not include any common commercial product within which artificial intelligence is embedded, such as a word processor or map navigation system.[10]

The Advancing American AI Act provisions mandate that the Secretary of Homeland Security shall issue policies and procedures for the Department of Homeland Security with the participation of the Chief Procurement Officer, the Chief Information Officer, the Chief Privacy Officer, and of the Officer for Civil Rights and Civil Liberties of the Department and any other person determined to be relevant by the Secretary of Homeland Security. These policies shall include

> B) considerations for *the risks and impacts related to artificial intelligence-enabled systems, including associated data of machine learning systems,* to ensure that full consideration is given to—
> (i) the privacy, civil rights, and civil liberties impacts of artificial intelligence-enabled systems; and

[9] Advancing American AI Act, S. 1260, 117th Cong. (2021), 4201-4207.
[10] Advancing American AI Act 2021, Sec. 4203 (4).

> (ii) security against misuse, degradation, or rending inoperable of artificial intelligence-enabled systems.[11]

The OSTP on November 10, 2021, announced the effort to engage the American public in order to "create a bill of rights for an automated society".[12]

The Department of Defense (DoD) in its AI Strategy of 2018 defined AI as

> … the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems.

The Food and Drug Administration (FDA) in its discussion paper of April 2019 "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback" adopted John McCarthy's 2007 definition of AI:

> It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

1.2 State level

Some efforts have been made at state level.

The State of Nevada attempted at defining AI in the field of autonomous vehicles: " 'Artificial intelligence' means the use of computers and related equipment to enable a machine to duplicate or mimic the behavior of human beings".[13] The statute was repealed in 2013.

The State of Louisiana defines "electronic agents" as "a computer program or an electronic or other automated means used independently to initiate an action or respond to electronic records or performances in whole or in part without review or action by an individual".[14] In an annotation to the Louisiana Uniform Electronic Transaction Act the drafters distinguish between different types of electronic agents:

> 5. "Electronic agent."
> (…).

---

[11] Advancing American AI Act 2021, Sec. 4204 (b)(1)(B).

[12] White House Office of Science and Technology Policy, *Join the Effort to Create A Bill of Rights for an Automated Society* (10 November 2021) < https://www.whitehouse.gov/ostp/news-updates/2021/11/10/join-the-effort-to-create-a-bill-of-rights-for-an-automated-society/ > accessed January 2022.

[13] NEV. REV. STAT. 482A.020 (repealed 2013).

[14] LA. STAT. ANN. § 9:2602 (2018).

> (b) An electronic agent, such as a computer program or other automated means employed by a person, is a tool of that person. As a general rule, the employer of a tool is responsible for the results obtained by the use of that tool since the tool has no independent volition of its own. However, an electronic agent, by definition, is capable within the parameters of its programming, of initiating, responding or interacting with other parties or their electronic agents once it has been activated by a party, without further attention of that party.
> (c) While this Chapter presupposes that an electronic agent is capable of performing only within the technical strictures of its preset programming, it is conceivable that in the future, electronic agents may be created with the ability to act autonomously, and not just automatically. That is, through developments in artificial intelligence, a computer may be able to "learn through experience, modify the instructions in their own programs, and even devise new instructions." If these developments occur, the courts may construe the definition of electronic agent accordingly, to recognize such new capabilities.

According to Martinez, the statute's annotation "articulates the differences between weak and strong AI" and "denotes legislative intent for how to qualify computer agents as either strong AI or weak AI machines".[15]

1.3 Definition elaborated by scholars (e.g., in the field of criminal law, civil law, administrative law, labor law)

In 2016, the One Hundred Year Study on Artificial Intelligence, an initiative of Stanford University, released it first report, where they adopted a working definition of AI: "AI can also be defined by what AI researchers do. This report views AI primarily as a branch of computer science that studies the properties of intelligence by synthesizing intelligence".[16]

In the 2021 report, the authors provide an alternative definition:

> An alternative definition is that artificial intelligence is about getting a machine to carry out behaviors that we think of as requiring intelligence. This view is useful in that it doesn't put a great deal of emphasis on the specifics of the machine or the technique used to create the behavior. It also captures an important yet frustrating aspect of artificial intelligence—once a machine can carry out a behavior, we tend to stop thinking of it as something that requires intelligence. Real-time navigation aids that decide when and how to describe upcoming turns to guide you to your destination are not thought of as AI, even within the field. But there's no question that it would have been considered an

[15] Rex Martinez, 'Artificial Intelligence: Distinguishing Between Types & Definitions' (2019) 19 NEV. L.J. 101, 1032.
[16] Stanford's "One Hundred Year Study on Artificial Intelligence" project (AI100), 'Artificial Intelligence and Life in 2030' (2016) 13.

AI problem just a few decades ago. This phenomenon is known as the "AI Effect," as mentioned in the 2016 report.[17]

Scherer adopts the following definition: "artificial intelligence" refers to machines that are capable of performing tasks that, if performed by a human, would be said to require intelligence".[18] The author himself notices that this entails defining AI "in a blissfully circular fashion".[19]

Some authors focus exclusively on robots.[20] For example, Ryan Calo in "Robotics and the Lessons of Cyberlaw" delivers a definition of *robots* that is based on the so-called sense-think-act paradigm: robots are defined as "mechanical objects that take the world in, process what they sense, and in turn act upon the world".[21] In his understanding, a "[robotic] system acts upon its environment to the extent that it changes that environment directly … It must *be* in some way. A robot in the strongest, fullest sense of the term exists in the world as a corporeal object with the capacity to exert itself physically … robots are best thought of as artificial objects or systems that process, and act upon the world to at least some degree".[22] Robots, according to Calo, express three essential qualities: embodiment, emergence, and social valence.

In the field of *criminal legal scholarship,* Abbott and Sarch in "Punishing Artificial Intelligence: Legal Fiction or Science Fiction"[23] adhere to the first definition of AI of 1995 given by McCarthy et. al: "[T]he artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving".[24]

1.3.1 Legal personhood and legal capacity

AI is not considered as a person or as possessing legal capacity, nor are there any efforts on part of the national legislator to grant AI such status.

The academic scene on the topic is rich in the variety of opinions.

---

[17] Stanford's "One Hundred Year Study on Artificial Intelligence" project (AI100), 'Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report' (2021) 78.

[18] Matthew U. Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29 (2) Harvard Journal of Law & Technology, 362.

[19] Ibid.

[20] E.g., Ying Hu, 'Robot Criminals' (2019) 52 U. Mich. J. L. Reform 487; Christina Mulligan, 'Revenge Against Robots' (2018) 69 South Carolina Law Review 579.

[21] Ryan Calo, 'Robotics and the Lessons of Cyberlaw' (2015) 103 CALIF. L. REV., 529.

[22] ibid 531.

[23] Alexander Sarch and Ryan Abbott, 'Punishing Artificial Intelligence: Legal Fiction or Science Fiction'(2019) UC Davis Law Review 377.

[24] John McCarthy and others, 'A Proposal for the Dartmouth Summer Research Project On Artificial Intelligence' (1955).

Lawrence B. Solum, in his landmark essay "Legal Personhood for Artificial Intelligences" of 1992,[25] conducted a thought experiment in order to transform the theoretical question on whether an AI could become a legal person into a legal one. Specifically, he explored the hypothetical scenarios of appointing AI as a legal trustee and of AI invoking the individual rights provided by the US Constitution. He theorized the introduction of a legal Turing Test which would be applied by courts to determine whether an AI can stand trial as a stand-alone legal agent. He focused on three objections to the idea of recognizing rights to artificial agents, or intelligences (AIs), namely, "AIs Are Not Human"; "The Missing-Something Argument"; and "AIs Ought to Be Property". He did not exclude the possibility of granting legal personhood to AI, but he noted that "the answer to the personhood question is likely to be found two places - in our experience with AI and in our best theories about the underlying mechanisms of the human mind".[26]

In 1994, California Superior Court Judge Curtis Karnow[27] proposed the introduction of a new legal subject called "electronic persona" (*eper*), based on an analogy between corporations and agents.[28]

Chopra and White adopt a pragmatic approach to the issue, which is similar to Solum's approach. They argue that "[w]hile artificial agents are not yet regarded as moral persons, they are coherently becoming subjects of the intentional stance, and may be thought of as intentional agents".[29] It follows, that "[a]n artificial agent with the right sorts of capacities—most importantly, that of being an intentional system—would have a strong case for legal personality … There is no reason in principle that artificial agents could not attain such a status, given their current capacities and the arc of their continued development in the direction of increasing sophistication."[30] In any case, they believe that the decision on whether to confer or not legal personhood to an AI system will be based on economic considerations/utilitarian arguments discussing the benefits vs. the estimated costs. They state that such a system for recognizing legal personality upon AI systems "may need to be set out by legislatures, perhaps through a registration system or "Turing register".[31]

---

[25] Lawrence B. Solum, 'Legal Personhood for Artificial Intelligences'(1992) 70 N.C. L. Rev. 1231.

[26] ibid 1285.

[27] Curtis E.A. Karnow, 'The Encrypted Self: Fleshing out the Rights of Electronic Personalities'(1994) XIII Journal of Computer and Information Law.

[28] See also Jean-Francois Lerouge, 'The Use of Electronic Agents Questioned under Contractual Law: Suggested Solutions on a European and American Level'(2000) 18 The John Marshall Journal of Computer and Information Law 403; Emily M. Weitzenboeck, 'Electronic Agents and the Formation of Contracts' (2001) 9 International Journal of Law and Information Technology 204.

[29] Samir Chopra and Laurence F. White, 'A Legal Theory for Autonomous Artificial Agents' (2011) University of Michigan Press, 189.

[30] Ibid.

[31] Chopra and White (n 29) 190.

Bryson, Diamantis, and Grant[32] claim that even though it would be feasible to recognize legal personhood upon a machine, it would also be "morally unnecessary and legally troublesome".[33] The authors argue that this legal fiction would create asymmetries in particular legal systems, together with a "legal black hole" in terms of accountability for damages, and that could lead to abuses at the expenses of existent legal persons.

In the field of American *business law,* Bayern[34] argues that it is possible to grant legal personality to an autonomous system in the US by putting it in control of a limited liability corporation (LLC). The LLC would then serve as a container (or a "legal alter ego"[35]) of autonomous systems (such as computer programs or robots) and it would obtain an "effective" legal status without leading to extensive legal reform. He distinguishes "effective" legal personhood from "real legal personhood". According to his proposal, "autonomous systems have neither legal "equality" with humans nor any direct, de jure legal personhood. They can simply operate or maneuver a legal person to achieve arbitrary legal ends".[36]

In the field of *tort law*, Scherer suggests the idea to "establish something akin to the legal fiction of corporate personhood, where AI systems would be capable both of owning assets and of being sued in court".[37] In a similar fashion, Vladeck contends that "[c]onferring "personhood" on these machines would resolve the agency question; the machines become principals in their own right, and along with new legal status would come new legal burdens, including the burden of self-insurance".[38]

In the field of *criminal law,* Hu argues in favor of recognizing legal personhood to "smart" robots, provided that they satisfy three conditions (to be held criminally liable): one, the smart robot must be equipped with a "moral algorithm" (i.e., one capable of making nontrivial morally relevant decisions); two, the smart robot must be capable of communicating its moral decisions to humans (including the options of action available prior to the decision, the weight placed on each available course of action and the chosen course of action); three, the smart robot must be able and allowed to act without immediate human supervision.[39] She directly addresses, and refutes, Ryan Calo's view.[40] Calo claims that recognizing legal personhood to robots in the criminal law field would lead to the antromorphizing of robots by the population. Indeed, as recognized by Hu

[32] Joanna J. Bryson, Mihailis E. Diamantis, Thomas D. Grant, 'Of, for, and by the people: the legal lacuna of synthetic persons (2017) 25 Artif Intell Law 273–291.
[33] ibid 289.
[34]Shawn Bayern, 'The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems'(2015) 19 Stanford Technology Law Review 93.
[35] ibid 112.
[36] Bayern (n 34) 112.
[37] Scherer (n 18) 399.
[38] David C. Vladeck, 'Machines Without Principals: Liability Rules and Artificial Intelligence'(2014) 89 Wash. L. Rev. 117, 150.
[39] Hu (n 20).
[40] Calo (n 21).

herself, Calo does not argue that these objections are sufficient to reject the idea of recognizing legal capacity to robots.

Abbott and Sarch claim that "any sort of legal personhood for AIs would be a dramatic legal change that could prove problematic" as it would lead to increased antromorphism of AI systems and consequently to higher expectations on AI capabilities.[41]

1.4 Regulating AI applications: preferred approach (general vs. sectoral)

The preferred approach is a sectoral-agency-by-agency approach. There is no general approach such as the one adopted by European legislators, but more a "patchwork of regulatory approaches".[42]

Examples of (proposed) agency level regulation/tools/guidelines are:

- The Food and Drug Administration (FDA) discussion paper of April 2019 on "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback";
- The Defense Innovation Unit (DIU) of the DoD launched the "Responsible AI Guidelines" on November 14, 2021. They provide step-by-step guidance for implementing the DoD Ethical Principles for AI in the whole development cycle of AI systems for military uses;
- The NIST, upon indications of the U.S. Congress, is developing the Artificial Intelligence Risk Management Framework (AI RMF). The AI RMF is meant as a tool to help designers, developers, users and evaluators of AI systems better manage risks across the AI lifecycle. In March 2022 NIST also published a paper titled "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" where it outlines its approach to identify and managing AI biases;
- The Federal Trade Commission (FTC) is taking steps towards AI rulemaking. On April 8, 2020, it released a set of guidelines on "Using Artificial Intelligence and Algorithms" which focus, amongst other things, on the validation and quality of datasets. On April 19, 2021, it released another set of guidelines on "Aiming for truth, fairness, and equity in your company's use of AI" where it identifies *three laws that are relevant for developers and users of AI (and which are enforced by the FTC*):
    - Section 5 of the FTC Act: prohibits unfair or deceptive practices;

[41] Abbott and Sarch (n 23) 377.

[42] Dan Reilly, 'White House A.I. director says U.S. should model Europe's approach to regulation' (*Fortune* 10 November 2021) < https://fortune.com/2021/11/10/white-house-a-i-director-regulation/ > accessed October 2022.

- Fair Credit Reporting Act: it would come into play in circumstances where an algorithm is used to deny people employment, housing, credit, insurance, or other benefits;

- Equal Credit Opportunity Act (ECOA): prohibits a company to use a biased algorithm that results in credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, or because a person receives public assistance.

On December 10, 2021, it filed an "Advanced Notice of Proposed Rulemaking" (ANPRM) regarding the consideration of initiating a rulemaking process on privacy and artificial intelligence).

1.5 Autonomous driving

With regards to Autonomous Vehicles (AVs), the Department of Transportation (USDOT) released on January 11, 2021, the Automated Vehicles Comprehensive Plan. One of its purposes is to modernize existing regulations to remove unintended and unnecessary barriers to innovation in the field of AVs.

The National Highway Traffic Safety Administration (NHTSA), an agency part of the USDOT has been very active[43] in the field of Automated Driving Systems (ADS).[44] On March 20, 2020, it delivered a NPRM (F.R.1764) relating the development of vehicles equipped with ADS. The same agency, on November 19, 2020, published the ANPRM "Framework for Automated Driving System Safety". The purpose of the framework is to objectively define, assess, and manage the safety of ADS performance while ensuring the needed flexibility to enable further innovation. On June 29, 2021, the NHTSA issued a Standing General Order requiring manufacturers and operators of vehicles equipped with SAE Level 2 advanced driver assistance systems, or SAE Levels 3-5 automated driving systems, to report crashes that occur on public roads in the US (in cases where the systems were engaged during or immediately before the crash). On March 10, 2022, the NHTSA published a ruling which amends the Federal Motor vehicle standards to

[43] See Department of Transportation - NHTSA, *Removing Regulatory Barriers for Vehicles With Automated Driving Systems Request for Comment* (2018) 83 FR 2607; NHTSA, *Removing Regulatory Barriers for Vehicles With Automated Driving Systems Advance Notice of Proposed Rulemaking* (2019) 84 FR 24433; NHTSA, *Occupant Protection for Automated Driving Systems Notice of Proposed Rulemaking* (2020) 85 FR 17624.

[44] "ADS, as defined by SAE International and as used in this notice, refers to driving automation Levels 3-5. SAE International J3016_201806 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On Road Motor Vehicles. An ADS is the hardware and software that are, collectively, capable of performing the entire dynamic driving task on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD). In less technical terms, an ADS maintains the control and driving functions within the situations that the system is designed to operate in", Framework for Automated Driving System Safety, 5.

account for future vehicles that will not display traditional manual controls associated with a human driver, as long as they comply with safety regulations.[45]

Different regulations on self-driving vehicles have been adopted at state level.[46] No state provides for a ban of AVs. Most of these laws regulate the testing of self-driving vehicles, the role of the human operator (if present) and the vehicle authorization procedure. As of 2021 the following states require that self-driving cars have a human operator in the vehicle: Connecticut, District of Columbia, Illinois, Massachusetts, New Hampshire, New York, Vermont and Washington D.C. Florida, Georgia, Nebraska, Nevada, North Carolina, North Dakota, Pennsylvania and Washington connect the requirement of a human operator the level of automation displayed by the vehicle. Some states allow for self-driving vehicles to operate without a human operator in the vehicle upon the fulfillment of specific conditions (Georgia, Iowa, Louisiana, Michigan, Texas).

1.5.1 Areas in which complete automated and autonomous decision-making processes carried out by AI systems are forbidden (or proposals in that direction)

Federal level

With regards to *facial recognition and biometric technologies*, on February 12, 2020, Sen. Jeff Merkley introduced the bill "Ethical Use of Facial Recognition Act" (S. 3284) to the U.S. Senate. The bill would prohibit any officer, employee, or contractor of a federal agency from engaging in specified activities with respect to facial recognition technology without a warrant until a congressional commission established by this bill recommends rules governing the use and limitations on both government and commercial use of the technology. Forbidden activities include setting up a camera to be used in connection with facial recognition technology, accessing or using information obtained from such technology, or importing such technology to identify an individual in the United States

On June 15, 2021, a group of senators from the Democratic party introduced to the U.S. Senate the Facial Recognition and Biometric Technology Moratorium Act of 2021 (S.2052). The Moratorium aims to prohibit biometric surveillance by the Federal Government without explicit statutory authorization and to withhold certain Federal public safety grants from State and local governments that engage in biometric surveillance.

State level

The state of Virginia on July 1, 2021, adopted a bill (H.B. 2031) providing a ban on the use of facial recognition technology by law enforcement. According to the law, "*No local law-enforcement agency shall purchase or deploy facial recognition technology unless the locality*

[45] Department of Transportation – NHTSA, *Occupant Protection for Vehicles With Automated Driving Systems* (2021) 49 CFR Part 571 Docket No. NHTSA-2021-0003 RIN 2127-AM06

[46] A complete overview is available on the website of the National Conference of State Legislature: < https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx >.

in which the law-enforcement agency is located has adopted an ordinance authorizing the use of facial recognition technology by local law-enforcement agencies" (art. 1, B). Similar provisions have been adopted by the state of Maryland, regarding the use of these technology for the purpose of creating a facial template during an applicant's interview for employment (H.B. 1202); by the state of Washington, regarding the use of facial recognition by the government (S.B. 6280), and California (A.B. 1215).

On February 8, 2021, the state of Washington introduced a new bill (S.B. 5116) to establish guidelines for government procurement and use of automated decision systems in order to protect consumers, improve transparency, and create more market predictability. If adopted, the bill would prohibit Washington public agencies from developing, procuring, or using an automated decision system[47] that discriminates against an individual, or treats an individual less favorably than another, in whole or in part, on the basis of a number of factors (enucleated in the Washington Law against discrimination, RCW 49.60.010) such as race, creed, color, national origin, sex, sexual orientation and age. It would also prohibit these actors from developing, procuring, or using an automated final decision system[48] to make a decision impacting the constitutional or legal rights, duties, or privileges of any Washington resident, or to deploy or trigger any weapon. Similarly, a public agency may not operate, install, or commission the operation or installation of equipment incorporating AI-enabled profiling in any public place or use AI-enabled profiling to make decisions that produce legal effects or similarly significant effects concerning individuals.

City level

Several cities adopted ordinances to ban the use of facial recognition technology by law enforcement, for example the "Stop Secret Surveillance" ordinance in the city of San Francisco (Ordinance No. 103-19), effective 31 May 2019.

2 Existing criminal offences and criminalization

2.1 Offences already applicable to illegal acts involving AI systems

From a theoretical perspective, existing criminal offenses are applicable to acts involving AI systems either as victims or means to commit a crime. In other words, offenses cannot be committed *by* AI systems themselves as criminal law is currently only applicable to

[47] "(6)(a) "Automated decision system" means any algorithm, including one incorporating machine learning or other artificial intelligence techniques, that uses data-based analysis or calculations to make or support government decisions, judgments, or conclusions that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of governmental interaction with that individual including, without limitation, benefits, protections, required payments, penalties, regulations, timing, application, or process requirements".

[48] "(7) "Automated final decision system" means an automated decision system that makes final decisions, judgments, or conclusions without human intervention".

human or corporate entities. There have been cases (discussed below) which have triggered the application of criminal law specifically in the field of automated driving.

2.1.1 Automated vehicles

According to Westbrook, a useful tool could be found in the jurisprudence and regulation in the field of *airplane autopilot systems*.[49] In a Minnesota Supreme Court case of 2012 (*Glorvigen v. Cirrus Design Corp*.), the court discarded compensations claims of the plaintiff against an airplane manufacturer arguing that "the manufacturer of a sophisticated technology, even one that purports to fly an airplane on autopilot, may not be held liable for failing to adequately train the operator on its use".[50] Westbrook argues that "AV technology manufacturers will face the same issues regarding the adequacy of warnings, with the added potential liability incurred when the product autonomously breaks the law".[51] He contends that, differently from Glorvigen, courts should recognize a "basic duty for car manufacturers to either accept liability for intent-based and strict-liability crimes or to lose the option to advertise vehicles that "drive themselves".[52]

The killing of Elaine Herzberg – Arizona (2018)

Elaine Herzberg was struck and killed by an automated Uber test vehicle transporting a human operator. The car failed to recognize whether Herzberg was a pedestrian, a vehicle or a bicycle. As reported by the Vehicle automation report redacted by the National Transportation Safety Board,[53] the ADS sensed the pedestrian 6 seconds before the impact but it never classified her as such – or predicted correctly her goal as a jaywalking pedestrian or cyclist – and its design did not include a consideration for jaywalking pedestrians at all. The ADS only determined that Herzberg was on the path of the vehicle 1.2 seconds before the impact; it recognized an emergency situation and the imminent collision. Per design, the system did not engage in emergency brakes but alerted the vehicle operator and initiated a plan for the vehicle to slow down.

The report concludes that the probable cause of the crash was the failure of the vehicle operator to monitor the driving environment and the operation of the automated driving system because she was visually distracted throughout the trip by her personal cell phone.

Contributing to the crash were the Uber Advanced Technologies Group's (1) inadequate safety risk assessment procedures, (2) ineffective oversight of vehicle operators, and (3)

---

[49] Clint W. Westbrook, 'The Google Made Me Do It: The Complexity of Criminal Liability in the Age of Autonomous Vehicles' (2017) 97 Mich. St. L. Rev. 97, 142.
[50]ibid 117.
[51] Westbrook (n 52) 117.
[52] ibid 142.
[53] National Transportation Safety Board, *Accident Report 'Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018* (2018) NTSB/HAR-19/03 PB2019-101402.

lack of adequate mechanisms for addressing operators' automation complacency—all a consequence of its inadequate safety culture.

In letter of March 4, 2019, the Yavapai County Attorney stated that there was no basis for criminal liability for the Uber corporation arising from the matter.[54] Thus, the back-up driver in the Uber car was charged and indicted with a count of negligent vehicular homicide by a Maricopa County Grand Jury on August 27, 2020 (A.R.S. §§ 13-1101, 13-1102, 28-3001, 28-3004, 28-3005, 28-3315, 13-701, 13-702, and 13-801). The indictment states that the offense is to be considered as a dangerous felony, in violation of A.R.S. §§ 13-105 and 13-704.

Negligent homicide is regulated in the Arizona Criminal Law Statute at §13-1102:

> A. A person commits negligent homicide if with criminal negligence the person causes the death of another person, including an unborn child.
> (…).
> C. Negligent homicide is a class 4 felony.

According to §13-105,

> (d) "Criminal negligence" means, with respect to a result or to a circumstance described by a statute defining an offense, that a person fails to perceive a substantial and unjustifiable risk that the result will occur or that the circumstance exists. The risk must be of such nature and degree that the failure to perceive it constitutes a gross deviation from the standard of care that a reasonable person would observe in the situation.

According to §13-702, the term of imprisonment for a class 4 felony shall fall within the following range:

Mitigated	Minimum	Presumptive	Maximum	Aggravated
1 year	1.5 years	2.5 years	3 years	3.75 years

In Arizona, the crime of "vehicular homicide" is not regulated by a specific article in the criminal law statute. In the Herzberg Case, vehicular negligent homicide was considered a class 4 dangerous felony offense, hence the applicable terms of imprisonment are the ones proscribed at §13-704:

Minimum	Presumptive	Maximum
4 years	6 years	8 years

[54] The letter is available at <https://s3.documentcloud.org/documents/5759641/UberCrashYavapaiRuling03052019.pdf>.

The Riad's case – Los Angeles (2019)

News report that in October 2021 prosecutors of the city of Los Angeles filed a felony complaint against Kevin George Aziz Riad for two counts of vehicular manslaughter.[55] Riad was the driver of a Model S Tesla that, while in autopilot mode, crashed with another car, killing the two passengers inside it. Riad's prosecution differs from the one in the Elaine Herzberg case since the vehicle was commercially available (and not under testing). Hence, it targets driving technology that is already accessible and used by the public. Admittedly, Riad's prosecution could work as a precedent to many other instances of prosecuting the human pilot for his overreliance on an autopilot system.

Vehicular manslaughter is a criminal offense regulated in the California Penal Code at § 192 (c)(2):

> 192**.** Manslaughter is the unlawful killing of a human being without malice. It is of three kinds:
> (a) Voluntary—upon a sudden quarrel or heat of passion.
> (b) Involuntary—in the commission of an unlawful act, not amounting to a felony; or in the commission of a lawful act which might produce death, in an unlawful manner, or without due caution and circumspection. This subdivision shall not apply to acts committed in the driving of a vehicle.
> (c) Vehicular—
> (1) Except as provided in subdivision (a) of Section 191.5, driving a vehicle in the commission of an unlawful act, not amounting to a felony, and with gross negligence; or driving a vehicle in the commission of a lawful act which might produce death, in an unlawful manner, and with gross negligence.
> (2) Driving a vehicle in the commission of an unlawful act, not amounting to a felony, but without gross negligence; or driving a vehicle in the commission of a lawful act which might produce death, in an unlawful manner, but without gross negligence.
> (…).

2.1.2 Other AI applications

According to Abbott and Sarch, instances where responsibility for harmful Al conduct is reducible to the culpable conduct of an individual human actor, such as the case of a hacker using Al to steal funds from individual bank accounts, could fall into existing criminal offenses (*fraud or computer crimes*).[56] In other words, AI systems could be considered as the *object* of a criminal offense.

---

[55]Tom Krisher and Stefanie Dazio, *LA Times* (Los Angeles, 18 January 2022) < https://www.latimes.com/california/story/2022-01-18/felony-charges-are-first-in-fatal-crash-involving-teslas-autopilot > accessed February 2022.

[56] Sarch and Abbott (n 23) 369.

2.1.3 *Computer Fraud and Abuse Act (CFAA), 18 U.S.C. §1030*

Add Acts involving AI systems could fall under the scope of application of the CFAA.

The CFAA establishes criminal liability for conducts of access to a computer without authorization (i.e., hacking) or of excess of authorization access. In Van Buren,[57] the U.S. Supreme Court established that "an individual 'exceeds authorized access' when he accesses a computer with authorization but then obtains information located in particular areas of the computer — such as files, folders, or databases — that are off limits to him".[58] The CFAA provides for seven criminal offenses:

Table 1. Summary of CFAA Penalties

Offense	Section	Sentence*
Obtaining National Security Information	(a)(1)	10 (20) years
Accessing a Computer and Obtaining Information	(a)(2)	1 or 5 (10)
Trespassing in a Government Computer	(a)(3)	1 (10)
Accessing a Computer to Defraud & Obtain Value	(a)(4)	5 (10)
Intentionally Damaging by Knowing Transmission	(a)(5)(A)	1 or 10 (20)
Recklessly Damaging by Intentional Access	(a)(5)(B)	1 or 5 (20)
Negligently Causing Damage & Loss by Intentional Access	(a)(5)(C)	1 (10)
Trafficking in Passwords	(a)(6)	1 (10)
Extortion Involving Computers	(a)(7)	5 (10)

* The maximum prison sentences for second convictions are noted in parentheses.

Figure 1. Source: U.S. Department of Justice, Computer Crime and Intellectual Property Section (CCIPS), Prosecuting Computer Crimes Manual, 2010.

Conspiracy and attempt to commit these offenses is also considered a crime (18 U.S.C. § 1030(b)).

2.1.4 *Electronic Communications Protection Act ("ECPA"), 18 U.S.C. § 2702*

Conducts of intentional access without authorization (i.e., hacking) (or exceed authorized access) to a facility that provides an electronic communication service are a criminal violation under the ECPA. The ECPA also punishes intentionally intercepting electronic communications in transit under the Wiretap Act (18. US.C. § 2511).

Conclusively, the following conducts could constitute a criminal offense under federal law, whether committed through or against and AI system: phishing, infection of IT

[57] *Van Buren v. United States* 141 S. Ct. 1648, 1652 (2021).
[58] ibid 1662.

systems with malware, identity theft or identity fraud (Federal identity Theft Statute, 18 U.S.C. § 1028); electronic theft; unsolicited penetration testing.

3 New offences

3.1 New offences related to designing, programming, developing, producing, functioning or making use of AI systems

No specific law (both at federal and state level) has been introduced with regards to conducts of designing, programming, developing, producing, functioning or making use of AI systems.

Westbrook argues that a new kind of liability – called *products culpability* – should be introduced in the field of AV to hold manufactures criminally liable. According to his theory, the passive human-operator-passenger should not be liable when the AV breaks the law.[59] With regards to *actus reus*, Westbrook argues that neither the human operator, nor the AV, or the programmer of the AV software would satisfy the *actus reus* element of a crime committed by the AV. Yet, the law should hold the AV and its manufacturer liable through the "products culpability cause of action". In cases of *strict-liability criminal infractions* (i.e., offenses that only require the *actus reus* element), new legislation that holds the manufacturer financially liable shall be introduced, provided that it is ascertained that the violation of the norm was caused by an error in the hardware or software. The AV shall be considered as an *agent* of the manufacturer. With regards to *intent-based crimes*, the author argues that "the mens rea element of the crime will nearly always be missing, resulting in no sustainable criminal charges".[60] With regards to *negligence-based crimes*, the author suggests holding the manufacturer of the AV liable only in cases where the manufacturer breached its duty of care. The breach could be proven by examining the programming, the hardware and the foreseeability of the malfunction, provided that the burden of proof lies with the State.

3.2 New offences related to acts committed through or against an AI system

No specific law (both at federal and state level) has been introduced with regards to criminal acts committed through or against an AI system.

Abbott and Sarch propose the creation of a *new constructive liability offense* called "Causing Harm Through Criminal Uses of AI"[61] for situations where a base crime is committed and then AI system commits a further crime (see. Q. E (3)). The "Causing Harm Through Criminal Uses of AI" model cannot be applied where an irreducible AI crime (i.e., when the crime cannot be reduced to a human agent) is committed. There would be no base crime, that is, an "underlying culpable conduct by the programmers and users of the AI", out of which to construct the liability of the human agent for the

[59] Westbrook (n 52) 127.
[60] ibid 136.
[61] Sarch and Abbott (n 23) 373.

unforeseeable harms that the Al caused.[62] For these reasons, the authors hypothesize the introduction of new *negligence crimes* directed at developers. The new offense would punish the conduct of developing systems that "foreseeably could produce a risk of *any* serious harm or unlawful consequence, even if a specific risk is unforeseeable".[63] The authors abandon this option for two reasons: first, every technology activity involves the risk of some kind of harm, hence it would be hard to identify an individually culpable conduct; second, the expansion of criminal law would hinder technological advancement.

The same authors argue in favor of a minimal extension of criminal law:

- *Introduction of an "AI Abuse Act"* which would criminalize humans for:
  - Conducts of malicious or reckless uses of AI;
  - Conducts of failing "responsibly design, deploy, test, train, and monitor the AIS one contributed to developing";[64]
- Introduction of the concept "*Responsible Person*", i.e., a "designated adjacent person" which should be registered in a federal registration system. As it is hard to determine which AI systems are capable of causing harm, and which are not, the authors hypothesize the designation of a Responsible Person for any AI system. The designation would work by default. This would entail:
  - That the Responsible Person (a natural person or a corporation) could be
    i. The *AI's manufacturer or supplier* if the AI system is a commercial product;
    ii. The *AI's owner, or developer if no owner exists, or user if no developer can be identified* if the AI system is non-commercial product;
    iii. (Only for cases of a non-commercial AI system without an identifiable owner, developer and user) an *ex-ante* designated Responsible Person.
  - The creation of new forms of criminal negligence such as failing to discharge (new) statutory duties of supervision and care (even through strict criminal liability constructs).

[62] Sarch and Abbott (n 23) 374.
[63] ibid.
[64] Sarch and Abbott (n 23) 378.

i. The Responsible Person would be liable also for harms caused by an AI system where the AI, if a natural person, would be criminally liable together with another individual. This entail that the Responsible Person would be liable also for acting as a *co-conspirator* of the AI system.

ii. Imposing strict criminal liability on the Responsible Person could be justified if the relevant punishments were only monetary (i.e., fines) and considering that states have a duty to society to provide special assurances that certain serious risks will be mitigated as much as possible. Nevertheless, the authors are not convinced that Hard AI Crime is significant enough to justify the application of strict criminal liability constructs to natural persons yet. Specifically, strict liability should be used only as a last resort tool for *unusually dangerous activities* – where instead there are areas where it would be unreasonable not to use AI systems because they would reduce the dangerousness of a certain activity.

- The creation of a new criminal offense (similar to the offense of driving without a license) which would punish programmers, developers, owners or users of an AI system capable of causing harm who did not registering *ex ante* a Responsible Person.

The authors believe that introduction of a Responsible Person regime could achieve the same (expressive) benefits as the conviction of an AI system, with less (legal fiction) costs.

3.3 Positive obligations for persons and/or legal person designing developing, producing, testing, selling or distributing AI systems

There are a number of positive oblgiations which apply to persons and/or legal person designing developing, producing, testing, selling or distributing AI systems, such as:

- With regards to AI systems deployed in healthcare, the Health Insurance Portability and Accountability Act (HIPAA) applies;

- Failure to respect the principles when using data and algorithms to make decisions about consumers could lead to a violation of the following acts:

  - the Fair Credit Reporting Act,
  - the Equal Credit Opportunity Act;
  - Title VII of the Civil Rights Act of 1964;
  - the Americans with Disabilities Act;
  - the Age Discrimination in Employment Act;
  - the Fair Housing Act;
  - the Genetic Information and Nondiscrimination Act

- and be generally pursued under the FTC Act regarding unfair and deceptive trade practices;

4 Applicability of Traditional Criminal Law Categories

4.1 The notion of "computer system" as defined by Article 1, lett. a) of Cybercrime Convention and/or Article 2, lett. a) of Directive EU/2013/40

Article 1 of the Cybercrime Convention defines "computer system" as "any device or a group of interconnected or related devices, one or more of which, *pursuant to a program*, performs *automatic processing of data"*.

Directive EU/2013/40 Art 2 a defines information system as "a device or group of inter-connected or related devices, one or more of which, *pursuant to a programme*, *automatically processes computer data*, as well as computer data stored, processed, retrieved or transmitted by that device or group of devices for the purposes of its or their operation, use, protection and maintenance".

The US has not adopted a definition of AI systems. By comparing the different definitions that have been set forth (see sec. 1) it is possible to theorize that AI systems could fall under the scope of the Cybercrime Convention.

For example, the definition contained in the Defense Authorization Act of 2019 focuses on the autonomy of the system ("without significant human oversight, or that can learn from experience and improve performance when exposed to data sets") and includes AI systems "developed in computer software". The National Artificial Intelligence Initiative Act of 2021 recalls that AI systems can abstract perceptions into models through analysis in an automated manner.

4.2 Other definitions applicable to AI systems

AI systems could fall contained in the definition of "Computer" of the federal Computer Fraud and Abuse Act (18 U.S.C. § 1030(e)(1):

> (1) the term "computer" means an electronic, magnetic, optical, electrochemical, or other high speed data processing device performing logical, arithmetic, or storage functions, and includes any data storage facility or communications facility directly related to or operating in conjunction with such device, but such term does not include an automated typewriter or typesetter, a portable hand held calculator, or other similar device;

4.3 Principle of legality and possible issues

The American criminal law system knows the concept of "principle of legality", often referred to with the Latin phrase "*nullum crimen sine lege, nulla poena sine lege"* (no crime or punishment without law).

The principle of legality is reflected in:

- The prohibition of *ex post facto laws* and *bills of attainder*[65] (U.S. Constitution art. I, §§ 9 and 10);
- The requirement of *strict construction* of criminal law statutes;
- The *void-for-vagueness* doctrine (Fifth and Fourteenth Amendment to the U.S. Constitution);
- The trend away from open-ended common law crimes.[66]

There is no competence at federal level to create new crimes – with the exception of federal territory – unless jurisdictional powers are expressly conferred (such as in the field of interstate commerce, to prosecute war, etc.) or implicitly inferred by the Constitutional "necessary and proper" clause (U.S. Const. art. I, § 8). The so-called "police power" to proscribe punishment lies mostly within states.

Matters regarding the principle of legality are raised by Sarch and Abbott when addressing the so-called "Eligibility Challenge" (i.e., "AI, like inanimate objects, is not the right thing to punish").[67] The authors discuss whether convicting AI systems of crimes which require *mens rea* would violate the principle of legality: since AI systems cannot fulfill the *mens rea* requirement, punishing them with a criminal offense would entail convicting a defendant of a crime when it is not proven that its conduct satisfied all the elements of the offense.[68] One of the solutions proposed by these authors is a legal fiction: they contend that principle of legality concerns can be superseded by applying the doctrine of *respondeat superior*, which implies that mental states possessed by an agent of a corporation can be imputed to the corporation itself, provided that the agent was acting within the scope of his employment and in the furtherance of corporate interests.

4.3.1 On analogy

As already mentioned, the American legal tradition is familiar with the concept of legality in criminal law: analogy represents a challenge to the concept of *nullum crimen sine lege*. Yet, the American legal system does not prohibit analogy at constitutional level (as other civil legal systems instead do). In this sense, American criminal legal doctrine refers more frequently to the principle of "strict construction of criminal statutes", rather than to prohibitions of reasoning by analogy.[69] The rule entails that criminal statutes must be strictly construed in favor of the defendant.[70] It is based on two notions: first, prospective offenders should be given fair warning before they engage in criminal

[65] "A legislative act which inflicts punishment without a judicial trial", *Cummings v. Missouri* 71 U.S. (4 Wall) 277 18 L.Ed. 356 (1867).
[66] W. LaFave, *Criminal Law* (4th edition, Thomson West, 2003), 11.
[67] Sarch and Abbott (n 23) 349.
[68] ibid
[69] Jessica L. Corsi, 'An Argument for Strict Legality in International Criminal Law' (2018) 49 Georgetown Journal of International Law, 1338.
[70] La Fave (n 66) 88.

conduct; second, legislatures have the power to define crimes, not courts. The principle of fair warning is the foundation for three further rules: A) vague criminal statutes violate due process; B) the prohibition against ex post facto laws and; B) the decision of states (and of the federal government) to abolish common law crimes.[71]

The prohibition of *ex post facto judicial decisions* (i.e., the retroactive application of a judicial decision to the disadvantage of a defendant in a criminal case) is not as strong as the prohibition of ex post facto statutes.[72] In Bouie v. City of Columbia, 378 U.S. 347 (1964), the U.S. Supreme Court has held that due process clause prohibits an appellate court from giving retroactive application to its new construction of a criminal statute, since it deprives the defendant of their right to fair warning of a criminal prohibition, in violation of the due process clause of the Fourteenth Amendment.

Analogy has not (yet) been used to criminalize illegal acts related to AI systems.

4.4 Joint-perpetration and participation

Complicity is defined in the MPC at §2.06 (3)-(4):

> (3)A person is an accomplice of another person in the commission of an offense if:
> (a) with the purpose of promoting or facilitating the commission of the offense, he
> (i) solicits such other person to commit it; or
> (ii) aids or agrees or attempts to aid such other person in planning or committing it; or having a legal duty to prevent the commission of the offense, fails to make proper effort so to do; or
> (b) his conduct is expressly declared by law to establish his complicity.
> (4) When causing a particular result is an element of an offense, an accomplice in the conduct causing such result is an accomplice in the commission of that offense, if he acts with the kind of culpability, if any, with respect to that result that is sufficient for the commission of the offense.

Complicity is not an offense itself (differently from conspiracy and solicitation), but a theory by which an accomplice is rendered liable for an offense committed by a perpetrator. The assistance provided by the accomplice does not have to be essential for the offense to be successfully realized nor there is a substantial threshold of gravity that needs to be reached. It implies an agreement to commit a crime. Any crime in furtherance of a conspiracy that is reasonably foreseeable may lead to criminal liability for any member of the conspiracy ("an overt act of one partner may be the act of all without any new agreement specifically directed to that act… o long as the partnership in crime continues, the partners act for each other in carrying it forward … the criminal intent to do the act is established by the formation of the conspiracy" (*Pinkerton v. United States*, 328 U.S. 640 (1946)).

[71] ibid

[72] La Fave (n 66) 116.

Following a "perpetration-through-another" model, a criminal plan is executed through the instrumental use of another person which is an innocent agent. It is a "late development of the concept of vicarious liability into a law of complicity". [73]

According to Hallevy, the *perpetration by another* doctrine is applicable to cases where a programmer designed an AI system to commit certain offences, but the system then exceeded the plan either quantitatively, qualitatively or in both ways.[74] The AI system is then considered a very sophisticated tool to commit the crime, i.e., an innocent agent. The innocent agent doctrine requires that the perpetrator must know or foresee that the AI system that is being used will cause harm. According to Hallevy, the liable perpetrator-through-another agent can be both a human agent (the programmer of the AI system, the user, or the end-user) and an AI agent. The model would fit two types of scenarios including AI systems:

- When the AI system (even if equipped with strong AI technology) is used to commit an offense without using its advanced capabilities;
- When the AI system used is a "weak" one.

4.5 Legal persons

Corporate criminal liability (CCL) is proscribed both at federal and at (some) state level. CCL is regulated in the MPC in sec. 2.07. A corporation may be held criminally liable according to the MPC when

> a) the offense is a violation or the offense is defined by a statute other than the Code in which a legislative purpose to impose liability on corporations plainly appears and the conduct is performed by an agent of the corporation acting in behalf of the corporation within the scope of his office or employment, except that if the law defining the offense designates the agents for whose conduct the corporation is accountable or the circumstances under which it is accountable, such provisions shall apply; or
> (b) the offense consists of an omission to discharge a specific duty of affirmative performance imposed on corporations by law; or
> (c) the commission of the offense was authorized, requested, commanded, performed or recklessly tolerated by the board of directors or by a high managerial agent acting on behalf of the corporation within the scope of his office or employment.

Hence, in theory, legal persons could be held criminally liable for AI-related crimes committed for their benefits.

---

[73] Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Springer 2015), 106.
[74] ibid 118.

According to Westbrook, in order to "determine a corporation's criminal liability for the crimes attributable to an AV, a court would have to determine the scope of the corporation's involvement in the crime, including the role of its agents … Due to the inconsistent body of law governing such liability, however, the court's task will prove difficult".[75]

Diamantis argues that the law ought to consider algorithmic conduct as corporate conduct and, as a consequence, the framework on CCL would kick in.[76] Accordingly, corporations are in the best position to benefit and to discipline/correct algorithms, as they are with employees.

4.6 Secondary liability

The American criminal legal system identifies different roles in the commission of a criminal offense by group: 1) principal in the first degree; 2) principal in the second degree; 3) accessory before the fact; 4) accessory after the fact. The accessory before the fact is the one who "orders, counsels, encourages, or otherwise aids and abets another to commit a felony and who is not present at the commission of the offense".[77] The accessory conduct might take place far away in time from the commission of the offense and the quantity of the aid is immaterial. Most states provisions, today, do not require nor the conviction or the prosecution of the principal actor in order to convict the accomplice. With regards to the *actus reus* element, this is described with different terms by states: aid, abet, advise, assist, cause, command, counsel, encourage, hire, induce, and procure. The MPC at §2.06 (3)(a)(i-iii) refers to conducts of soliciting, aiding or agreeing, failing to make the proper effort to prevent the commission of an offense (in presence of a legal duty to so). With regards to the *mens era* element, it is certain that some level of mental connection with the offense committed by another is required (mere assistance or encouragement is not sufficient, see *Hicks v. U.S.*, 150 U.S. 442, 14 S.Ct. 144, 1893). The level of *mens rea* required, though, varies state by state. The MPC prescribes that the accomplice must act with the purpose of promoting or facilitating the commission of the offense (§2.06 (3)(a)).

Vicarious liability statutory crimes and strict liability statutory crimes are compatible with the American Constitution. LaFave traces the following distinction amongst the two: "With strict liability, there must be a showing that the defendant personally engaged in the necessary acts or omissions; only the requirement of mental fault is dispensed with altogether. By contrast, with vicarious liability it is the need for a personal *actus reus* that is dispensed with, and there remains the need for whatever mental fault the law requires on the part of the employer".[78] The MPC (§ 2.05 (1)(a)) permits strict

[75] Westbrook (n 52) 121.
[76] Mihail Diamantis, 'Algorithms Acting Badly: A Solution from Corporate Law' (2021) 89(4) George Washington Law Review 801.
[77] LaFave (n 66) 666.
[78] LaFave (n 66) 694.

liability offenses only for offenses which constitute violations, i.e., not criminal offense which can be punished only by a fine or another civil penalty.

It follows that forms of secondary liability might be applicable to AI-related crimes: with regards to *accomplice liability*, this would entail considering the AI system as *capable* of fulfilling both requirements of a criminal offense; with regards to *vicarious liability*, it would entail regarding an AI system as capable of fulfilling the *mens rea* requirement; with regards to strict liability it would entail regarding an AI system as capable of fulfilling the *actus reus* requirement (which seems to be the predominant opinion in legal doctrine).

4.7 Illegal acts committed through or against an AI system

The Computer Fraud and Abuse Act (H.R. 4718) criminalizes access and other illicit conducts committed against a "protected computer" (18 USC § 1030(e)(1)).

Protected computers are computers "exclusively for the use of a financial institution or the United States Government, or, in the case of a computer not for such use, used by or for a financial institution or the United States Government and the conduct constituting the offense affects that use by or for the financial institution of the Government". The CFAA also protects any computer, whether or not connected to the government, "which is used in or affecting interstate or foreign commerce or communication, including a computer located outside the United States that is used in a manner that affects interstate or foreign commerce or communication of the United States." The term "protected computer" has been interpreted broadly by US courts, especially in cases of "public safety" concerns.

Ryan Calo et al.[79] discuss whether the CFAA – which is considered the paradigmatic U.S. federal anti-hacking law – could be applied to adversarial ML. They describe five case studies (1-planting adversarial sound commands in ads; 2-causing a car crash by defacing a stop sign to appear like a speed limit; 3-shoplifting with anti-surveillance makeup; 4-poisoning a crowd-sourced credit rating system; 5-data inversion across international borders) and argue that strong arguments both in favor and against applying CFAA could be made. The authors conclude that "a comparison between the leading anti-hacking law and adversarial machine learning reveals is ambiguity. It simply isn't clear how or when the CFAA or similar laws applies to "tricking" a robot as opposed to "hacking" it".[80] The ambiguity could lead to: uncertainty and prosecutorial overreach; a chilling effect on research on testing systems for resilience; (in case adversarial ML is not considered hacking) less incentive for firms to ensure that their systems are attack-resilient.

[79] Ryan Calo and others, 'Is Tricking a Robot Hacking?' (2018) University of Washington School of Law Research Paper.
[80] ibid 15.

In general, the wording of existing cybercrime and computer offenses (which include also provisions of the Wiretap Act and other relevant federal laws) could be applied to illegal acts committed through or against an AI system.

4.8 *Mens rea* aspects

The MPC defines four levels of culpability: purposely, knowingly, recklessly, and negligently.

The MPC regulates intent at § 2.02. An agent is acting *purposely* if, according to sec. 2.02(2)(a) MPC, "it is his conscious object to engage in conduct of that nature or to cause such a result" and if he is aware of the "attendant circumstances" that satisfy the other elements of the crime. An agent is acting *knowingly*, according to sec. 2.02(2)(b) MPC, when "he is aware that it is practically certain that his conduct will cause such a result" and if he is aware of the "attendant circumstances" that satisfy the other elements of the crime. An agent is acting *recklessly*, according to sec. 2.0.2(c) MPC, if he "consciously disregards a substantial and unjustifiable risk . . . [that] involves a gross deviation from the standard of conduct that a law-abiding person would observe in the actor's situation". An agent is acting *negligently*, according to sec. 2.0.2(d) MPC, ""when he should be aware of a substantial and unjustifiable risk that the material element exists or will result from his conduct". The actor's failure to perceive the risk must involve a gross deviation from the standard of conduct that reasonable person would observe in the actor's situation. The distinction between knowledge and recklessness lies in the "degree of risk—"practically certain" versus "substantial risk"— of which the defendant is aware".[81] The distinction between recklessness and negligence is that the latter requires "something less" (i.e., the awareness of the risk). The common denominators of the different existing definitions of negligence are that there must be (a) a degree of risk that must be created by the conduct of the defendant and (b) a comparable objective reasonable man standard.[82]

The question then is: 1) whether the "conscious object" behind someone's behavior to cause a result should include the exact and concrete modus operandi of the AI system (*intent*); 2) whether the awareness that a certain conduct will cause a result should include the exact and concrete modus operandi of the AI system (*knowledge*).

According to Abbott and Sarch,[83] recklessness or negligence liability could apply to cases where a human agent does not intend or foresee that the AI system will cause harm, provided that the AI systems creates a "foreseeable risk of a prohibited harm". For example, "if the developers or users of Al foresee a substantial and unjustified risk that an Al will cause the death of a person, these human actors could be convicted of reckless

[81] Kevin J. Heller and Markus D. Dubber, M. (eds.), *The handbook of comparative criminal law* (SUP 2010), 574.
[82] LaFave (n 66) 263.
[83] Sarch and Abbott (n 23) 370.

homicide *[MPC § 210.3(a)]*. If such a risk was merely reasonably foreseeable (but not foreseen), then lower forms of homicide liability would be available *[such as negligent homicide as proscribed at § 210.4]*. Similar forms of recklessness or negligence liability could be adopted where the Al's designers or users actually foresaw, or should have foreseen, a substantial and unjustified risk of other kinds of harms as well - such as theft or property damage".[84]

5 Case law

5.1 Judgments concerning AI systems, relevant for possible criminal consequences – Torts law

5.1.1 Hudson v. Tesla Inc, Circuit Court of the Ninth Judicial Circuit, In and For Orange County, Florida, 2017

Case regards the complaint filed by Hudson against Tesla for injuries suffered by the failure of the autopilot feature of a Tesla S to detect presence of a vehicle on the roadway, causing the car to crash into the vehicle at approximately 80 mph.[85] Plaintiff claimed that Tesla was strictly liable for designing manufacturing, producing, distributing and selling the car, as the model was defective in its design, manufacture and warning – making it unreasonably dangerous for its intended and reasonably foreseeable use. Plaintiff also claimed Tesla was negligent due to a breach in its duty of care to design, manufacture, produce, distribute, and sell the Model S and the Model S's autopilot system in a condition that was not defective and unreasonably dangerous. Tesla also owed a duty of care to adequately test, inspect, and ensure the quality of the Model S and the Model S's autopilot system prior to placing these products into the stream of commerce.

In 2018 Tesla settled a class action lawsuit regarding its Model S and Model X cars with a 5 million dollars settlement.[86]

5.1.2 Huang et al, v. Tesla, Inc.

The family of Walter Huang, who died after an accident caused by the slamming of its Model X Tesla on autopilot feature against a concrete wall, filed a complaint against Tesla in 2018. The complaints were many: negligence, wrongful death, strict liability, negligence (post-sale), defective product design, failure to warn, intentional and negligent misrepresentation, and false advertising. Specifically, plaintiffs complained that Tesla was negligent in failing and omitting to provide adequate instructions and warnings to protect against injuries occurring as a result of vehicle malfunction and the absence of an effective automatic emergency braking system. They also claimed Tesla

[84] ibid

[85] *Hudson v. Tesla, Inc.* Complaint Filing # 80052957 (2018).

[86] Tina Bellon, 'Tesla agrees to settle class action over Autopilot billed as "safer"' (*Reuters* 25 May 2018) < https://www.reuters.com/article/us-tesla-autopilot-lawsuit-idUSKCN1IQ1SH > accessed April 2022.

was strictly liable as the Tesla X constitute a defective product according to California law.

5.1.3 Nilsson v. General Motors LLC, U.S. Dist. Ct. N.D. California, 2018.

Case regards a negligence suit filed by Oscar Nilsson, a motorcyclist who was involved in a collision with a car in self-driving mode in December 2016 (Compl. ¶¶ 15-16, Nilsson v. Gen. Motors, Jan. 22, 2018 (No. 18-471) (N.D. Cal.), ECF No. 1). The car suddenly changed lanes and collided with the motorcycle driven by the plaintiff. The plaintiff complained that General Motors had a duty of care in having its Self-Driving Vehicle operate in a manner in which it obeys the traffic laws and regulations The case was settled.

6 Adaptation of Traditional Criminal Law Categories and academic debate

6.1 Legal issues concerning the traditional categories of the general part of the criminal law

Matters of *actus reus* and *causality* seem to have gained less attention in American criminal legal doctrine as compared to matters of *mens rea*. As mentioned by Kingston, "It is relatively simple to attribute an actus reus to an AI system. If a system takes an action that results in a criminal act, or fails to take an action when there is a duty to act, then the actus reus of an offence has occurred".[87]

6.1.1 Actus reus

The MPC defines an *act* as a bodily movement (§ 2.01), which could be both voluntary or not. The requirement of an act is not safeguarded by any constitutional provision.[88]

Hallevy abandons the doctrine which submits that the criminal act need be the result of a willed bodily movement. He considers it a "mongrel requirement" belonging to the past.[89] In other words, according to this author there is no space for any mental element requirement when examining *actus reus*. Following this reasoning, Hallevy claims that "for the question of performing an act in order to satisfy the conduct component requirement, any material performance through factual–external presentation is considered an act, whether the physical performer is strong artificial intelligence entity or not". [90] The same reasoning is also applied to commission-by-omission scenarios.

[87] John K.C. Kingston, 'Artificial Intelligence and Legal Liability' in Max Bramers and Milos Petridis (eds), *Research and Development in Intelligent Systems XXXIII: Incorporating Applications and Innovations in Intelligent Systems XXIV* (Springer 2016), 4.
[88] Lima, 679.
[89] Hallevy (n 73) 61.
[90] Hallevy (n 73) 62.

From this statement, it follows that even the simplest machines could perform "conduct under the definition and requirements of criminal".[91]

Dafni Lima is amongst the few scholars to analyze whether and AI system can act from a criminal legal perspective.[92] She argues that AI actions could hardly fall into the definition of acting: "Even if we set aside as obsolete the "bodily" dimension of acting, which by definition could never apply to a machine, an intelligent agent's movements could neither be seen as "socially relevant" nor as "voluntary" in the sense that criminal law implies".[93] Moreover, the AI system cannot be said to act voluntarily:

> Voluntariness … implies the ability to act otherwise, and an agent that is programmed to choose A when it encounters B is not necessarily choosing. Thus, Al agents do not yet seem to possess the potential for fully independent, even self-destructive decisions. In other words, no one would regard a robot's choice to change its route when stumbling upon a table as voluntary, so long as the robot is simply following an algorithm, however intricate, that dictates it to change route when encountering a physical obstacle-or to put it more simply, so long as the robot does not have the choice to keep hitting at the obstacle if it so wishes. This holds true even when this choice is the only reasonable one and in the Al agent's "benefit" of achieving its objective.[94]

6.1.2 Causality

According to the MPI, causation is ascertained following two requirements:

- the but-for test: the conduct must be "an antecedent but for which the result in question would not have occurred. (§ 2.03 (1)(a)),"
- The legal or proximate cause test: the harm must be "not too remote or accidental in its occurrence to have a [just] bearing on the actor's liability or on the gravity of his offense (§§ 2.03(2)(b) and (3)(b)).[95]

Hallevy adheres to a *conditio sine qua non* theory of causation.[96] Consequently, he claims that since AI technology is "capable of committing conduct of all kinds, in the context of criminal law, it is capable of causing results out of this conduct".[97]

---

[91] ibid 63.
[92] Dafni Lima, 'Could AI Agents Be Held Criminally Liable: Artificial Intelligence and the Challenges for Criminal Law'(2018) 69 S. C. L. REV. 677.
[93] ibid 682.
[94] Lima (n 92) 683.
[95] Heller and Dubber (n 81) 572
[96] Also referred to as "ultimate cause theory" or "but for" test. According to this theory, to prove that the conduct A cause the result B, the judge has to remove element A from the factors which led to the event and ask herself if event B would still have happened.
[97] Hallevy (n 73) 66.

6.1.3 Principle of culpability (nullum crimen sine culpa) and mens rea

When discussing *mens rea* of AI systems, Hallevy asks himself whether artificial intelligence technology has on one hand, the capability of "being aware of conduct, circumstances or possibility of the results' occurrence, in the context of criminal law" [98] and, on the other hand, the capability of consolidating will.[99] The author answers positively to both questions.

With regards to liability of *human agents*, Hallevy theorizes two models: 1) the perpetration-through-another model; 2) the natural probable consequence model.

The *perpetration-through-another* model would be applicable to situations where the AI system is used by a human being as a (sophisticated) tool to commit an offense. The AI system is treated as an innocent agent and therefore the model considers the action committed materially by the AI system as if it had been the action of the user or of the programmer.

The *natural probable consequence model* would be used in cases where the AI system was not designed to commit the specific offense, but the offense was committed by the artificial intelligence technology, nonetheless.[100]

According to the "natural and probable consequence doctrine", members of a criminal enterprise are held liable for *unplanned* harms, i.e., outcomes that differ from intended harms (acts that are the natural and probable consequence of the criminal scheme the accomplice encouraged or aided).[101] Such a model requires two conditions: first, that the unplanned offense is a consequence of the planned offense; second, that the unplanned offense is probable (hence foreseeable by the relevant party) as a consequence of the planned offense. As mentioned by Hallevy,[102] "American criminal law imposes full criminal liability for the unplanned offense equally upon all parties of the planned offense as long as the unplanned offense is the *probable consequence* of the planned one".[103]

As stated in U.S. v. Powell, 929 F.2d 724 (D.C.Cir.1991), the expression "natural and probable consequences" does not communicate the level of likelihood that the forbidden act must have in the eyes of the accomplice: "It could signify any position within a broad range: for example, all acts with a substantial probability of occurrence (e.g., one chance in five); acts that are more probable than not to occur; acts of very high probability (e.g.,

---

[98] ibid 89.
[99] Hallevy (n 73) 94.
[100] ibid 119 ss.
[101] Wayne R. LaFave, *Criminal Law* (2nd ed, Thomson West 1988), § 6.8, 590.
[102] Hallevy (n 73) 117.
[103] *People v. Prettyman*, 14 Cal.4th 248, 58 Cal.Rptr.2d 827, 926 2d 1013 (1996); *Chance v. State*, 685 A.2d 351 (Del.1996); *Ingram v. United States*, 592 A.2d 992 (D.C.Ap 1991); *Richardson v. State*, 697 N.E.2d 462 (Ind.1998); *Mitchell v. State*, 114 Nev. 1417, 971 2d 813 (1998); *State v. Carrasco*, 122 N.M. 554, 928 2d 939 (1996); State v. Jackson, 137 Wash.2d 712, 976 2d 1229 (1999); *United States v. Powell*, 929 F.2d 724 (D.C.Cir.1991).

90%); and acts so likely that their occurrence is a practical certainty". Hence, the "natural or probable consequence" model is normally used to prosecute accomplices to a crime. In case no conspiracy is demonstrated, it is possible, according to US law, to prosecute the accomplice in case the criminal acts of the perpetrator were the natural or probably consequence of a scheme "that the accomplice encouraged or aided, as long as the accomplice was aware that some criminal scheme was under way".[104]

With regards to AI systems, the model would be applicable in situations where there is a "deep involvement of the programmers or users in the AI entity's daily activities, but without any intention of committing any offense via the AI entity".[105] Notably, this model shall be used in *aberratio delicti* situations, i.e., "unplanned developments of a planned delinquent event".[106]

Hallevy provides the following example:

> [A] medical expert artificial intelligence system is used for diagnosis of certain types analysis is based on machine learning, which inductively analyses and generalizes specific cases. The system fails to diagnose correctly one case, and that reveals to wrong treatment, which worsens the patient's situation and finally causes the patient's death. The analysis of the artificial intelligence system's activity reveals negligence of it,[107] and it fulfills both factual and mental elements requirements of the relevant negligence offense (negligent homicide). At this point arises the question of the *programmer*'s [emphasis added] criminal liability for that offense. His criminal liability is not related to the decision to use the artificial intelligence system, to follow its diagnosis, etc., but it is related to the very initial programming of the system. If the programmer would have programmed the system to kill patients and instrumentally used it for this purpose, it would have been perpetration-through- another of murder, but this is not the case here. For the programmer's criminal liability in this case the probable consequence liability may be relevant.[108]

---

[104] Kingston (n 87) 3. Kingston bases his reflections on American law.

[105] Gabriel Hallevy, 'The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control' (2010) 4 (2) Akron Intellectual Property Journal, 181.

[106] Hallevy (n 73) 119.

[107] Hallevy theorizes that it is possible to ascertain an AI system's negligence by analyzing the machine learning process which led to the mistaken decision. He defines negligence as "unawareness of the factual component in spite of the capability to form awareness, when reasonable person could and should have been aware of that component", Hallevy (n 73) 124. He summarizes the test in three general questions, which would have to been ascertained by a court (with the aid of a computer scientist expert):
(a) Was the artificial intelligence system unaware of the factual component?
(b) Has the artificial intelligence system the general capability of consolidating awareness of the factual component?
(c) Could a reasonable person have been aware of the factual component?

[108] ibid 134.

By applying this model, users or programmers would be held criminally liable if they knew that "a criminal offence was a natural, probable consequence of their programs/use of an application".[109] The natural probable consequence model does not fit situations of *pure negligence*, such as the ones where the programmer had no criminal intent whatsoever, but he was negligent in the programming of the AI system, which lead to the realization of a criminal offense.

The models theorized by Hallevy can be combined.

According to Abbott and Sarch, criminal law has the tools to handle cases where the human agents was not negligent, and AI-related harms were not foreseeable. They make the following example:

> Imagine hackers use an Al to drain a fund of currency, but this ends up unforeseeably shutting down an electrical grid which results in widespread harm. The hackers are already guilty of something namely, the theft of currency (if they succeed) or the attempt to do so (if they failed). Therefore, our question here is whether the hackers can be convicted of any further crime in virtue of their causing harm through their Al unforeseeably taking down an electrical grid. [110]

They argue that one could blame the hackers by applying *constructive liability* crimes, i.e., offenses that "consist of a base crime base crime which require mens rea, but where there then is a further result element as to which no mens rea is required".[111] The paramount example of a constructive liability crime is felony murder, where the liability for murder is "attached" to the commission or attempted commission of a felony: think of the burglar that breaks into a home, thinking that the house is empty. He then finds out that the homeowner is home. As a consequence, the homeowner is frightened to death by the burglar (he has a heart attack and dies). According to this doctrine, the burglar is guilty of murder. This type of construction, according to some, is only justified if the base crime in question carries the "risk of the same general type of harm as the constructive liability element at issue (death)".[112] According to Abbott and Sarch, constructive criminal liability could be extended to the case of the hackers and the electrical grid.

6.2 Proposal for a possible form of criminal liability or a direct sanctioning of AI systems or AA

Gabriel Hallevy theorizes a *direct liability model* which would be applicable in situations where the AA makes a decision to commit an offense based on its own accumulated experience or knowledge or based on advanced calculations of probabilities.

---

[109] Kingston (n 87) 4.
[110] Sarch and Abbott (n 23) 371.
[111] ibid 372.
[112] ibid

With regards to *punishment*, Hallevy argues that retribution and deterrence would prove useless in the case of punishing robots but could prove valuable when punishing the human participants in the offence.[113] On the other hand, he states that rehabilitation and incapacitation are relevant from an AI-punishment perspective. Rehabilitation could function for machines as it functions for humans: it may be used to refine the machine learning process, as a way to lead AI systems to make better decisions. The rehabilitated AI system, then, would be able to perform better, same as rehabilitated defendants should have better tools to face reality.[114] Incapacitation would be the last resort measure directed at those systems that have proven to be incapable of changing their ways through their inner processes (i.e., via machine learning).[115]

Ying Hu makes a positive case for imposing direct criminal liability on "smart robots", that is, robots that fulfil three threshold conditions. These conditions are: "the robot must be (1) equipped with algorithms that can make nontrivial morally relevant decisions; (2) capable of communicating its moral decisions to humans; and (3) permitted to act on its environment without immediate human supervision".[116] Hu argues in favor of legal personhood for robots. Yet, she distances herself strongly Hallevy: the Israeli author, she claims, fails at describing in detail the type of robot on which to impose criminal liability and at explaining way why we should adopt the criminal tool in the first place.[117] Moreover, he "appears to assume that, since we already impose criminal liability on non-human entities such as corporations, extending such liability to robots requires little justification".[118]

Hu theorizes the introduction of a Criminal Code for Robots. The reasons for its introduction are twofold: first, there are grounds to hold smart robots to a higher moral standard than humans. To support her claim, she argues that smart robots could be held liable for failure to act not only when they have a legal duty to do so.[119] Yet, this would entail diverting heavily from the principle of culpability. Second, smart robots might prompt new moral questions which were never faced by humans, as they are able to act in ways that are physically impossible for a human being. Hu claims that it would introduce a minimum set of moral standards decided collectively by society. What is more, it would work as a legal basis for holding "robot manufacturers" (subjects who participate in creating the algorithms) and "robot trainers" (subjects who train the

[113] Hallevy (n 73) 210.
[114] ibid 211.
[115] Hallevy (n 73) 211.
[116] Hu (n 20) 490.
[117] ibid 492, note 13.
[118] Hu (n 20) 492.
[119] Think for example of a smart robot which, through its sensors, detects that a person is drowning. He could be held accountable for not trying to save the person. The same couldn't be said with regards to a human being unless she held a specific duty of care – such as in the case of working as a lifeguard on the specific beach. Cfr. Hu (n 19).

algorithms) criminally liable for failing the duty of care, i.e., for preventing smart robots from behaving in a way that is against the code.

She subsequently presents her case for imposing criminal liability on robots based on three arguments:

- criminal punishment has a censuring (or expressive) function. It communicates the disapproval of the community towards a morally wrongful conduct and this function acquires greater importance when no human being is at fault for the robot's misconduct.
- punishing robots would provide emotional relief to victims of smart robots' misbehavior.
- it would be a useful to identify culpable (human) individuals, since those who are not at fault would be inclined to cooperate with investigations, therefore pinpointing to those who are. It would also nudge towards the creation of self-policing mechanisms for robot manufactures and users, who would put in place safeguards against robot harm to avoid sanctions on the robot. [120]

Hu hypothesizes four forms of punishments for robot criminals:

- physically destroying the robot (the robot equivalent of a "death sentence");
- destroying or re-writing the moral algorithms of the robot (the robot equivalent of a "hospital order");
- preventing the robot from being put to use (the robot equivalent of a "prison sentence"); and/or
- ordering fines to be paid out of the insurance fund (the robot equivalent of a "fine"). [121]

Christina Mulligan argues that imposing punishment ("revenge" or "vengeance", as defined by the author) on AI systems would result in retributive benefits consisting in the psychological satisfaction of victims of AI-misbehavior. [122] She claims that the goal of criminal law is to create psychological satisfaction for the victims of the robot, and this can be done most effectively by introducing a modern version of the Middle Age practice of "noxal surrender". Noxal surrender would involve handing over the faulty robot to the victim or to her family so that they to do what they think its best with it in order to fit their satisfaction. As a matter of fact, a wronged party "may indeed be quite justified in dragging a robot out into an empty field and walloping it with a baseball bat". [123]

---

[120] Hu (n 20) 490.
[121] ibid 529.
[122] Christina Mulligan, 'Revenge Against Robots'(2018) South Carolina Law Review 69, 578.
[123] ibid 595.

Abbott and Sarch created the term "Hard AI Crime".[124] According to the authors, AI system behavior expresses four main features, which are relevant features from a criminal law perspective:

- Unpredictability, i.e., the capacity for the system to engage in activities which were not intended or foreseen by its creators;
- Unexplainability, i.e., the incapacity to explain why the AI system chose a certain pattern of behavior;
- Autonomy, i.e., the capacity for the AI systems to act independently of human control and therefore to cause harm without being under the direct control of an individual;
- Complexity, i.e., the fact that the AI system is the output of the contribution of many individuals over a long period or that its conduct is the result of a training based on huge databases coming from heterogeneous sources.

The combination of these factors might lead to irreducibility i.e., the impossibility to reconnect the crime to a liable person. The authors distinguish between "AI Crimes" and "Hard AI Crimes". AI Crimes are defined as "cases in which an AI would be criminally liable if a natural person had performed the same act".[125] Hard AI Crimes are instead defined as "scenarios where crimes are functionally committed by machines and there is no identifiable person who has acted with criminal culpability".[126] Hard-AI Crime make the strongest case for punishing artificial intelligence.

These authors provide a comprehensive theory for the foundation of criminal punishment of AI systems. They anchor their reflections building on a theory of punishment, which is, in turn, based on affirmative (pluralist) benefits. They contend, indeed, that punishing AI directly might lead to significant affirmative benefits. First, they argue that it could obtain general deterrence. Punishment of AI systems could lead to unrestricted general deterrence of other subjects, namely of developers, owners or users of AI. Moreover, AI punishment could have some expressive benefits, in the sense that it would convey a "message of official condemnation that could reaffirm the interests, rights, and ultimately the value of the victims of the harmful AI".[127]

With regards to AI systems and *mens rea*, the authors contend that that the preponderant theory regarding culpability is based on whether the individual manifests "insufficient regard for legally protected interest or values"[128]. Accordingly, corporations can be

[124] The term AI-Crime (AIC) appears also in Thomas C. King and others, 'Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions' (2019) Science and Engineering Ethics 26 89

[125] Sarch and Abbott (n 23) 332

[126] ibid 328

[127] Sarch and Abbott (n 23) 346.

[128] ibid 355.

accounted as being directly criminally culpable through the "information-gathering, reasoning and decision-making procedures" of their employees.[129] In a similar manner, AI systems can gather information, process it and determine with autonomy how to complete pre-established goals. If these systems were programmed to follow certain rules, for example not to hurt humans, and they diverge from this rule to reach their goals, it could be argued that they display *intent*. Abbott and Sarch introduce the "Belief Desire Intention" (BDI) model for intent, which was first ideated in the 1980s in the field of cognitive science by Micheal Bratman.[130] This theory entails that an agent intends an outcome when he guides her conduct in the direction of causing that outcome. The concept of "direction of action" implies that the agent will adjust his behavior to make the outcome more likely and that it will monitor the surroundings to find ways, which will increase the probability of the outcome. Judges would have to ask themselves whether the system was adjusting and "*guiding* its behavior as to make *[a certain prohibited]* outcome more likely".[131] To conclude, the authors believe that while it is possible to make a coherent theoretical case for punishing AI, it is not justified considering the existence of less "disruptive" alternatives, which can provide the same benefits.

7 Alternatives to criminalization and non-criminal sources

7.1 Compulsory civil insurance for damages resulting from the use of an AI system and other means

Currently there is no form of compulsory civil insurance specifically for damages resulting from the use of an AI system. Other sectorial norms apply.

Shankar and Nagle make a case for AI/ML specific insurance, which would be different from cyber insurance. Cyber insurance typically covers model stealing attacks and data leakages, where instead AI/ML insurance would cover bodily harm due to AI failure; brand damage; damages to physical properties.[132]

Furthermore, there are no rules or regulation which provide for mandatory technical means for combating harm or abuses of AI systems such as reprogramming or destruction of the systems. The only applicable sanctions are those already in place for natural or legal persons and which are applicable to goods and services.

7.2 The role of users

Generally speaking, citizens (and therefore users) do not have a duty to rescue a person that is in peril. Certain states provide for exemptions to this general rule: for example, in

[129] Ibid.
[130] Michael Bratman, *Intention, plans, and Practical Reason* (Harvard University Press 1987).
[131] Sarch and Abbott (n 23) 358.
[132] Ram Shankar and Siva Kumar, 'The Case for AI Insurance' *Harvard Business Law Review* (29 April 2020) < https://hbr.org/2020/04/the-case-for-ai-insurance > accessed March 2022.

Minnesota, according to the Good Samaritan Law (604A.001), "[a] person at the scene of an emergency who knows that another person is exposed to or has suffered grave physical harm shall, to the extent that the person can do so without danger or peril to self or others, give reasonable assistance to the exposed person. Reasonable assistance may include obtaining or attempting to obtain aid from law enforcement or medical personnel".

Self-protection of users could amount to an intervening cause which breaks the chain of legal cause. In this sense, it could exempt producers of liability, provided that the behavior of users was unforeseeable. Moreover, self-protection of the user-victim against harms committed by the AI system could qualify as self-defense, provided that the user-victim reasonably believes that he is in an immediate danger of unlawful bodily harm from the AI system and that the use of the force is necessary to avoid the danger. Regardless of whether the AI system could qualify as "aggressor", the conduct of self-protection could exempt the user from criminal liability against accidental injury to third persons, provided that he was not reckless.

7.3 Product liability

There is no uniform product liability statute or law in the US. Each state defines product liability according to its own standards. There are no criminal offenses related to defective products. Rather, product liability in the US comprises of a mix of tort and contract law obligations.

According to the Restatement (Third) of Torts § 402a (Am. Law Inst. 1965):

> (1) One who sells any product in a defective condition unreasonably dangerous to the user or consumer or to his property is subject to liability for physical harm thereby caused to the ultimate user or consumer, or to his property, if
> (a) the seller is engaged in the business of selling such a product, and
> (b) it is expected to and does reach the user or consumer without substantial change in the condition in which it is sold.
> (2) The rule stated in Subsection (1) applies although
> (a) the seller has exercised all possible care in the preparation and sale of his product, and
> (b) the user or consumer has not bought the product from or entered into any contractual relation with the seller.

Product liability can be connected to specific theories of liability that identify a source of a product flaw (such as design defect or manufacturing defects) or through strict liability constructs. According to some authors, product liability would be efficient in addressing

AI-harm.[133] Other authors argue that more sophisticated autonomous system should be subject to an *ad hoc* regime.[134]

7.4 Cybersecurity and data protection

The US lacks a comprehensive federal law on privacy.

On June 17, 2021, U.S. Senator Gillibrand introduced the Data Protection Act of 2021 (S.2134) which would create an independent federal agency (the Data Protection Agency). The agency would focus on the protection of individuals' privacy elated to the collection, use and processing data. The bill defines Automated Decision System as: "a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision, or facilitates human decision making" (§ 2 (3)).

Most of the regulation does not address AI directly but it mentions automated decision-making or provides for rights of data subjects which could be impacted by AI systems.

AI systems applied in the field of healthcare have to comply with HIPAA Regulation.

At state level, relevant legislation includes: the California Consumer Privacy Act; the California Privacy Rights and Enforcement Act (CPRA); the Virginia Consumer Data Protection Act (VCDPA); the Colorado Privacy Act (CPA) (also taking effect in 2023).

The states of Illinois (Illinois Biometric Information Privacy Act, 740 ILCS 14/1m), Texas (Texas Business and Commerce Code Sec. 503.001 'Capture or Use of Biometric Identifier) and Washington (Title 19 of the Revised Code of Washington, Chapter 19.375, 'Biometric Identifiers') have enacted laws which provide for data protections for biometric information.

7.5 The role of the human agent

The human agent can take different roles in AI system processes and there is no general regulation in this regard. The role of the human agent is often discussed with regards to human-in-the-loop or human-on-the-loop solutions. Moreover, research has focused on human-agent teams (HAT) to obtain Meaningful Human Control (MHC).[135] This would entail that the human expert oversees (through MHC) the AI system's behavior when performing a morally charged decision. By doing so, the human agent is held accountable for the decision.

---

[133] John Villasenor, 'Products liability as a way to address AI harms'(*Brookings*, 31 October 2019) <https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/> accessed December 2021.

[134] Karni A. Chagal-Feferkorn, 'Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers'(2019) Stanford Law & Policy Review, 30.

[135] Jasper van der Waa and others, 'Moral Decision Making in Human-Agent Teams: Human Control and the Role of Explanations'(2021) Front Robot AI 8.

7.6 Technical rules and standards

The International Organization for Standardization (ISO) and the International Electro technical Commission (IEC) have formed a joint technical committee in 2017 with the purpose of achieving standardization in the field of AI (ISO/IEC JTC 1/SC 42). The Institute of Electrical and Electronics Engineers (IEEE) is building a set of standards on "Ethically Aligned Design" of Smart Information Systems/Autonomous and Intelligent Systems. Other European Standards Organizations (ESOs) and international Standards Development Organizations (SDOs) have proposed and adopted different standards.[136]

With regards to *American federal agencies*, as mentioned above, the NIST has released a plan for prioritizing federal agency engagement in the development of standards for artificial intelligence (AI) per the February 2019 Executive Order on Maintaining American Leadership on Artificial Intelligence (EO 13859).

8 Final evaluations and future developments

The US has joined the global efforts in the regulation of AI, opting for a sector-specific (or agency-based) approach to regulation, rather than for the production of broad regulation (as it is the case instead in the European Union). The regulatory landscape in the US is further muddied, on the one hand, by the delicate balance of powers between the federal and the state government and, on the other hand, by AI's intrinsic trasversal nature. Similarly to other countries, the focus of AI regulation in the US has not been on criminal legal regulation

It is undoubted that AI systems trigger the application of criminal law. For the time being, this interaction is limited to the liability of humans and to malicious uses of AI. Thus, the US legal system can rely on a longstanding tradition of corporate criminal liability and this might facilitate the application of criminal sanctions to corporations for criminal acts committed by AI systems.

[136] An analysis is available at: Stefano Nativi and Sarah De Nigris, 'AI Standardisation Landscape: state of play and link to the EC proposal for an AI regulatory framework' (2021) Publications Office of the European Union EUR 30772 JRC125952.