



University of Tehran Press

Comparative Law Review

Homepage: <https://jcl.ut.ac.ir>

Online ISSN: 2423-3404

Volume: 17, Issue: 1
Spring & Summer
2026

Autonomy: The Indefensible Assumption of Artificial Intelligence Criminal Liability Theory

Abbas Shiri^{1✉} | Mohammadreza Barzegar² 

1. Corresponding author; Associate Professor, Department of Criminal Law and Criminology, Faculty of Law and Political Science, University of Tehran, Tehran, Iran. Email: ashiri@ut.ac.ir
2. PhD Student in Criminal Law and Criminology, Department of Criminal Law and Criminology, Faculty of Law and Political Science, University of Tehran, Tehran, Iran. Email: mrbarzegar@ut.ac.ir

Article Info

Article Type:
Research Article

Received:
2025/03/11
Received in revised form:
2025/05/29
Accepted:
2025/10/02
Published online:
2026/08/23

Keywords:
*Artificial Intelligence,
Autonomy, Agency,
Criminal Liability.*

Abstract

Purpose and Research Problem

The fast development of artificial intelligence (AI) capacities has presented significant conceptual and normative difficulties to accepted legal systems on criminal liability. Certain Western legal theories supporting the expanded concept of "independent criminal liability for AI" call for a basic review of the presumptions of agency and will within the criminal justice system. The main causes of conflict are around the fundamental idea of "autonomy" as the basis for assigning criminal guilt to an intelligent system. This paper investigates whether artificial intelligence has enough autonomy in its present operational and developmental settings to be categorized as an autonomous criminal agent under substantive and procedural criminal law. The relevance of this study is enhanced by its direct link to pragmatic consequences—more especially, the feasibility or inaccessibility of defining the causal link between human actions (as creators or users) and the criminal outcomes resulting from artificial intelligence performance, so precisely determining the human agent's position in the chain of liability. Although international legal literature has made significant theoretical contributions, domestic research has sometimes neglected a thorough and critical analysis of this fundamental condition, so underlining the need of a thorough basic study. This paper recognizes the evolutionary character of criminal liability but contends that any liability system—traditional or creative—made for intelligent systems must establish a baseline of "legal agency," with autonomy being absolutely necessary.

Research Method

Accomplishment This essay aims to do a comprehensive and informed examination of AI autonomy using the descriptive-analytical technique and its data. The analytical framework presented is founded on a comprehensive examination of substantial library resources pertaining to the ethics and philosophy of law, criminal law theory, computer science, and artificial intelligence, encompassing both ancient and modern writings. The fundamental components of "autonomy" in

Article Info**Abstract**

philosophical and legal discourse are initially outlined. We then analyze its adherence to or deviation from contemporary "intelligent systems" technical attributes, concentrating on elements such as "emergent behavior" and "explainability." The two-tier AI, progressing along a spectrum from simple tool utilization to autonomous existence, poses challenges for criminal law doctrine, being classified under terms such as "heteronomous autonomy" and "bounded autonomy." The independent segment of the research involves a typological analysis of actual harmful or criminal cases related to AI, considering instances of algorithmic bias, safety failures, and deliberate malpractices, to empirically assess the operational autonomies of these systems and their capacity to perpetrate harmful actions. The results of these research establish a preliminary basis for the conclusion that the autonomy constraint is insufficient for a theory of autonomous AI criminal culpability and for the formulation of relevant laws.

Results

1. Criminal liability primarily hinges on the agent's autonomy in executing a criminal act—an autonomy characterized by self-determination in decision-making and action, free from external influence. Evaluating this criteria about artificial intelligence necessitates a technological study that integrates "emergent behavior" and "explainability," which provides significant insights. Despite exhibiting operational complexity and exceeding initial programming, emergent behaviors in advanced artificial intelligence systems cannot achieve the requisite autonomy and legal agency for criminal liability, as they are constrained by their architecture, algorithms, training data, and human-defined objectives. Simultaneously, the "explainability" challenge—particularly the "black box" issue present in numerous deep learning models—prevents these systems from providing a "legal or narrative explanation," which necessitates an elucidation of the motivations, objectives, and beliefs driving a specific action in a manner evaluable by the legal system. The limitations of a "precise scientific-mechanical explanation" of the system's internal workings hinder the assessment of mens rea and other mental elements crucial for establishing criminal guilt or applying legal principles. Consequently, both critical criteria under examination indicate that current artificial intelligence has not attained the level of autonomy necessary for autonomous criminal liability.

2. The present level of artificial intelligence cannot be fully ascribed to autonomous agency nor can it be reduced to simple functionality. Although humans set and limit these overall objectives and constraints, ideas like "heteronomous autonomy" and "bounded autonomy" more fairly depict this intermediate condition, when the intelligent system has operational freedom to reach its goals. The difference between artificial intelligence and conventional passive instruments exposes a "liability gap" in classical criminal liability theories since the intelligent system cannot be considered as an independent criminal agent and assigning liability to human agents is difficult due of the impracticality or major constraints of direct control

Article Info	Abstract
	and total predictability of human behavior. Algorithmic opacity and the "black box" phenomena compromise the evaluation of mens rea and causality for people engaged in the system, so aggravating this problem. Important legislative and regulatory systems mostly ignore the different criminal liability of artificial intelligence and concentrate on improving and clarifying mechanisms of human liability; they also show shortcomings in both theory and practice. 3. Consistently, a typological analysis of criminal or detrimental events associated with artificial intelligence—including instances of "algorithmic bias" (such as discrimination in facial recognition or credit scoring), "safety deficiencies" (including accidents involving autonomous vehicles), and "misuse or unintended consequences"—such as the generation of malicious content by chatbots or the application of deepfake technology—reveals that in none of these cases has AI autonomously perpetrated a criminal act devoid of human influence, design, or training. The main causes of these events are either deliberate exploitation of the functionalities of the system by individuals, human error during the design and development phases, inadequate training data, or human error during the operational stage. This empirical study validates the lack of autonomy and agency needed to assign independent criminal liability to artificial intelligence in its current state. 4. Underlying empirical data and theoretical analysis, this paper shows the limits of the "autonomy" criterion for autonomous criminal liability for artificial intelligence. Though their capacity to learn, adapt, and show originality is rising, intelligent systems still depend on simple algorithms and data and are constrained by human-defined goals. More precisely reflecting the present state of this technology are the terms "bounded autonomy" and "heteronomous autonomy". Notwithstanding the technological difficulties and future developments in artificial intelligence requiring a constant review of legal presumptions, the technical features of artificial intelligence and accepted criminal law doctrines prevent these systems from being acknowledged as having independent criminal agency. The general international laws and rules reflect this result. Given the complexity and risks associated with every system, the Iranian lawmaker should give the establishment of accurate and efficient frameworks for spotting and assigning liability to human agents—including designers, developers, providers, and users—for crimes and damages resulting from the operation of artificial intelligence top priority rather than on building an autonomous criminal persona for this technology.
How To Cite	Shiri, Abbas; Barzegar, Mohammadreza (2026). Autonomy: The Indefensible Assumption of Artificial Intelligence Criminal Liability Theory. <i>Comparative Law Review</i> , 17 (1), 145-171. DOI: https://doi.com/10.22059/jcl.2025.390243.634723
DOI	10.22059/jcl.2025.390243.634723
Publisher	The author(s) retain the copyright.

