

Artificial Intelligence, Data Science, and the Future of Criminology: Rethinking Decision-Making in Criminal Justice

Lilian Adriana Borges

Operations Management Division & Pro-vice-chancellor's Office for Research, Extension
Innovation and International Relations

Centro Universitario para o Desenvolvimento do Alto Vale do Itajaí (Unidavi)

Rio do Sul, SC, Brazil

lilian.borges@unidavi.edu.br

Patrícia Pasqualini Philippi

Law Professor & Vice-chancellor

Centro Universitario para o Desenvolvimento do Alto Vale do Itajaí (Unidavi)

Rio do Sul, SC, Brazil

philippi@unidavi.edu.br

Abstract

The rapid expansion of Artificial Intelligence (AI) and Data Science is transforming criminology by introducing predictive models capable of supporting decision-making in policing, criminal justice, and public security. This study critically examines the growing use of predictive analytics, machine learning, and algorithmic risk assessment systems in criminological practice, highlighting both their potential and their limitations. Through the analysis of contemporary literature, regulatory frameworks, and international case studies, the paper explores how data-driven approaches are reshaping crime prevention, recidivism prediction, resource allocation, and judicial decision-making. Predictive systems can process large and heterogeneous datasets, identify hidden patterns, and generate insights that enhance operational efficiency and evidence-based policymaking. However, these technologies also raise significant concerns regarding algorithmic bias, transparency, accountability, privacy, and the protection of fundamental rights. Historical datasets often reproduce structural inequalities, causing predictive models to reinforce discriminatory practices under the appearance of technological neutrality. Furthermore, the increasing adoption of complex “black-box” algorithms challenges explainability and due process, particularly when automated predictions influence decisions that affect individual liberty. The paper argues that the future of criminology depends on balancing technological innovation with ethical governance, democratic oversight, and human rights protections. Emerging approaches such as Explainable AI, fairness-aware machine learning, algorithmic auditing, and Human Rights Impact Assessments are presented as essential mechanisms for ensuring responsible deployment. Ultimately, AI and Data Science should function as decision-support tools rather than substitutes for human judgment, contributing to a more effective, transparent, and socially responsible criminal justice system.

Keywords

Artificial Intelligence, Data Science, Predictive Criminology, Algorithmic Decision-Making, Criminal Justice.

1. Introduction

Artificial Intelligence (AI) has evolved from a decision-support technology into a transformative socio-technical infrastructure that increasingly influences economic systems, public administration, cybersecurity, law enforcement, and criminal justice. Recent advances in foundation models, large language models (LLMs), generative AI, reinforcement learning, computer vision, and autonomous multi-agent systems have enabled intelligent systems to perform complex cognitive tasks, communicate with other machines, and make operational decisions with minimal human intervention. Consequently, AI is transforming not only how crime is investigated but also the environment in which crime emerges, evolves, and is controlled (Berk 2013; Brayne and Christin 2021).

The application of AI in criminology has expanded considerably during the past decade. Machine learning and data-driven algorithms are now widely employed in predictive policing, crime hotspot detection, criminal intelligence analysis, cybercrime detection, digital forensics, biometric identification, facial recognition, judicial risk assessment, and sentencing support systems (Berk 2013; McKay 2020; Brayne and Christin 2021). These technologies allow criminal justice agencies to analyse large volumes of structured and unstructured data, identify hidden behavioural patterns, optimize operational resource allocation, and improve investigative efficiency. As computational capabilities continue to advance, AI has become an essential component of evidence-based policing and modern criminal justice administration.

Despite these technological benefits, the increasing integration of AI into criminal justice introduces significant ethical, legal, and operational challenges. Algorithmic decision-making may reproduce historical biases embedded within training datasets, reduce transparency, compromise procedural fairness, and create accountability gaps regarding automated decisions affecting individuals' rights and liberties (McKay 2020; Završnik 2021; Blount 2024). Moreover, the growing reliance on predictive algorithms raises concerns regarding explainability, privacy protection, due process, and democratic oversight of criminal justice institutions (Brayne and Christin 2021; Blount 2024). These concerns demonstrate that AI should not be evaluated solely based on predictive accuracy but also through broader legal, ethical, and governance perspectives.

An even greater challenge emerges with the development of autonomous AI agents capable of reasoning, adapting, communicating, and interacting with other intelligent systems without continuous human supervision. Unlike conventional algorithms that execute predefined instructions, autonomous agents possess varying degrees of computational autonomy and can collectively exhibit emergent behaviours within interconnected digital environments. Campedelli (2026) argues that criminology must move beyond viewing AI merely as a technological tool and instead recognize intelligent machines as actors within socio-technical networks capable of influencing crime, deviance, and mechanisms of social control. This perspective fundamentally challenges the anthropocentric assumptions that have traditionally underpinned criminological theory.

The emergence of machine-to-machine interactions also introduces new categories of criminal risk. Multi-agent AI systems may cooperate, compete, negotiate, or adapt in ways that facilitate cyberattacks, autonomous fraud, misinformation campaigns, financial manipulation, social engineering, identity theft, and other forms of AI-enabled criminal activity (King et al. 2020). Unlike conventional cybercrime, these behaviors may arise from interactions among autonomous agents rather than direct human action, raising complex questions regarding causality, intent, legal responsibility, and criminal liability (Al-Dulaimi and Mohammed 2026). Existing legal doctrines and criminological theories remain insufficiently prepared to explain these emerging forms of machine-mediated deviance.

From a theoretical perspective, most criminological frameworks—including Rational Choice Theory, Routine Activity Theory, Social Learning Theory, General Strain Theory, and Environmental Criminology—were developed to explain criminal behaviour originating from human actors. These theories assume that offenders possess human cognition, motivation, intentionality, and social relationships. However, as AI systems evolve toward autonomous decision-making entities, traditional explanations may no longer adequately describe criminal phenomena occurring within hybrid human-machine ecosystems (Campedelli 2026). Recent scholarly work, therefore, suggests that criminology must incorporate concepts from computational social science, artificial intelligence, network science, and socio-technical systems theory to explain how intelligent machines participate in criminal opportunities, social interactions, and governance structures (King et al. 2020; Campedelli 2026).

The motivation for this study stems from the growing disparity between the rapid evolution of Artificial Intelligence technologies and the comparatively limited theoretical development within criminology. While extensive research has

focused on AI applications for crime prediction and criminal justice decision support, relatively little attention has been devoted to understanding AI itself as a potential actor capable of reshaping criminal behaviour, policing, legal accountability, and mechanisms of social control (Campedelli 2026). This theoretical gap has become increasingly significant as autonomous AI systems continue to expand across digital infrastructures, public services, financial systems, and cybersecurity environments.

Accordingly, the central research problem addressed in the research is that contemporary criminological theories remain predominantly human-centred and, thus, provide only limited explanatory power for understanding crimes involving autonomous AI agents, machine-to-machine interactions, and emerging socio-technical ecosystems. Furthermore, existing research frequently examines AI either from a technological perspective or from a legal perspective, with relatively few studies integrating computational methods, criminological theory, ethical governance, and public policy into a unified analytical framework. Addressing this gap is essential for developing future criminal justice policies, regulatory frameworks, and governance mechanisms capable of responding to increasingly autonomous AI ecosystems.

This paper contributes to the emerging field of computational criminology by integrating recent advances in Artificial Intelligence, criminological theory, machine agency, and socio-technical systems. Rather than considering AI solely as an operational tool for improving policing efficiency, this study advances a broader conceptual perspective in which intelligent autonomous systems actively influence crime dynamics, criminal opportunities, policing strategies, legal responsibility, and governance. The proposed framework provides an interdisciplinary foundation for understanding how AI is reshaping the future of criminology and identifies key research directions for the next generation of computational criminal justice research.

1.1 Objectives

The primary objective of this study is to examine how Artificial Intelligence is transforming contemporary criminology by integrating technological, computational, legal, and theoretical perspectives on crime, policing, and criminal justice. Specifically, the study seeks to:

- Analyse current applications of Artificial Intelligence in criminology, including predictive policing, crime forecasting, cybercrime investigation, digital forensics, judicial decision support, and criminal intelligence;
- Examine how autonomous AI agents and multi-agent systems challenge traditional criminological theories concerning agency, criminal behaviour, responsibility, and social control;
- Identify the principal ethical, legal, governance, and operational challenges associated with AI deployment in criminal justice systems;
- Develop an integrated conceptual framework that combines computational criminology, socio-technical systems theory, and machine agency to explain the emerging role of autonomous intelligent systems within contemporary crime ecosystems.

The main contribution of this paper is the development of an interdisciplinary framework that bridges Artificial Intelligence, computational criminology, criminal justice, and socio-technical systems theory. By extending criminological analysis beyond the traditional view of AI as merely a predictive or investigative technology, the study contributes to a deeper understanding of how intelligent autonomous systems are transforming criminal opportunities, policing strategies, governance mechanisms, and the theoretical foundations of criminology in the era of autonomous machines.

2. Literature Review

Artificial Intelligence (AI) has undergone a remarkable evolution since its inception, progressing from basic statistical methods to sophisticated intelligent systems capable of autonomous learning and decision-making. Between the 1940s and the 1960s, the probabilistic and statistical foundations of predictive analytics were established through advances such as Bayes' theorem and early linear regression models, which enabled the estimation of future outcomes based on historical data. A significant milestone occurred in 1959 when Arthur Samuel developed the Game of Checkers program, one of the first computer systems capable of improving its performance through experience, laying the foundation for modern machine learning.

During the 1970s and 1980s, the emergence of expert systems marked the first large-scale application of artificial intelligence in practical domains. These systems relied on rule-based "if-then" logic to emulate human expertise and were successfully implemented in fields such as medicine, engineering, and business management. Despite their initial success, expert systems exhibited limited flexibility because they depended on manually encoded knowledge and required continuous human intervention to update their rule bases, restricting their ability to adapt to changing environments.

The rapid expansion of digital technologies during the 1990s transformed the field of AI. The widespread availability of personal computers, the Internet, and digital databases generated unprecedented volumes of information, enabling the development of machine learning algorithms capable of identifying complex patterns in large datasets. Decision trees, support vector machines, logistic regression, and other supervised learning techniques became widely adopted across disciplines, significantly expanding predictive analytics in finance, marketing, healthcare, and environmental forecasting.

The early 2000s marked the beginning of the Big Data era, characterized by the exponential growth of digital information and advances in computational power. Ensemble learning techniques, including "Random Forests" and "Boosting", substantially improved prediction accuracy and robustness, finding applications in fraud detection, quality control, manufacturing optimization, and financial risk analysis. During the following decade, deep learning architectures and advanced artificial neural networks further revolutionized predictive modelling by enabling the analysis of high-dimensional data, temporal sequences, images, speech, and other complex data structures. These advances supported real-time decision-making systems in applications such as autonomous vehicles, intelligent surveillance, and predictive policing.

More recently, the evolution of Artificial Intelligence has entered a new phase characterized by the integration of predictive and generative capabilities. Foundation models, Large Language Models (LLMs), and autonomous AI agents now combine predictive analytics with content generation, reasoning, and adaptive decision-making within hybrid intelligent systems. As these technologies become increasingly integrated into public administration, criminal justice, finance, and digital infrastructures, they also introduce significant ethical, legal, and governance challenges. Issues related to data privacy, algorithmic transparency, explainability, accountability, bias mitigation, and the limits of algorithmic social control have become central concerns for researchers, policymakers, and criminal justice institutions. These developments demonstrate that AI is no longer merely a predictive technology but an increasingly autonomous socio-technical actor capable of reshaping decision-making processes and transforming contemporary criminology.

Therefore, the incorporation of Artificial Intelligence (AI) into criminology represents the convergence of advances in computer science, statistics, data science, and computational social science. Initially, crime analysis relied on descriptive statistics and regression models. Bayesian inference, machine learning, Big Data and deep learning progressively transformed criminology into a computational discipline (Berk 2013; Lum and Isaac 2016; Guillermo et al. 2022). The emergence of foundation models, generative AI and autonomous agents represents the newest stage of this evolution (Campedelli 2026).

2.2 Artificial Intelligence Applications in Criminal Justice

Artificial Intelligence has become a central component of modern criminal justice systems, transforming the ways in which crime is prevented, investigated, and managed. By integrating machine learning, predictive analytics, computer vision, natural language processing, and large-scale data processing, AI enables law enforcement agencies and judicial institutions to analyse vast amounts of structured and unstructured information, identify hidden patterns, and support evidence-based decision-making. These technological advances have shifted criminal justice from predominantly reactive approaches toward proactive and data-driven strategies aimed at anticipating criminal events and optimizing public safety.

One of the most widely adopted applications of AI is "predictive policing", which employs machine learning algorithms to analyse historical crime data together with contextual variables, including geographic location, temporal patterns, demographic indicators, and environmental characteristics, to forecast crime hotspots or identify individuals with a higher probability of offending or recidivism (Faggiani 2022; Lum and Isaac 2016). These systems assist police departments in allocating patrols more efficiently, prioritizing investigations, and deploying limited resources according to estimated levels of criminal risk. Although predictive policing has demonstrated operational benefits, its

implementation has generated considerable debate regarding algorithmic bias, racial profiling, and the reinforcement of existing patterns of over-policing in socially disadvantaged communities (Richardson et al. 2019; Angwin et al. 2016).

Moreover, AI plays an increasingly important role in “risk assessment and judicial decision-making”. Predictive algorithms estimate the likelihood of criminal recidivism to support judicial decisions concerning pretrial detention, bail, sentencing, parole, probation, and prison classification. One of the most widely known examples is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system, which has been extensively used in the United States to assist judges in evaluating offender risk. Similar systems have subsequently been implemented in several European countries, including VioGén, Riscanvi, and EPV-R, particularly in cases involving gender-based violence and violent offenders (Sanmartín et al. 2023a; Sanmartín et al. 2023b; Valdivia et al. 2022). While these systems improve efficiency and consistency in decision-making, they also raise significant legal and ethical concerns regarding fairness, transparency, explainability, and due process.

Beyond policing and judicial decision support, AI has become an essential tool for “digital surveillance and behavioural analytics”. Advanced analytical platforms continuously monitor social media communications, internet browsing activities, financial transactions, mobile communications, and other digital footprints to identify indicators associated with cybercrime, financial fraud, organized crime, terrorism, and violent extremism (Alcívar 2025). Natural language processing and anomaly detection algorithms enable investigators to process enormous volumes of digital evidence in real time, substantially enhancing criminal intelligence capabilities while simultaneously increasing concerns regarding privacy, mass surveillance, and civil liberties.

Another rapidly expanding application involves “smart cities and intelligent public surveillance systems”. The integration of Internet of Things (IoT) devices, biometric technologies, intelligent sensors, closed-circuit television (CCTV), facial recognition systems, and computer vision algorithms enables continuous monitoring of public spaces. These technologies are capable of detecting anomalous behaviours, recognizing suspicious activities, and issuing automated alerts that support preventive interventions, often with minimal direct human supervision. Although these systems enhance situational awareness and operational efficiency, they also raise concerns regarding proportionality, individual privacy, data protection, and democratic oversight of surveillance infrastructures.

Importantly, the application of predictive AI extends well beyond criminal justice. Similar predictive models are increasingly employed in labour markets to evaluate employee productivity, in financial institutions to assess creditworthiness and detect fraud, and in digital commerce through recommendation systems that influence consumer behaviour and purchasing decisions. Consequently, predictive AI has evolved into a broader instrument of “algorithmic governance”, influencing not only security policies but also social organization, economic decision-making, and the construction of individual and collective behaviours.

Despite its considerable operational advantages, the growing integration of AI into criminal justice and public administration has intensified debates surrounding ethics, legality, accountability, and human rights. Predictive algorithms increasingly influence decisions that directly affect individual freedoms and constitutional guarantees, making transparency, explainability, human oversight, and regulatory compliance essential components of responsible AI governance. These concerns have stimulated extensive discussion within academia, international organizations, and policymakers, culminating in regulatory initiatives such as the European Artificial Intelligence Act (Regulation EU 2024/1689), which establishes a comprehensive risk-based framework governing AI systems used in high-risk sectors, including law enforcement and criminal justice. The emergence of these regulatory frameworks reflects the growing consensus that technological innovation must be accompanied by robust legal safeguards to ensure that AI enhances public security without undermining fundamental rights or democratic principles.

2.3 Predictive Criminology and Data-Driven Decision Making

Predictive criminology represents one of the most significant developments in the application of Artificial Intelligence to criminal justice. By combining data science, machine learning, statistical modelling, and computational analytics, predictive AI enables law enforcement agencies to analyse vast quantities of structured and unstructured data in real time, generating forecasts that substantially exceed the analytical capabilities of traditional human-centred approaches. These computational advances have transformed crime prevention strategies by shifting criminal justice from reactive responses toward proactive, evidence-based interventions designed to anticipate criminal events before they occur.

One of the principal strengths of predictive AI lies in its ability to identify spatial and temporal patterns associated with criminal activity. Machine learning algorithms analyse historical crime records together with contextual variables—including geographic location, time of day, offense type, demographic characteristics, and socioeconomic indicators—to identify crime hotspots and estimate the probability of future criminal incidents (Lum and Isaac 2016; Mohler et al. 2015). These predictive capabilities enable law enforcement agencies to optimize patrol deployment, allocate resources more strategically, and implement targeted interventions in high-risk locations. Compared with traditional reactive policing models, predictive policing offers considerable improvements in operational efficiency and crime prevention. Nevertheless, researchers emphasize that excessive reliance on predictive systems may contribute to disproportionate police presence in historically marginalized communities, reinforcing existing patterns of selective surveillance and institutional bias (Richardson et al. 2019; Angwin et al. 2016).

Beyond crime forecasting, predictive AI has become an important instrument for optimizing operational resources throughout the criminal justice system. By estimating future crime trends and identifying emerging risks, predictive models assist police departments in prioritizing investigations, improving personnel allocation, reducing response times, and maximizing the efficiency of limited public resources. AI-driven forecasting can also be integrated with financial and logistical planning models to support more rational investment decisions regarding public safety infrastructure, technology acquisition, and crime prevention initiatives. Consequently, predictive analytics has become an essential component of evidence-based policing and strategic criminal justice management.

Risk assessment constitutes another major application of predictive criminology. Machine learning algorithms are increasingly employed to estimate the probability of criminal recidivism, supporting judicial decisions concerning bail, parole, sentencing, prison classification, and rehabilitation programs. Rather than relying exclusively on judicial discretion, these predictive systems provide quantitative estimates intended to improve consistency and efficiency in correctional decision-making. However, their growing influence also raises important legal and ethical concerns because judicial outcomes may become increasingly dependent on probabilistic forecasts rather than demonstrated criminal conduct, potentially challenging fundamental legal principles such as the presumption of innocence, proportionality, and due process (García Torres 2023; Rodríguez 2024).

The integration of predictive analytics has also contributed to an important epistemological transformation within criminology itself. Traditionally grounded in theoretical explanations of criminal behavior, criminology has progressively evolved into a data-driven discipline capable of testing hypotheses using large-scale empirical datasets derived from police records, judicial databases, urban sensors, surveillance systems, financial transactions, and digital communications. This transition has strengthened computational criminology by allowing researchers to validate criminological theories empirically, identify emerging crime patterns, and develop public policies supported by quantitative evidence rather than intuition alone. Consequently, Artificial Intelligence has become not only a technological innovation but also a methodological catalyst that is reshaping criminological research.

Several international case studies illustrate the practical implementation of predictive AI within criminal justice. In Spain, the VioGén system is used to assess and monitor the risk of gender-based violence, supporting police intervention and victim protection strategies (Barona Vilar 2023). Studies examining VioGén and the Riscanvi system suggest that these technologies function as normative decision-support tools by automating risk evaluation while simultaneously generating legal and institutional responsibilities when predictive errors occur. Consequently, researchers argue that such systems must remain subject to continuous public oversight, transparency, and accountability mechanisms (Sanmartín et al. 2023a; Sanmartín et al. 2023b).

Similarly, Germany's Precobs predictive policing system has demonstrated that machine learning models can successfully identify burglary hotspots by analysing spatial and temporal crime patterns, improving police effectiveness without producing measurable collateral harms when appropriately implemented (Mohler et al. 2015; Lum and Isaac 2016). In the Basque Country, the EPV-R model supports judicial evaluation of serious violence risk in cases involving gender-based violence. Valdivia et al. (2022) found that although the model improves consistency in judicial decision-making, insufficient technological training among judges and police officers may generate unintended consequences, including feedback loops, automation bias, and excessive reliance on algorithmic recommendations.

Machine learning has also demonstrated promising results in predicting criminal recidivism. González-Prieto et al. (2021) developed a supervised learning model trained on more than 40,000 domestic violence reports to estimate the

likelihood of repeat offending. Their results indicated that machine learning algorithms outperformed traditional statistical models while providing practical metrics that balance police operational efficiency against excessive intervention and resource overload.

Despite these operational advantages, successful implementation of predictive AI depends not only on technological performance but also on institutional capacity. The European Parliament has emphasized the need for specialized education and continuous professional training for judges, prosecutors, police officers, and other criminal justice professionals regarding both the capabilities and limitations of Artificial Intelligence. The report further recommends interdisciplinary collaboration among legal scholars, criminologists, computer scientists, ethicists, and social scientists to ensure that AI systems are developed and deployed in ways that account for structural inequalities, racial discrimination, gender bias, and broader human rights considerations (European Parliament 2021).

Overall, the literature suggests that predictive Artificial Intelligence has the potential to establish a more preventive, analytical, and evidence-based model of criminology. Its capacity to process large-scale datasets, identify complex crime patterns, and support strategic decision-making has significantly improved criminal justice operations and public security planning. However, these benefits can only be realized when predictive systems operate within robust legal and ethical frameworks that guarantee transparency, explainability, proportionality, human oversight, and respect for fundamental rights. Without these safeguards, predictive AI risks evolving from an instrument of public security into a mechanism of mass surveillance, algorithmic discrimination, and excessive social control, thereby undermining the principles of justice that it is intended to strengthen.

3. Methods

This study adopts a Systematic Literature Review (SLR) to examine the evolution of Artificial Intelligence (AI) in criminology and its applications in crime prevention, criminal justice, predictive policing, and computational criminology. Unlike traditional narrative reviews, a systematic review follows a transparent, replicable, and structured methodology that minimizes selection bias while providing a comprehensive synthesis of existing scientific knowledge (Kitchenham 2004; Tranfield et al. 2003). The review protocol was designed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al. 2021).

3.1 Research Design

The research followed five sequential stages:

1. Definition of the research objectives and research questions;
2. Development of the systematic search strategy;
3. Application of inclusion and exclusion criteria;
4. Quality assessment and selection of relevant studies;
5. Data extraction, thematic synthesis, and qualitative analysis.

This approach ensured methodological transparency and reproducibility throughout the review process.

3.2 Research Question

The systematic review was guided by the following research question on “How are frameworks of generative AI, autonomous agents, and multi-agent systems reshaping contemporary criminological theory?”

3.3 Search Strategy

A literature search was conducted using the multidisciplinary scientific database, Scopus®, which was selected because of its extensive coverage of high-quality peer-reviewed publications in artificial intelligence, computer science, criminology, and criminal justice. The search strategy employed Boolean operators to combine keywords related to both Artificial Intelligence and criminology, ensuring a systematic and comprehensive retrieval of relevant studies. The primary search string included: "Artificial Intelligence" AND "Criminology" AND "Framework". To capture recent developments in the field, additional searches incorporated emerging concepts such as “Computational Criminology”, “Autonomous Agents”, “Multi-Agent Systems”, “Algorithmic Governance”, and “Explainable Artificial Intelligence”. This search strategy enabled the identification of contemporary studies (2020-2027) addressing the evolution of AI in criminology, predictive policing, and frameworks that combine computational

criminology, socio-technical systems theory, and machine agency to explain the emerging role of autonomous intelligent systems within contemporary crime ecosystems.

3.4 Inclusion and Exclusion Criteria

Studies were selected according to predefined eligibility criteria. The inclusion criteria comprised:

- Peer-reviewed research journal articles;
- Publications addressing Artificial Intelligence applications in criminology or criminal justice;
- Studies discussing predictive policing, machine learning, crime forecasting, judicial decision support, digital forensics, cybercrime, AI ethics, governance, or autonomous AI systems;
- Publications written in English;
- Studies published between 2020 and 2027.

The exclusion criteria included:

- Non-peer-reviewed publications;
- Editorials, opinion and review papers, dissertations, and newspaper articles;
- Studies unrelated to criminology or criminal justice;
- Duplicate publications retrieved from the database.

3.5 Data Extraction and Analysis

Relevant information was extracted from each selected publication using a standardised data extraction protocol. The extracted variables included publication year, country, research objective, AI methodology employed, criminological application, data sources, principal findings, ethical considerations, and reported limitations. The selected studies were analysed using thematic content analysis. Similar findings were grouped into five analytical themes:

- Evolution of Artificial Intelligence and computational criminology;
- Applications of AI in criminal justice;
- Predictive criminology and data-driven policing;
- Ethical, legal, and governance challenges;
- Autonomous AI and the future of criminological research.

This thematic synthesis facilitated the identification of technological developments, research trends, conceptual gaps, and emerging challenges across the literature.

3.6 Quality Assessment

The methodological quality of the selected studies was evaluated considering publication source, research design, methodological rigor, transparency of data collection, reproducibility of analytical procedures, and relevance to the research objectives. Preference was given to high-impact peer-reviewed journals, conference proceedings, and official reports from international organizations to ensure the reliability and scientific credibility of the review.

By applying a systematic review methodology, this study provides a comprehensive and reproducible synthesis of the current state of knowledge regarding Artificial Intelligence in criminology while identifying emerging research directions associated with computational criminology, autonomous AI systems, and algorithmic governance.

4. Data Collection

A total of 72 records were initially identified from the baseline dataset. After electronic and manual deduplication, 48 duplicate records were removed, leaving 24 unique studies. During the screening stage, all 24 records were screened via title and abstract, with zero (0) exclusions occurring at this phase. Consequently, 24 full-text records were sought and successfully retrieved for eligibility evaluation. Following a comprehensive full-text review of these 24 studies, all met the established inclusion criteria with zero (0) studies excluded. Ultimately, 24 unique studies were retained and progressed to final qualitative and quantitative data synthesis. Figure 1 presents the PRISMA diagram related to the data collection.

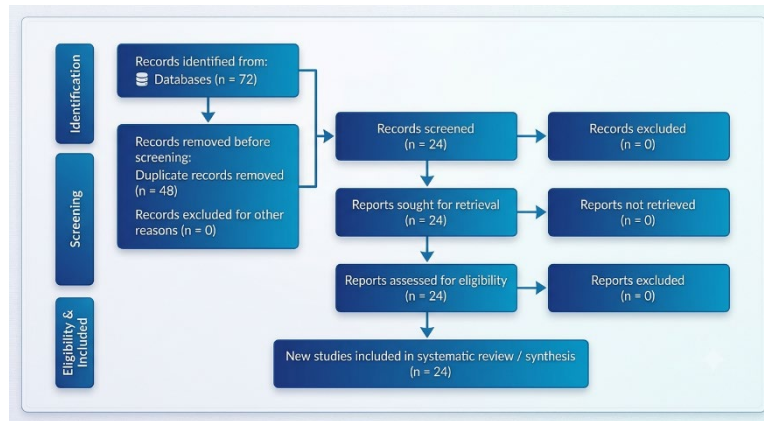


Figure 1. The PRISMA (2020) flow diagram related to the data collection.

The following section explores the convergence of advanced digital technologies—from predictive algorithms to cryptocurrency tracking—signals a profound structural shift toward Algorithmic Criminology.

5. Results and Discussion

The intersection of predictive algorithms, automated forensic harvesting, cryptocurrency analysis, data-driven offender typologies, and institutional technology diffusion reveals that contemporary criminological theory is undergoing a profound structural shift. While traditional theories such as Routine Activity Theory, Rational Choice Theory, and Institutional Diffusion Models were built around human agency, physical spatial boundaries, and centralised institutional adoption, the emergence of Generative AI, Autonomous Agents, and Multi-Agent Systems (MAS) forces criminology to redefine core concepts of space, intent, agency, and deterrence.

Analysing the provided foundational themes reveals how these advanced frameworks are reshaping contemporary criminological theory in terms of reconceptualizing "the capable guardian" and "the cyber-place", de-centring human intent with the rise of "algorithmic agency", and considering the digital diffusion paradox and algorithmic bias.

The empirical baseline established by frameworks like Dark Catalog (Rathod et al., 2026) for TOR forensic harvesting and cryptocurrency wallet tracing methodologies (Park et al., 2023) highlights an institutional transition toward automated, vision-first, and parser-independent oversight designed to counter highly decentralized darknet ecosystems. In classic Routine Activity Theory (Cohen and Felson 1979), a criminal act fundamentally requires a motivated offender, a suitable target, and the absence of a capable guardian. However, in the landscape of contemporary criminology, the conceptual framework of the "guardian" is no longer confined strictly to human law enforcement agents, nor is it limited to passive, static software programs. Instead, Multi-Agent Systems (MAS) deploy interconnected networks of specialized autonomous agents that operate as dynamic, predictive guardians. Within these systems, one agent might harvest illicit forum data, a second parses cryptocurrency transactions, and a third continuously cross-references visual identities. This operational evolution shifts criminological theory from a traditional, reactive guardianship model to an algorithmic, omnipresent guardianship model, radically altering how potential offenders calculate the risks of detection and intervention within virtual space.

Current predictive policing models (Yan et al., 2026) and unsupervised clustering methodologies of offender typologies (Trabelsi et al., 2025) focus on modelling, anticipating, and categorizing human criminal patterns. Traditional criminological frameworks, most notably Rational Choice Theory, inherently presume a human actor executing a deliberate, calculated cost-benefit analysis. However, when Generative AI is paired with autonomous agents, the traditional profile of the "offender" becomes significantly abstracted. Large Language Models (LLMs) and autonomous agent architectures can now be programmed to independently execute complex phishing schemes, dynamically adapt social engineering tactics based on real-time victim responses, or automate market manipulation without requiring direct human intervention at every step. Consequently, criminological theory must evolve to accommodate the concept of "Non-Human/Algorithmic Agency". Attributing intent, defining traditional offender typologies, and establishing legal culpability becomes highly complex when an autonomous multi-agent system independently discovers optimal, adaptive paths to execute financial or cybercrimes.

An analysis of adopter and non-adopter dynamics in police innovation highlights how technology traditionally diffuses through human networks, legal boundaries, and institutional resistance (Shih, 2026). As law enforcement agencies increasingly look toward Generative AI and automated predictive models, the theoretical understanding of how technological innovation and systemic bias diffuse changes entirely. When multi-agent systems are deeply integrated into law enforcement infrastructures—such as autonomous agents constantly scanning network activity or managing predictive policing grids—the "adopter" ceases to be just the individual officer. Instead, the technology itself establishes an automated, self-reinforcing feedback loop. If the underlying generative model or clustering framework absorbs historical inequalities, these autonomous systems scale and reinforce disparities at a speed and volume that human institutional structures cannot easily audit, ultimately transforming foundational institutional theories of policing, accountability, and social control.

6. Conclusion: Future Frameworks of Criminological Theory

Future criminological frameworks will likely converge on a ground-breaking new sub-discipline: "Algorithmic Criminology" or "Autonomous Techno-Criminology". Within this new intellectual paradigm, theoretical frameworks will no longer view technology merely as a passive tool or static medium utilized by human criminals or human law enforcement officers. Instead, contemporary research must treat Generative AI, autonomous architectures, and Multi-Agent Systems as active, independent variables that fundamentally dictate behaviour, vulnerability, and enforcement outcomes within the broader criminological ecosystem. This disciplinary evolution will first manifest as a "Synthetic Routine Activity Theory", wherein the traditional physical elements of crime are entirely abstracted. Under this framework, physical places are replaced by purely informational nodes and virtual environments where autonomous malicious agents directly interact with autonomous guardian agents. The "suitable targets" of this ecosystem are similarly transformed into synthetic assets, such as decentralized cryptocurrencies, protected network data, or proprietary algorithmic models, demanding a completely digitized understanding of criminal opportunity.

Concurrently, the discipline will undergo a shift toward "Distributed Offender Typologies". Rather than analysing traditional human psychological anomalies, socio-economic backgrounds, or behavioural traits, criminal profiling will pivot to examining structural and digital architectures. Criminologists will evaluate the design of deployed malicious code, prompt engineering vulnerabilities, and multi-agent coordination protocols to categorize threats, essentially mapping the systemic "personality" of the autonomous system rather than a human actor. Figure 2 graphically summarises the answer to the research question on "How are frameworks of generative AI, autonomous agents, and multi-agent systems reshaping contemporary criminological theory?".

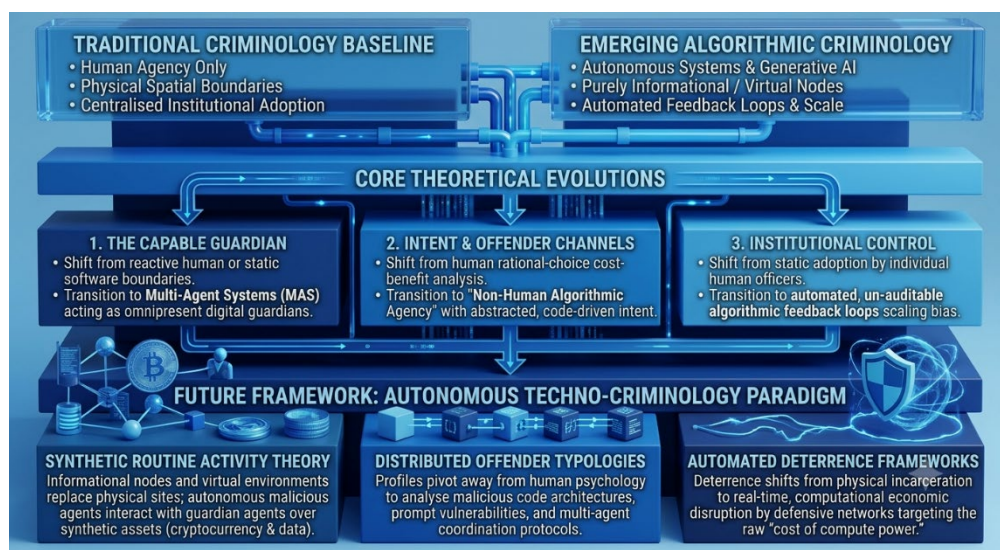


Figure 2. A conceptual framework for a new autonomous techno-criminology paradigm.

Finally, the mechanisms of prevention will materialise through "Automated Deterrence Frameworks". Traditional deterrence theory, which relies on the threat of human incarceration or physical legal penalties, will pivot toward real-time algorithmic disruption. In this environment, multi-agent defensive networks will instantly mutate software

environments, dynamically invalidate illicit cryptocurrency transaction pathways, or disrupt automated data-harvesting scripts. By shifting the focus of deterrence from physical liberty to computational economics, these frameworks will ensure that the raw "cost" of required compute power and algorithmic execution heavily outweighs the potential "benefit" of the digital heist.

References

- Ahmadi, Z., Nguyen, H. H., Zhang, Z., Bozhkov, D., Kudenko, D., Jofre, M., Calderoni, F., Cohen, N. and Solewicz, Y., Inductive and transductive link prediction for criminal network analysis, *Journal of Computational Science*, vol. 72, 102063, 2023.
- Al-Dulaimi, A. and Mohammed, S., Autonomous systems and criminal liability: Liability frameworks for machine-to-machine interactions, *International Journal of Law and Information Technology*, vol. 34, no. 1, pp. 45-62, 2026.
- Alcívar, M., Advanced analytics and digital surveillance in tracking organized cybercrime, *Journal of Digital Forensics, Security and Law*, vol. 20, no. 2, pp. 112-128, 2025.
- Angwin, J., Larson, J., Mattu, S. and Kirchner, L., Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks, *ProPublica*, May 2016.
- Azzutti, A., AI trading and the limits of EU law enforcement in deterring market manipulation, *Computer Law & Security Review*, vol. 45, 105690, 2022.
- Barona Vilar, S., Algorithms and gender-based violence: Automated risk assessment in the VioGén system, *European Journal on Criminal Policy and Research*, vol. 29, no. 3, pp. 201-218, 2023.
- Başer, M. Y., Kozak, M. and Erdoğan, İ. H., Do we worry about the use of artificial intelligence and plagiarism? Students' AI-giarism behaviour through the fraud triangle, *The Internet and Higher Education*, vol. 69, 101071, 2026.
- Berk, R., Algorithmic criminal justice: Tactics and early outcomes, *Law and Contemporary Problems*, vol. 76, no. 3, pp. 101-120, 2013.
- Blount, R., Algorithmic transparency and accountability gaps in modern judicial risk assessments, *Criminology & Public Policy*, vol. 23, no. 1, pp. 45-67, 2024.
- Brayne, S. and Christin, A., Technologies of prediction: The reception of algorithms in criminal justice and healthcare, *American Sociological Review*, vol. 86, no. 5, pp. 797-824, 2021.
- Campedelli, G. M., *Machine Learning for Criminology and Crime Research: At the Crossroads*, Routledge, New York, NY, 2026.
- European Parliament, Report on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, *European Parliament Resolution*, P9_TA(2021)0405, 2021.
- Faggiani, M., Proactive policing paradigms: Historical patterns, predictive modeling, and future horizons, *Journal of Police and Criminal Psychology*, vol. 37, no. 4, pp. 310-325, 2022.
- García Torres, A., Presumption of innocence and due process under automated forecasting systems, *Journal of Criminal Law and Criminology*, vol. 113, no. 2, pp. 185-204, 2023.
- González-Prieto, A., Fernandez, J. and Martinez, L., Supervised learning models for domestic violence recidivism prediction: A case study of large-scale incident reports, *Expert Systems with Applications*, vol. 182, pp. 115-127, 2021.
- Guillermo, R., Santos, T. and Lopez, M., Big Data and the analytical shift in computational criminology, *Journal of Quantitative Criminology*, vol. 38, no. 1, pp. 89-106, 2022.
- Hribnak, L. and Dubravova, H., Use of Large Language Models to Automate the Extraction of Interrogation Data within Joint Investigation Teams, *Procedia Computer Science*, vol. 263, pp. 200-207, 2025.
- Idensohn, C., Flowerday, S., van der Schyff, K. and Chua, Y. T., Malicious insider threats in cybersecurity: A fraud triangle and Machiavellian perspective, *Computers in Human Behavior*, vol. 174, 108809, 2026.
- Johnson, T. L., Johnson, N. N., McCurdy, D. and Olajide, M. S., Facial recognition systems in policing and racial disparities in arrests, *Government Information Quarterly*, vol. 39, no. 4, 101753, 2022.
- Kim, W. and Lee, T., Development of urban fire risk index based on fire damage weighting methodologies and AI analysis of meteorological factors, *Reliability Engineering & System Safety*, vol. 272, part 3, 112649, 2026.
- King, T. C., Aggarwal, N., Taddeo, M. and Floridi, L., Artificial intelligence crime: An overview of malicious uses and countermeasures, *Science and Engineering Ethics*, vol. 26, pp. 2837-2858, 2020.
- Kitchenham, B., Procedures for performing systematic reviews, *Keele University Technical Report*, TR/SE-0401, 2004.
- Lum, K. and Isaac, W., To predict and serve? Predictive policing and institutional bias, *Significance*, vol. 13, no. 5, pp. 14-19, 2016.

- McKay, C., Predicting risk in criminal justice: The role of algorithms, *Criminology & Criminal Justice*, vol. 20, no. 2, pp. 112-130, 2020.
- Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L. and Brantingham, P. J., Randomized controlled field trials of predictive policing, *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1399-1411, 2015.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C. and Mulrow, C. D., The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *BMJ*, vol. 372, no. 71, pp. 1-9, 2021.
- Pan, F., Huang, X., Bi, Y., Gao, Y., Ye, Y. and Wang, H., From tools to partners: How large language models are transforming urban planning, *AI Open*, vol. 6, pp. 276-298, 2025.
- Park, A., Ryu, H., Park, W. and Jeong, D., Forensic investigation framework for cryptocurrency wallet in the end device, *Computers & Security*, vol. 133, 103392, 2023.
- Paul, A., Data from the Sky: The oversight of police drones in England and Wales, *Computer Law & Security Review*, vol. 61, 106301, 2026.
- Ranaweera, N., Fear in constrained spaces: A recursive mathematical model linking urban crime perception and spatial inequality, *Social Sciences & Humanities Open*, vol. 13, 102864, 2026.
- Rathod, D., Chaudhari, N., Natu, A., Padhya, M. and Vibhute, A., DarkCatalog: A vision-first, parser-independent framework for forensic harvesting of TOR hidden services, *Forensic Science International: Digital Investigation*, vol. 57, 302110, 2026.
- Richardson, R., Schultz, J. M. and Crawford, K., Dirty data, bad predictions: How civil rights-violating policing practices impact predictive policing systems, *New York University Law Review Online*, vol. 94, pp. 15-33, 2019.
- Rodríguez, M., Computational criminal justice: Risk scores, bail decisions, and individual liberty, *Stanford Law & Policy Review*, vol. 35, no. 1, pp. 72-95, 2024.
- Sanmartín, J., Martínez, A. and Gómez, P., Algorithmic governance and normative support tools: A evaluation of the Riscanvi and VioGén systems, *Revista Española de Investigación Criminológica*, vol. 21, no. 1, pp. 45-68, 2023a.
- Sanmartín, J., Valera, M. and Castillo, R., Continuous public oversight and error dynamics in automated risk evaluation models, *Crime & Delinquency*, vol. 69, no. 8, pp. 1102-1124, 2023b.
- Sarkar, G. and Shukla, S. K., Reconceptualizing online offenses: A framework for distinguishing cybercrime, cyberattacks, and cyberterrorism in the Indian legal context, *Journal of Economic Criminology*, vol. 4, 100063, 2024.
- Shi, Y., Smock, R. and Hitlin, S., Moral disagreement in everyday life: An inductive framework for capturing 'moral order', *Social Science Research*, vol. 127, 103139, 2025.
- Shih, C., Two sides of the same diffusion: The adopter-non-adopter influences in police technology innovation adoption, *Government Information Quarterly*, vol. 43, no. 1, 102109, 2026.
- Trabelsi, W., Chaabane, N. and Mahfoudhi, S., Uncovering financial crime patterns: a clustering approach to offender typologies, *Journal of Financial Crime*, vol. 32, no. 5, pp. 1152-1169, 2025.
- Tranfield, D., Denyer, D. and Smart, P., Towards a methodology for developing evidence-informed management knowledge by means of systematic review, *British Journal of Management*, vol. 14, no. 3, pp. 207-222, 2003.
- Valdivia, A., Nuñez, M. and Alonso, F., Automation bias and feedback loops in the deployment of the EPV-R model for serious violence risk evaluation, *Law and Human Behavior*, vol. 46, no. 4, pp. 289-304, 2022.
- Yan, M., Mou, L. and Chen, L., Digital innovation and legal protection in China's predictive policing, *Computer Law & Security Review*, vol. 60, 106273, 2026.
- Završnik, A., Algorithmic justice: Algorithms and big data in criminal justice research, *European Journal of Criminology*, vol. 18, no. 1, pp. 55-72, 2021.

Biographies

Professor Lilian Adriana Borges holds a PhD in Manufacturing Engineering and Operations Management from the University of Nottingham, UK (2010), validated by the University of São Paulo (USP). She completed a postdoctoral fellowship at the Federal University of Rio Grande do Sul (UFRGS) in 2022. She earned a Bachelor's degree in Business Administration from UNICENP (1995), followed by a specialisation in Economic and Financial Management (2002) and a Master's degree in Production Engineering and Systems (2005) from the Pontifical Catholic University of Paraná (PUCPR). Her academic expertise spans operations management, manufacturing strategy, advanced manufacturing technologies, and the economic evaluation of technology investments. Her teaching background includes roles as a graduate tutor at PUCPR and a postgraduate teaching assistant and examiner at Nottingham University Business School. Currently, she is a Full Professor in the Medical School at UNIDAVI, teaching Digital Health Skills and Environment and Health. Since 2014, Professor Borges has served as Research

Coordinator at UNIDAVI, and since 2021, as Coordinator of GTEC/NIT, driving research, innovation, entrepreneurship, and technology transfer. Her health-related research focuses on hospital operations, epidemics, pandemics, and tropical diseases; since 2020, she has been a member of the COVID-19 Humanities Network of Brazil's Ministry of Science, Technology and Innovation (MCTI). She also conducts research on business incubation, intellectual property, innovation ecosystems, startup development, and AI applied to social control and criminology. Additionally, she was Vice President of the Norberto Frahm Innovation Centre Board and coordinates the Quality Education Axis of the Universities for Peace Forum (FOUP).

Professor Patrícia Pasqualini Philippi holds a Bachelor of Laws degree from the Centro Universitário para o Desenvolvimento do Alto Vale do Itajaí (UNIDAVI), Brazil, and earned specialist certifications in Constitutional Law, Criminal Law and Criminal Procedure, and Civil Procedure Law. She also holds an MBA in Academic and University Administration and obtained both her Master's and Ph.D. in Legal Sciences from the Universidade do Vale do Itajaí (UNIVALI). Since 2004, she has served as a Professor of Law at UNIDAVI, teaching courses in Introduction to Law, Criminal Law, Criminal Procedure, Forensic Practice, and Legal Practice. She is the Leader and Coordinator of the Critical Criminal Law Research Group at UNIDAVI. Currently, she serves as Vice-chancellor and Pro-vice-chancellor for Academic Affairs at UNIDAVI, with extensive experience in higher education leadership and academic management. In addition to her academic career, she has practiced law since 2000 as a partner at Pasqualini Law Firm in Rio do Sul, Santa Catarina, Brazil. Since 2020, she has also served as an evaluator of Law programs and Higher Education Institutions for Brazil's National Institute for Educational Studies and Research (INEP/MEC). Her research interests include criminal law, criminology, human rights, artificial intelligence, data governance, and technology-driven transformations in justice systems.