Available online at: http://journal.uinsgd.ac.id/index.php/kh
Khazanah Hukum, Vol. 7 No. 3: 312-324

DOI: 10.15575/kh.v7i3.45201

Using Big Data in Criminal Investigations: Between Privacy and Efficiency

Oleksii Tavolzhanskyi[1]*, Olena Shumeiko[1], Oleksandr Burda[1], Kostiantyn Orobets[1], Maksym Struchaiev[1]

[1]Yaroslav Mudryi National Law University, Ukraine
*Corresponding Author Email: oleksii_tavolzhanskyi@edu-knu.com

ABSTRACT

The growing complexity of criminal activity and the exponential expansion of digital data necessitate the integration of Big Data technologies into criminal investigations. This paper examines the legal, technological, and ethical implications of using Big Data in criminal justice systems, with a focus on balancing investigative efficiency and individual privacy rights. The research applies a combination of philosophical and normative legal analysis, along with systemic and historical methods, to assess how these technologies are transforming investigative procedures. Findings highlight the potential of Big Data to enhance investigative accuracy, especially through data mining and predictive analytics, but also underscore serious risks related to data protection and regulatory ambiguity. The paper calls for clearer legal standards, international cooperation, and ethical frameworks to guide the application of Big Data in criminal proceedings. Furthermore, it emphasizes the need for institutional accountability and judicial oversight to prevent misuse, ensure transparency, and uphold the rule of law in an increasingly data-driven legal landscape.

Keywords: Big Data, Information, Criminal justice, Criminal Offences, Information Flow, Investigation.

ABSTRAK

Meningkatnya kompleksitas aktivitas kriminal dan ekspansi eksponensial data digital mengharuskan integrasi teknologi Big Data ke dalam investigasi kriminal. Artikel ini mengkaji implikasi hukum, teknologi, dan etika dari penggunaan Big Data dalam sistem peradilan pidana, dengan fokus pada penyeimbangan efisiensi investigasi dan hak privasi individu. Penelitian ini menerapkan kombinasi analisis hukum filosofis dan normatif, bersama dengan metode sistemik dan historis, untuk menilai bagaimana teknologi ini mengubah prosedur investigasi. Temuan menyoroti potensi Big Data untuk meningkatkan akurasi investigasi, terutama melalui penambangan data dan analisis prediktif, tetapi juga menggarisbawahi risiko serius yang terkait dengan perlindungan data privasi dan ambiguitas peraturan. Artikel ini juga menyerukan standar hukum yang lebih jelas, dan tepat terkait kerja sama internasional, dan kerangka etika untuk memandu penerapan Big Data dalam proses pidana. Selain itu, diperlukan penguatan mekanisme akuntabilitas institusional dan pengawasan yudisial guna mengawasi, mencegah, dan mengevaluasi penyalahgunaan, menjamin transparansi, serta memastikan bahwa inovasi digital tetap selaras dengan prinsip-prinsip negara hukum.

Kata kunci: Aliran informasi, Big Data, Informasi, investigasi, Pelanggaran pidana, Peradilan pidana.

Received: April 25, 2025; Revised: June 06, 2025; Accepted: June 13, 2025

INTRODUCTION

The term "Big Data" typically refers to datasets characterized by high volume (amount of data), velocity (speed of data generation), and variety (types of data). More recent models include veracity (data quality) and value (data utility) as critical dimensions of the Big Data paradigm (Laney, 2001; Dakes, 2022). Rise of "Big Data" technologies mark a paradigm shift in how modern societies collect, store, and analyse information - a shift that is increasingly reshaping criminal investigations. Law enforcement agencies now rely on data-driven methods to manage vast information flows, detect criminal patterns, and anticipate future threats. However, this digital transformation raises critical challenges at the intersection of technology, law, and ethics. In particular, the potential benefits of Big Data must be weighed against the risk of violating privacy rights, breaching legal safeguards, and undermining public trust.

In many jurisdictions, including Ukraine, the legal and institutional framework remains ill-equipped to regulate the collection and processing of large-scale personal data in the context of criminal justice. Although public authorities are beginning to embrace digital tools, the absence of consistent legal standards and operational guidelines has resulted in fragmented and often arbitrary practices. This study addresses these issues by investigating the technological, legal, and organisational dimensions of using Big Data in criminal investigations, aiming to enhance investigative efficiency while protecting fundamental rights. The use of modern information technologies, in particular Big Data, to collect and evaluate huge amounts of data on criminal law regulation is one of the approaches to modernising communication around criminal law. However, national perspectives on the application of this methodology are limited by organisational, legal and technological factors, as well as by the reluctance of legislative and law enforcement levels to recognise the results of scientific research.

To overcome these barriers, communication strategies for the development of criminal justice regulation should consider organisational and technological factors in addition to the willingness of the legislative and law enforcement environment to use new data and technologies. Arguments about Big Data and other new data technologies can also be added to the academic debate in this area. Due to its recent emergence and rapid development, the concept of Big Data is both extremely relevant and, at the same time, poorly understood. People are exposed to a daily number of new sources of information, social media platforms and online applications that require personal information in one way or another. The news often contains stories about technological advances that will improve some aspect of life. The phenomenon of Big Data in Ukraine is poorly regulated both legally and otherwise. Even if public authorities make extensive use of Big Data, the law and society as a whole should standardise all public information about this phenomenon.

Several scholars have conducted research in this and related areas. Dunaeva (2023) investigated the introduction of new technologies in cybercrime investigations. Despite challenges such as privacy and data protection, these tools are essential for collecting and analysing digital evidence. They facilitate decision-making in court proceedings and improve the efficiency of investigations while ensuring data security and compliance with legal requirements. Belova and Belov (2023) analyse the implementation of artificial intelligence in pre-trial criminal investigations based on international experience, highlighting the benefits and challenges of this process. The aim is to explore the potential and limitations of artificial intelligence in this area and to promote further development.

Severyn et al. (2023) investigated using information and telecommunication tools in criminal analysis during martial law. They analysed the possibilities of these tools to improve the efficiency of law enforcement agencies, highlighted ethical aspects and proposed strategies for their optimal implementation. It was investigated the use of new technologies in the investigation of crimes using theory, practice and regulations, in particular, the materials of the National Police of Ukraine. This is a useful study for law enforcement officers, teachers, students and anyone interested in forensics and investigative activities. Demidov (2020) investigated the impact of Big Data technology on modern politics. The hypotheses confirmed that these technologies are effective and spreading, which may change the approach to data processing in political processes (Blahuta & Movchan, 2020)

The purpose of the study is to examine the use of Big Data technology in criminal investigations about the balance between privacy and efficiency, in particular, to identify technological, organisational and legal aspects that affect the confidentiality of personal data when using this methodology and to achieve maximum efficiency in the investigation of crimes.

RESEARCH METHOD

The methodological basis of the study is modern general and special methods of scientific knowledge. Their application is based on a systematic approach, which makes it possible to study problems in the unity of their social content and legal form. The basis for the development of the conceptual apparatus within the framework of scientific research is the philosophical formal and logical approach. This strategy helps to accurately define and organise the concepts related to the research topic. This ensures that the concepts, their properties and interactions are defined in a way that makes sense. This method allows for a deeper analysis of the essence of the research topic and helps to avoid misconceptions and ambiguities. In general, the systematic and logical development of conceptual apparatus within a particular scientific topic is based on a philosophical formal logical process.

Big Data has been studied using a variety of logical methods, including analysis, synthesis, analogy, induction, deduction, generalisation and modelling. These methods facilitated an understanding of the basic ideas underlying Big Data, identifying usage patterns and trends, and creating models for analysing and forecasting future data growth. The author identifies important areas for enhancing the use of Big Data in the investigation of criminal offences using the systemic and structural methods. The phenomenon of Big Data is analysed using a historical approach, which helps to understand the impact of this phenomenon on modern culture, as well as the evolution of data processing, use and nature over time.

This study employs three core research methods to investigate the intersection of Big Data and criminal investigations: the philosophical-formal logical method, the systemic-structural method, and the normative legal method. The philosophical-formal logical method is used to define and refine key interdisciplinary concepts such as “Big Data,” “data mining,” and “criminal investigation,” ensuring conceptual clarity and consistency throughout the analysis. The systemic-structural method enables the examination of how Big Data technologies function within broader institutional and legal systems, helping to identify the interdependencies between technological applications and procedural safeguards in law enforcement. The normative legal method is applied to analyse international and national legal instruments - such as the General Data Protection Regulation (GDPR), the International Covenant on Civil and Political Rights (ICCPR), and relevant criminal procedure codes - to evaluate whether existing legal frameworks adequately address the challenges posed by Big Data use in criminal proceedings. Together,

 ISSN 2715-9698 (online)

these methods provide a comprehensive analytical foundation, allowing the study to assess both the potential and the limitations of integrating Big Data into contemporary criminal justice systems.

The potential of Big Data to change the course of criminal investigations is explored using a logical and normative method, with a particular focus on legal issues of data protection and privacy. In the context of the use of Big Data in criminal investigations, this method helps to determine what legal frameworks and standards should be put in place to protect personal data and preserve confidentiality. It helps to build strategies that effectively protect data security and privacy in the judicial sector by analysing logical and regulatory factors.

RESULTS

Big Data as it exists today is not something new or exclusive to humanity. Data analysis has been used by humans for centuries to support governments, wage wars, and make choices. For example, to determine the best way to allocate their forces, Roman government officials scrutinised their military data. All the established norms of human life changed dramatically with the advent of the Internet in the late 20th century. The evolution of Big Data has been significantly influenced by the slow but steady development of technology. Big Data is now easier to share thanks to the speed, accessibility and openness of the Internet. Previously, information was a very scarce resource that could only be accessed by a select few businesses or government officials. The volume of information coming in these days makes data especially important. The proliferation of Big Data has simplified and increased the mobility of people, created new opportunities, and reduced the time and effort required for various tasks (Hancock & Khoshgoftaar, 2020).

Since the 1990s, governments have been actively using stored and codified data for their "practical, legal and administrative purposes". By the beginning of the 21st century, websites, online portals, social networks and blogs had become so widespread that public authorities were able to freely use open access information for decision-making. The concept of big data has not been properly regulated not only in Ukraine but also around the world. Unfortunately, as of today, scientists have not found a consensus on the definition of this term, despite the large number of proposed definitions. The list of key features of big data, recognised by the scientific community, helps to understand the essence and value of this phenomenon and to formulate one's own approach to its content (Jerome, 2020).

With regard to the legislative regulation of Big Data, the result is again not positive, as Big Data does not have a clear legal definition, which creates a certain field for personal interpretation and potential risks of manipulation. Questions about the definition, main features and place of Big Data in modern life are extremely important for further understanding of its impact on public administration. Big Data is used not only by large corporations and private companies, but also by public authorities. Thanks to the development of technology, it is now possible to obtain, record, archive, measure and otherwise use a significant amount of information that previously could not be perceived by a person due to the limitations of his or her perceptual system; could not be properly obtained, recorded, concentrated and stored for an indefinite purpose; was available but remained out of the person's attention due to lack of interest or a clearly defined purpose of processing, etc (Naeem et al., 2022).

Sources of Big Data include devices and sensors of industrial and other facilities, production lines, infrastructure systems (energy, communication, trade, transport, security, etc.), social

networks, mobile communications, any household monitoring and control sensors, video surveillance and medical examination equipment, databases of enterprises, institutions and organisations, etc. Big Data can also include any type of information, including its excerpts, GPS coordinates, search engine queries and responses, the length of time a certain content is viewed, reactions to it (likes, dislikes, approvals, objections, comments) in an open or anonymous mode, etc.

Due to the diversity and networked nature of sources in highly technologically advanced societies, information is monitored, recorded, generated, processed and stored on an unprecedented scale that humanity has never experienced before. It is particularly telling that information is being accumulated in large volumes not because it is ordered or monitored by anyone, but because of the availability of modern technology, as in many cases it is easier and less expensive to record and store all incoming information (for example, from a CCTV camera or a sensor in a smart home) than to filter and distribute it according to certain criteria. On the other hand, the volume and speed of generating and receiving certain data necessitates storing it in its original raw (unordered, fast, inaccurate, unstructured) form as a better alternative to losing it. Thus, according to the ideology of Big Data, all information that is received or generated and entered the processing system in one way or another should remain there and be stored just in case for as long as possible for an indefinite future.

This approach to the organisation and implementation of information activities is fundamentally different from all previous ones, just as if every drop of drinking water suddenly began to be used by humanity to the maximum extent possible without any losses due to the emergence of an advanced technology called Big Water (Wang et al., 2020). Along with its unprecedentedly large volume, another extraordinary feature of Big Data is the ability to reuse information that was previously collected and stored for another purpose but continues to contain some unobvious value. The discovery of a new aspect of this value of previously obtained and stored information expands the horizons for efficient data processing, but also gives rise to new legal challenges.

Information activity under the slogan "lose nothing, use all types of data and as many times as it is deemed effective today or in the future" opens a new era of initial and repeated value of any information, exceptional transparency and de-anonymisation, and replacement of the basis for making legally significant decisions from establishing causation to clarifying correlations. Therefore, as an attempt to further develop the conceptual apparatus, there are grounds to propose considering Big Data as a complex phenomenon that includes: 1) an extremely large amount of information that exceeds all previously available for obtaining, storing and processing; 2) the ability to use the full amount of information, instead of a separate part (sample); 3) the ability to identify and use the re-value of any information that is obtained and stored, including just in case; 4) new grounds for drawing conclusions, in particular, the transition from establishing causality to identifying correlations (Emmert-Streib et al., 2020).

As with any achievement of scientific and technological progress, the Big Data phenomenon is neither positive nor negative in itself, but is a logical result of human activity to understand the world around us. Meanwhile, it: 1) it really exists and is gaining momentum, so it would be imprudent to make a different view; 2) it has the ability to have a profound, large-scale and unpredictable impact on most aspects of human life and activity, including the areas of economics, politics, culture and law; 3) law may undergo fundamental changes, which will result in a revision

of many concepts and approaches to solving existing and future problems (Emmert-Streib et al., 2020). There is a gradual transition from the times of "small data" to the newest era of Big Data. The old rules that were in place before and proved their effectiveness should be critically reviewed and changed for the benefit of society, taking into account current realities.

Despite some differences of opinion about the primacy of Doug Laney in the use of the term "Big Data", there is no doubt about the characteristic features of Big Data proposed by him. Accordingly, it was Doug Laney who identified the following three commonly used features of big data: volume, velocity and variety, the so-called three Vs (from English: volume, velocity, variety). As for volume, it is clear that the array of big data is infinite. Every day, various public entities and private organisations can collect information from a variety of sources, including social media, mobile applications, online business transactions, weblogs, etc. A single click on any website already transmits information about the person who made it. This characteristic feature of Big Data helps to collect, structure and analyse it, as well as to separate it from other information. An important component of this feature is that people "give" their data to the so-called system on their own. Velocity - the speed at which data is distributed around the world. This characteristic is inextricably linked to volume, as the speed of data dissemination contributes to its growth (Bulgakova, 2022).

The diversity of big data lies in the variation of its sources. These include text, audio, video messages, social media posts, financial transactions, participation in court proceedings, and filling out property declarations. In addition to the fundamental three characteristics of Big Data, some scientists and researchers of the Big Data concept add value and accuracy to this list. Thus, another approach to defining the characteristics of big data is the so-called five Vs. Value, according to Jean-Pierre Dakes, reflects the "economic importance of data", because the selection of valuable and important information from a large amount of information is one of the stages of data processing. Accordingly, the quality of the information depends on the benefits it can bring to the state, corporations and society as a whole. Veracity is an indicator of the reliability of the data obtained. The trust of people who use information for work, personal and social needs must be high. Given the amount and variety of data, it is especially important to pay attention to its veracity in order to achieve the desired processing purpose.

Guided by the above definitions and permanent features of Big Data, it is possible to distinguish such characteristics of this phenomenon as variability and easy accessibility, because the sources of information, its relevance and novelty are rapidly changing, and obtaining the desired and necessary data at any time and in any city is not a problem today (Dunaeva, 2023). The definition of the range of crimes is complicated by the heterogeneity and number of their types that can be considered in this context. Big Data and Data Mining cover a wide range, from simple theft to international criminal activity. At the same time, information about suspects can be obtained and stored in different countries and cover significant periods of time. It follows that solving such crimes usually requires cooperation with foreign countries and coordination of international organisations, such as Europol or Interpol. Conceptual documents contain only an approximate list of such crimes, for example, in Annex 1 to Regulation (EU) 2018/1727 of the European Parliament and of the Council. It is believed that the relevant criminal offences can be classified according to the following criteria: a) territorial affiliation; b) nature of the act; c) substantive content. At the same time, the classification groups do not exclude each other, but describe actions in different aspects. The main common feature of all acts can be considered: a) the complexity of their

investigation, which necessitates the use of the latest technologies; b) the danger, which is why they are classified as serious (grave) crimes punishable by imprisonment (Završnik, 2020).

Big Data and Data Mining are methods used to predict crimes and to conduct criminal investigations. Given that crime is an offence, these two activities are closely related. In this context, crime forecasting should be viewed as a rational process for the purpose of identifying and further investigating certain criminal offences. At the level of a particular crime, the patterns identified in the course of crime analysis are also important. Estimating the probability of non-obvious connections between pairs of objects can help to realise one of the main goals of researching complex social systems: predicting connections. This can be a useful method for identifying hidden connections in criminal networks and organised crime syndicates (Belova & Belov, 2023).

The need to use Big Data and Data Mining to predict and investigate crimes is acute because of the large number of data sources that contain a full range of structured and unstructured data. These sources can be permission-only or publicly available. They can be owned by the state, private companies, government agencies, law enforcement agencies or individuals. These are the raw materials that law enforcement organisations process through data mining to create multiple databases or databanks that are actively used in criminal investigations. This is of particular importance for global requests. The INTERPOL General Secretariat maintains numerous databases that include information on wanted persons, stolen vehicles and documents, lost art and valuables, as well as DNA and fingerprint data to help identify criminals and investigate crimes. Databases containing pornographic photographs, including those of children, and information on credit card theft have also been provided (Tymchyshyn et al., 2022).

In terms of technology, Big Data is processed using computer tools and data analysis methods. The analyst chooses a specific set of methods based on the nature of the task and the specifics of the criminal offence under investigation. The most commonly used data mining method for analysing offences is classification, but the most commonly used methods in criminal procedures are a) pattern recognition; b) cluster analysis (clustering); c) associative analysis; d) classification; e) social media analysis. In turn, the main technologies used to implement data analysis methods in practice are visualisation and machine learning.

Visualisation helps to collect data at the beginning of a project and is used to identify dependencies, common patterns and exceptions. Then, the project that is currently running is checked for dependencies using machine learning. The results of the use of data mining methods and technologies in criminal cases must be subject to certain limitations. They are determined by the tasks of various components of the expert's and investigator's work. Thanks to data mining, results can be obtained with the least amount of user involvement. This helps practitioners and analysts make important decisions about how to investigate crimes (Severyn et al., 2023).

In particular, it is about objectifying the process of putting forward versions, choosing investigation tactics, etc. Information exchange between law enforcement agencies is an important component of Big Data and Data Mining. The nature of the crimes that Big Data and data mining are used to investigate requires that the results of their application allow law enforcement agencies in several countries to communicate with each other. In general, the main result should be evidence in criminal proceedings that the court recognises as reliable, admissible and adequate to resolve the case on the merits. Human rights concerns are particularly important in this scenario. The use of forensic innovations in criminal proceedings determines the discourse on ensuring human rights

and freedoms both in relation to participants in criminal proceedings and to persons whose interests may be affected by the investigation (Brayne & Christin, 2021).

Protection of personal information and confidentiality are two of the weakest areas. The usefulness of data mining as a forensic technique depends to a large extent on the availability of data. Nevertheless, there are problems when it comes to maintaining the confidentiality of information. At the EU level, regulations specifically address this issue. It is noted that the processing of personal data must be carried out in a manner that is fair, lawful and transparent in relation to the data subject; the data must also be relevant and limited to the purposes for which it is collected; finally, the data must be stored in a manner that guarantees its security, including protection against unauthorised or unlawful processing (Blahuta & Movchan, 2020). When processing personal data in criminal proceedings, controversial issues may arise regarding disproportionate interference with the privacy of vulnerable groups of people (children, the elderly, people in need of international protection, etc.). In addressing such issues, human integrity and dignity should be fully respected. The issue of freedom of speech and expression on the Internet is raised through the use of this data mining technique to analyse social media (Holovkin et al., 2023a).

These features have allowed us to pinpoint key opportunities to improve the use of data mining and Big Data in criminal investigations, including crime prediction. It is believed that these are related to the standardisation of practices for their use. These standards should be based on the following three ideas: procedural, organisational and ethical. In light of the above, it is appropriate to discuss the creation of standard operating procedures for the use of data mining and ig Data in criminal investigations. Organisational, procedural and ethical considerations should form the basis of these practices. It is advisable to set out the relevant framework procedures in practical recommendations for authorised law enforcement officials, noting that violation of their principles entails liability. This will allow for the active application of Big Data and Data Mining in criminal proceedings and the use of the results for the needs of national and international justice (Zhylka, 2021).

It is possible to agree with the general consensus that the development of criminal behaviour has led to the use of advanced technologies for both committing crimes and evading punishment. In this sense, the automated recognition of patterns and relationships in massive criminal data sets has been made possible by Big Data and Data Mining, which seek to identify accurate, simple, relevant and understandable patterns and models. It is reasonable to say that these techniques have a major impact on how decisions are made to prevent and combat crime.

Only by combining data from multiple sources of information can data mining be much more successful. As a result, a strategy that combines data mining techniques for both structured and unstructured data is inspiring. This research has shown that there are real challenges associated with the use of Big Data and data mining (Kishinska & Neklesa, 2022).

Nevertheless, one cannot agree with the dominance of technological features, such as the creation of clusters (regions) based on the average risk of becoming a victim of specific illegal acts or the study of data sources with inadequate information on specific categories of criminal behaviour. Because some crimes are so unusual, it may not be possible to use the latest technology to solve them because there is insufficient information to analyse. In certain situations, it is impossible to resolve the legal side of the issue.

The lack of attention in the literature to the micro-level or specific criminal offence is another worrying feature. Despite the fact that criminals are constantly improving their illegal operations, they continue to invent new ways to break the law and avoid social control. Thus, technology cannot predict crime. It merely serves to discriminate against disadvantaged groups, undermine public trust in law enforcement and increase the likelihood of crime (Podilchak et al., 2023).

These opinions tend to be exaggerated and negate the scientific methodologies used to investigate criminal offences. On the contrary, the debate on the human rights implications of using Big Data and data mining for criminal investigations is more grounded in reality. Huge amounts of data can now be secretly collected and analysed. People with criminal records, who are considered to be at higher risk of committing crimes, are severely punished for using opaque data processing methods. It is therefore crucial to recognise that these organisational and ethical issues need to be addressed before new technologies can be widely adopted and integrated into all criminal justice processes. These factors can generally serve as a basis for the organisational, legal and procedural aspects of using Big Data and data mining for criminal investigations (Holovkin et al., 2023b).

A jurisdictional example that illustrates both the potential and the risks of Big Data use in criminal investigations is predictive policing in the United States. Several police departments, such as the Los Angeles Police Department (LAPD), have used predictive algorithms to forecast high-crime areas and possible offenders based on historical crime data. While such systems increase operational efficiency, critics highlight that they often rely on biased historical data, leading to over-policing of minority communities. The empirical case of LAPD's now-defunct "Operation LASER" program exemplifies how algorithmic tools, when used without transparency and oversight, can undermine public trust and civil liberties (Bureau of Justice Assistance, 2025).

DISCUSSION

The use of Big Data has the potential to revolutionise criminal investigations by providing law enforcement organisations with new tools to identify, profile and predict criminal activity. On the other hand, the privacy and security of personal data is also at risk. Big Data analytics can help police better fight crime and protect the public. Police can detect, investigate, and prosecute crimes by analysing huge amounts of data from social media, phone records, video surveillance, and other sources. Police can predict the locations and dates of future crimes by analysing past data and taking into account a range of circumstances. This allows them to allocate resources more efficiently and take preventive measures. Police can develop comprehensive profiles to help identify and investigate criminal activity by analysing the characteristics of crimes and criminals. Police can detect terrorist or cyber-attack plots and suspicious behaviour by analysing vast amounts of data from social media and the Internet. Based on the data analysis, police can create and implement preventive actions and programmes aimed at reducing the number of crimes in certain locations. Big Data analytics in general allows law enforcement agencies to fight crime more successfully, maintaining public safety and order (Demidov, 2020).

At the international level, the GDPR of the European Union provides a detailed framework for personal data processing, emphasizing lawfulness, transparency, and accountability. Article 22 of the GDPR, which limits automated decision-making, is particularly relevant when applying predictive analytics in criminal investigations. Similarly, Article 17 of the International Covenant on Civil and Political Rights (ICCPR) protects individuals from arbitrary or unlawful interference

with privacy. National constitutions and criminal procedure codes must be interpreted in light of these instruments to ensure that Big Data use does not override fundamental rights. For example, in Germany, constitutional jurisprudence strictly limits data surveillance practices that lack sufficient statutory basis or proportionality. These legal safeguards should inform the design and implementation of Big Data systems in law enforcement.

However, there are many legal hurdles to overcome before Big Data can be used in criminal investigations. Law enforcement organisations should only process information that is necessary to fulfil the defined purpose of the investigation. This implies a minimum amount of personal data that is collected, stored and used to fulfil law enforcement tasks. Modern technologies and security standards should be used by law enforcement organisations to protect against unauthorised access, loss or destruction of personal data. Information about what data is collected, how it will be used, and the rights of individuals whose data is processed should be made public by law enforcement agencies.

Any kind of profiling or discrimination based on a person's ethnicity, gender, religion or political views should be prohibited by law enforcement agencies. Law enforcement should ensure that any processing of personal information complies with human rights laws and privacy standards. The processing of personal data by law enforcement agencies should be subject to an effective system of control and oversight, which should include an independent entity responsible for this process (Mulyk, 2024).

As a result of the use of machine learning algorithms based on statistical data, people can be profiled based on attributes such as skin colour, gender, age, ethnicity, etc., which can lead to discrimination. Inaccuracies in the data used to train machine learning algorithms can lead to systematic errors and inaccurate profiling of people with certain traits. Ethical risks related to algorithmic bias are not merely theoretical concerns. Machine learning tools trained on historical crime data often replicate and amplify existing societal biases, particularly against racial or socio-economic groups. This raises questions not only about fairness, but also about the admissibility of such data as evidence in criminal trials. For example, risk assessment tools used in pre-trial detention decisions have been shown to disproportionately classify minority defendants as high risk, despite lack of individualised evidence. These practices may inadvertently contribute to discriminatory policing, undermine due process, and conflict with principles of presumption of innocence. Therefore, it is essential to incorporate ethical review boards and bias-mitigation protocols at all stages of algorithm development and deployment in criminal justice.

It is crucial to thoroughly examine and validate the effectiveness of machine learning algorithms before applying them to law enforcement. In addition, it is crucial to identify and eliminate any cases of discrimination or profiling. It is important to recognise that machine learning algorithms may not always be able to accurately understand the details and context of each scenario, which requires further research and validation of the findings. When using machine learning algorithms, law enforcement agencies should be transparent about their methods and be prepared to take responsibility for any adverse consequences that may arise, such as invasion of privacy and human rights. Researching and developing machine learning algorithms that guarantee fairness and non-discrimination, taking into account diversity and inclusion, is essential (Shevchuk, 2023).

The conditions and limitations on the collection and use of personal data in criminal investigations should be clearly defined in laws and regulations. This includes specifications on

proportionality, lawfulness and explanation of data processing. Laws should ensure the protection of individuals' personal information, including their right to restrict access to it and ensure its secure use and storage. Openness and transparency in the use of Big Data in criminal investigations should be guaranteed by legal oversight. This means that law enforcement agencies are obliged to inform the public about the data they collect, how they plan to use it, and the security measures they take (Završnik, 2021).

Independent systems for monitoring and controlling the use of Big Data by law enforcement agencies are necessary for effective legal control. These can be commissions, special committees or other organisations authorised to review potential violations and conduct inspections. Legislation that contains procedures for legal redress is crucial in case of violations of citizens' rights when using big data for criminal investigations. In addition, accountable persons and governance systems need to be recognised. The legal environment should be actively developed to meet the current technologies and challenges in the use of Big Data in criminal investigations. This may include the development of new laws and standards, as well as the adaptation of existing legislation to new needs. In summary, the use of Big Data has the potential to significantly improve the effectiveness of criminal investigations; however, this will require a careful balance between the effectiveness of law enforcement tactics and the preservation of private and individual rights (Holovkin et al., 2021).

The strategic implementation of Big Data in criminal justice systems should follow three guiding principles: legality, proportionality, and accountability. First, legal frameworks must define the permissible scope and limits of data collection and algorithmic processing in criminal investigations. Second, the use of Big Data must be proportionate to the severity and urgency of the crime being investigated, avoiding excessive surveillance of low-risk populations. Finally, there must be robust accountability mechanisms, including independent audits, impact assessments, and public transparency requirements, to ensure compliance and protect against misuse. These principles should form the core of any national or international policy initiative aimed at governing Big Data in law enforcement.

CONCLUSION

Although data analytics has been around for centuries, Big Data has become an important tool in the modern world thanks to the Internet and technological advances. The ease and speed with which data can now be collected and accessed via the Internet has greatly increased the importance and ubiquity of huge amounts of information. In addition to making life easier and speeding up procedures, Big Data offers tremendous opportunities for several industries, including government and industry.

Combining data from different information sources significantly increases the effectiveness of data mining. A strategy that combines data mining techniques for both structured and unstructured data is particularly promising. However, there are real challenges associated with the use of Big Data and data mining, such as poor data quality and difficulties in analysing specific forms of criminal activity. The micro-level or specific criminal behaviour requires more attention, and organisational and ethical issues need to be addressed. Taking these factors into account can help make the use of Big Data and data mining in criminal justice more practical and effective.

All indications point to the possibility that the use of Big Data in criminal investigations could fundamentally change the way law enforcement agencies operate, providing officers with new resources

for profiling, identifying and predicting criminal activity. Vast amounts of data from a variety of sources can improve the ability of police to detect, investigate, and prevent crime. These prospects, however, raise issues of secrecy and privacy. To maintain a balance between the protection of individual rights and freedoms and the effectiveness of law enforcement, clear legal norms and control systems are needed. The use of machine learning algorithms requires special attention when making a decision to avoid bias and stop profiling people based on illegal characteristics. The successful integration of Big Data into the criminal justice system depends on the creation of a modern legal framework that takes into account the requirements and obstacles of the digital era.

Future research should focus on comparative empirical studies assessing how different jurisdictions integrate Big Data into their investigative processes while upholding legal safeguards. There is also a growing need to evaluate the long-term societal impacts of predictive policing, the effectiveness of legal oversight bodies, and the development of algorithmic transparency standards. These efforts would contribute to building a more equitable and accountable digital criminal justice system.

REFERENCES

Belova, M. V. & Belov, D. M. (2023). Implementation of artificial intelligence in pretrial investigation of criminal cases: international experience. Analytical and Comparative Jurisprudence, (2), 448-453.

Blahuta, R. I. & Movchan, A. V. (2020). The latest technologies in the investigation of crimes: The current state and problems of use. Lviv: Lviv National University of Internal Affairs.

Brayne, S. & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. Social Problems, 68(3), 608-624.

Bulgakova, D. A. (2022). Protection of commodified data on digital platforms. Analytical and Comparative Jurisprudence, (1), 208-212.

Bureau of Justice Assistance. (2015). Los Angeles Police Department – Operation LASER: Leveraging data for smarter policing. Retrieved from https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/losangelesspi.pdf

Demidov, D. Yu. (2020). Use of Big Data technology in political research. Zaporizhzhia: Zaporizhzhia National University.

Dunaeva, T. (2023). Certain aspects of the use of technologies in the investigation of cybercrimes. Current Issues in Modern Science, 12(18), 502-512.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. Frontiers in Artificial Intelligence, 3, 4.

Hancock, J. T. & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. Journal of Big Data, 7(1), 94.

Holovkin, B. M, Cherniavskyi, S., & Tavolzhanskyi, O. (2023b). Factors of Cybercrime in Ukraine. Relacoes Internacionais no Mundo Atual, 3(41), 464-488. https://doi.org/10.21902/Revrima.v3i41.6401

Holovkin, B. M., Semenyshyn, M., Tavolzhanskyi, O. V., Lysodyed, O. V. & Smetanina, N. V. (2023a). Fight against Corruption-Related Crimes in Wartime in Ukraine. International Annals of Criminology, 61(3-4), 384-409. https://doi.org/10.1017/cri.2023.31

Holovkin, B. M., Tavolzhanskyi, O. V., Tavolzhanskyi O.V., & Lysodyed, O. V. (2021). Connections, 20(2), 75-87. https://doi.org/10.11610/Connections.20.2.07

Jerome, B. (2020). Criminal investigation and criminal intelligence: Example of adaptation in the prevention and repression of cybercrime. Risks, 8(3), 99. https://doi.org/10.3390/risks8030099

Kishinska, I. O. & Neklesa, O. V. (2022). Analytical complexes in criminal analysis. In S. S., Cherniavskyi, D. I., Ovsianiuk, & V. V., Korolchuk (Eds.). Current Issues and Prospects for the Development of Criminal Analysis in the Law Enforcement System of Ukraine (pp. 243-245). Kyiv: National Academy of Internal Affairs.

Mulyk, K. (2024). Improving efficiency in the investigation of crimes in Ukraine and foreign countries, their innovations and measures. New Ukrainian Law, 6, 33-36. https://doi.org/10.51989/NUL.2023.6.5

Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., De-la-Hoz-Franco, E., & De-La-Hoz-Valdiris, E. (2022). Trends and future perspective challenges in big data. In Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications (15–18 October 2019, Arad) (pp. 309-325). Singapore: Springer.

Podilchak, O., Dunaeva, T., & Kulyk, K. (2023). Information technologies for the provision of criminological policy. Current Issues in Modern Science, 11(17), 663-678.

Severyn, O., Shaposhnyk, M., & Kiselyov, A. (2023). The use of information and telecommunication tools in criminal analysis during martial law: features and possibilities. UNIVERSUM, (5), 102-106.

Shevchuk, V. (2023). Possibilities and prospects of using artificial intelligence technologies in the investigation of war crimes. In The use of artificial intelligence technologies in combating crime (pp. 86-94). Kharkiv: Pravo.

Tymchyshyn, A., Semeniaka, A., Bondar, S., Akhtyrska, N., & Kostiuchenko, O. (2022). The use of big data and data mining in the investigation of criminal offenses. Revista Amazonia Investiga, 11(56), 278-290.

Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big data service architecture: a survey. Journal of Internet Technology, 21(2), 393-405.

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. ERA Forum, 20(4), 567-583.

Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. European Journal of Criminology, 18(5), 623-642.

Zhylka, M.P. (2021). Protection of labour rights of employees by courts. In M.P., Zhylka & V.V. Zhernakov (Eds.). Human Security and Implementation of the Right to Work in Modern Living Conditions (pp. 124-130). Kharkiv: Yaroslav Mudryi National Law University.