FLASH REPORT 20

AI AND IMPLICATIONS ON VICTIMS

Main Authors

Dorothea Tsatsou (CERTH)

Sylvie Dias (PJ)

Isabela Maria Rosal (KU Leuven)

April 2026

ENACT.
European Network Against Crime and Terrorism

03

About ENACT

ENACT is a knowledge network focused on the fight against crime and terrorism (FCT). The network is funded under the Horizon Europe Framework Programme in Cluster 3 – Civil Security for Society. The project addresses the call topic HORIZON-CL3-2022-SSRI-01-02 'Knowledge Networks for Security Research & Innovation', aiming to collect, aggregate, process, disseminate and make the most of the existing knowledge in the FCT area.

The project aims to satisfy two major ambitions,

- Provide evidence-based support to the decision-makers in the EU research and innovation (R&I) ecosystem in the FCT domain, targeted explicitly at enabling more effective and efficient programming of EU-funded R&I for the fight against crime and terrorism.

- Act as a catalyst for the uptake of innovation by enhancing the visibility and reliability of innovative FCT security solutions.

Report Feedback

We're collecting feedback on this report through the EU Survey Platform, if you'd like to share your thoughts anonymously please use the following link. https://ec.europa.eu/eusurvey/runner/enact-report-feedback

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Misuse awareness: This report does not intend any type of support or endorsement of use of the technologies described herein in malicious activities. All scenarios described in this report are abstract, hypothetical, non-operational and not intended to describe exploitable vulnerabilities.

Copyright

Funded by
the European Union

04

Scene Setter

Artificial intelligence (AI) is reshaping contemporary societies by automating decision-making, enhancing predictive capabilities and scaling services across sectors such as healthcare, finance, law enforcement and social services. In particular, the growing use of AI-driven predictive policing systems, designed to forecast crime risks and allocate policing resources, illustrates both the promise and risks of AI in security and justice contexts. While such systems are often justified as tools for efficiency and prevention, recent research shows that they can amplify existing social inequalities due to biased or incomplete data, including factors such as race, socio-economic status and prior policing practices. Moreover, predictive models may overestimate risks or misclassify individuals and communities, leading to disproportionate surveillance and intervention. Consequently, these systems raise concerns about discrimination, lack of transparency and forms of anticipatory harm, where individuals or groups may be targeted not for actions taken but based on algorithmic predictions. As a result, they exemplify how AI can simultaneously support and undermine victim protection. This shifts the nature of victimisation from reactive to predictive and pre-emptive, challenging fundamental principles such as the presumption of innocence.

Nowadays, the real gap is not just what AI can do, but what victims actually need versus what systems provide: current versus emerging victim needs. The biggest shift is that victims are no longer only harmed by individuals or institutions, but by complex socio-technical systems where responsibility is diffuse, harm is scalable and visibility is low. Classifying victims of AI involves identifying who is harmed, the nature of the harm and the method of victimisation. Based on current AI incident taxonomies and real-world incidents, AI victims can be classified across several dimensions, including individuals, groups and society at large. This report aims to reflect on and refocus the issue of AI within the context of victims, exploring the different aspects related to the relationship between victims and AI systems across multiple dimensions, effectively examining **how AI:**

- **shapes victimisation**, i.e., how harm is produced and scaled,
- **affects recognition and response**, including technology and operational aspects,
- **redistributes risks and responsibilities**, including market dynamics, and
- **challenges accountability and rights**, in relation to European legal and ethical frameworks.

The report will present different examples to illustrate some of the novel scenarios that surfaced with emerging AI tools and systems, exploring some of the contexts where AI may affect victims. Although AI enhances capabilities to support victims, it is plausible that technological limitations, market incentives and legal gaps may undermine effective protection and redress, requiring a shift toward victim-centred AI governance.

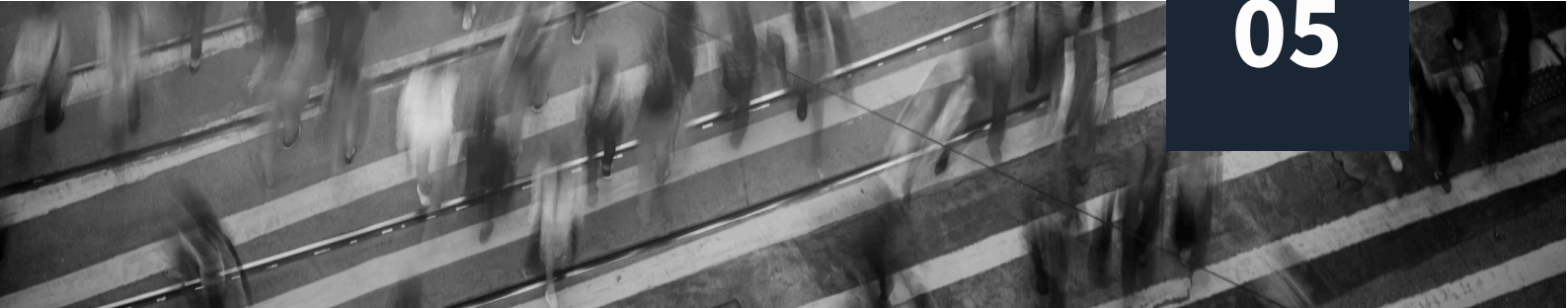
Releated References

- Almasoud, A. S., & Idowu, J. A. (2025). Algorithmic fairness in predictive policing. AI and Ethics, 5(3), 2323-2337.
- Kondapalli, P., Singh, P., Malik, A., & Lesmana, C. T. (2025). A Literature Review: Bias Detection and Mitigation in Criminal Justice. Engineering Proceedings, 107(1), 72.
- Bartoletti, I., & Xenidis, R. (2023). Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. Council of Europe.
- Hasinoff, A. A., & Schneider, N. (2022). From scalability to subsidiarity in addressing online harm. Social Media+ Society, 8(3), 20563051221126041.
- Li, S., & Heine, K. (2026). Developing a harm-based approach to understand digital vulnerability in the era of AI: A perspective of the European Union. Computer Law & Security Review, 60, 106266.
- Citron, D. K. (2022). The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age. W. W. Norton & Company.
- Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Acknowledgement

This report was requested by the International Network Supporting Victims of Terrorism and Mass Violence (INVICTIM) to support the 2026 INVICTIM Symposium, which has the theme Reimagining Victim Support in a Changing World: Learning from the Past, Succeeding in the Present, Preparing for the Future.

Created in 2016, INVICTM brings together trusted experts dedicated to improving support for victims of terrorism and mass violence. The strength of INVICTM lies in the ability to share lessons learned, leverage the knowledge and expertise of its members and their networks, to influence change and turn research into action within our own countries and globally. The group of over 30 international members includes NGOs, law enforcement agencies, civil society members and other experts that provide information based upon their background, country and professional perspective.



Victims' needs in the age of Artificial Intelligence (AI)

In crime contexts, including terrorism and mass violence, victims' needs are increasingly shaped by both physical and digital harms. The integration of AI into policing, justice and support systems does not replace these needs, but reshapes how they are addressed and often where systems fall short.

Persistent victims' needs (reinforced by AI)

Core victims' needs remain consistent but are intensified in AI-driven contexts:

- **Access to justice and redress:** AI complicates causation and accountability.
- **Transparency and explainability**: decisions are often opaque.
- **Privacy and data protection**: increased exposure through data-driven systems.
- **Fairness and non-discrimination**: risk of amplifying structural bias.
- **Human-centred support**: essential in trauma contexts but not always offered.
- **Protection of vulnerable victims**: systems often fail to account for differentiated needs.

Illustrative Harm Scenario One
Automated welfare decision systems
Automated welfare systems have wrongly denied benefits due to flawed data or rigid models. Victims, often already vulnerable, face loss of income, stigma and limited ability to challenge decisions due to a lack of transparency. In crisis contexts, such failures risk excluding victims from essential support, illustrating how automation can intensify existing vulnerabilities.

Emerging victims' needs in AI-driven environments

AI introduces new forms of harm, generating novel needs among victims and requiring new protections.

- **Content creation and dissemination:** Protection from misinformation, disinformation and narrative manipulation.
- **Harmful materials:** Protection from AI-generated content on synthetic identity fraud, abuse and sexual exploitation, particularly affecting women and minors.
- **Algorithmic accountability and contestation rights:** (New) instruments are needed to guarantee transparency and explainability of AI systems and remedies to allow the revision of automated decisions, including possible obligations regarding human participation in the review process.

- **Protection from large-scale and rapid harm:** AI-enabled harms can occur instantly, spread globally and be replicated at scale. Victims, therefore, require rapid response mechanisms, effective automated detection systems and stronger platform accountability.
- **Digital and AI literacy:** Information about digital transformation should be offered to different groups of society, including how AI systems work, what they can do and how they are applied. Literacy methods should consider the targeted audience and different methodologies and means.
- **Identity and reputation protection:** Focus on prevention rather than reparation, considering that AI-related harms can be long-lasting and may affect the identity and reputation of individuals or even entities. Specific remedies should be created and provided to mitigate these risks.
- **Protection from predictive and pre-emptive harms:** AI systems offer outputs before concrete action. New remedies should be provided for when these results create harms.

Illustrative Harm Scenario Two
AI-Generated online abuse, gender-based violence and sexual exploitation
Generative AI enables large-scale production of deepfake sexual content and synthetic abuse material, disproportionately targeting women and minors. Victims face psychological harm, reputational damage and coercion, compounded by rapid dissemination and weak enforcement. Responses usually take time and are insufficient in providing the tools and remedies needed to the victims. This reflects a shift toward scalable, digitally mediated victimisation requiring new legal and support responses.

Structural needs: toward systemic protection

AI-related harms reveal systemic gaps requiring structural reform.

- **Burden of proof on victims:** victims often must build their own case to prove any AI-related harms, even limiting the presumption of innocence. Victims should not be required to demonstrate complex AI causation.
- **Weak oversight and accountability:** lack of human participation, assessments, documentation and other instruments that allow the oversight of AI systems (e.g., regulators, auditors, watchdog bodies).
- **Lack of collective redress mechanisms:** AI systems may affect not one specific individual but rather a group, thus enabling class actions and other tools to address large-scale AI harm.

These structural gaps are particularly harmful for vulnerable victims, who often face greater barriers in accessing justice, understanding automated decisions and seeking redress.

Illustrative Harm Scenario Three
Large-scale data breach and algorithmic harm
In cases where AI systems misuse or expose personal data. Data breaches in AI systems can expose sensitive information and affect large populations simultaneously. Victims may face identity theft, long-term privacy violations and loss of trust. For victims, such breaches can lead to re-identification and re-traumatisation, highlighting the need for systemic safeguards.

From reactive to proactive protection

The central transformation is a shift from reactive victim support to proactive risk governance, preventing harms from materialising. Protection must be embedded in system design, deployment practices and regulatory frameworks. Victims are increasingly harmed by complex socio-technical systems, where responsibility is diffuse and harm is scalable.

For victims of crimes, this shift is particularly significant: protection must extend beyond immediate response to include long-term, system-level safeguards against both physical and digitally mediated harm, ensuring that technological innovation strengthens rather than undermines trust in justice and support systems.

In this sense, improving victim protection in the age of AI is not only a matter of individual rights, but also a cornerstone of democratic resilience, particularly in ensuring that vulnerable populations are not disproportionately exposed to harm or excluded from protection in increasingly digital and automated systems. Trust in institutions increasingly depends on the perceived fairness, accountability and reliability of the systems that govern security and justice.

Releated References

- Victim Support Europe (2026). Victim Support Europe Strategy 2026 – 2030.
- Lupo, G., & Pacifico, G. (2025). Addressing the Needs of Victims: the Design of a Multi-Role AI-Driven Application for Victims of Crime Access to Justice. European Journal of Law and Technology, 16(2).
- Impunity Watch (2025). Artificial Intelligence and Transitional Justice: Framing the Connections.
- United Nations Interregional Crime and Justice Research Institute (2024). Generative AI: A New Threat for Online Child Sexual Exploitation and Abuse.
- Internet Watch Foundation IWF (2024). What has changed in the AI CSAM landscape?
- National Crime Agency UK (2024). Virtual Global Taskforce (VGT) AI statement.
- AI4People (2018). AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.
- UNESCO (2021). Recommendations on the Ethics of Artificial Intelligence.
- Newman, J. (2026). "Technology-Facilitated Harm": How AI Chatbots Are Failing Abuse Survivors.
- Romero, I. B. (2026). UNICEF. Deepfake abuse is abuse. UNICEF press release.

AI technologies and harm risks

AI technologies, including machine learning, natural language processing, computer vision and generative models, introduce distinct operational characteristics that shape how harm emerges, evolves and is detected in increasingly digital and data-driven environments. From a technology perspective, the focus is not only on what these systems do, but on how their underlying properties influence patterns of victimisation and the ability to identify and respond to harmful situations.

AI systems combine scalability, automation and data-driven inference, enabling the generation and detection of harmful patterns at scale. At the same time, their opacity, reliance on data and limitations in contextual understanding constrain the reliability and completeness of outputs. Understanding these characteristics is therefore essential to assess how these systems influence the detection, interpretation and handling of harm in complex and rapidly evolving environments.

Core technological characteristics

AI systems exhibit a set of defining features that directly affect how harmful situations are produced and interpreted:

- **Black-box decision-making:** Many AI models operate with limited explainability, making it difficult to trace how specific outputs are generated. This affects the ability to understand how harmful situations are classified or prioritised.
- **Data dependency:** AI systems rely heavily on training and operational data. Incomplete, biased, or unbalanced datasets can shape how harms are recognised, potentially reinforcing blind spots or over-representing certain patterns.
- **Unpredictability, hallucinations and model variability:** AI systems, particularly generative models, can produce unexpected, fabricated or inconsistent outputs. This variability affects the reliability of identifying and interpreting harmful behaviours or content.
- **Bias and error propagation:** AI systems can embed and amplify biases present in training data, while also producing systematic or sporadic errors (related also to the previous point). These can potentially lead to misrepresentation and over-/under-detection of harmful situations, as well as to unequal treatment across different groups.
- **Autonomy and continuous operation:** Once deployed, AI systems can operate with limited human intervention, continuously analysing data streams and updating outputs, limiting sanity and reliability checks.

Technological manifestations of harm

These characteristics translate into specific ways in which harm can emerge or be amplified:

- **Pattern amplification:** AI systems can reinforce existing trends in data, leading to repeated identification of similar types of cases while overlooking less visible or emerging forms, relying on historical patterns that do not necessarily translate the current context.
- **Synthetic content generation:** Generative AI enables the production of realistic but artificial text, images, audio and video. This expands the range and scale of deceptive or abusive material that can circulate and increases the risks of not identifying what is real and what is artificially created.
- **Data manipulation and system interference:** AI systems are vulnerable to data and model manipulation (e.g., poisoning or adversarial inputs), which can distort outputs and affect reliability.
- **Cross-system propagation:** Interconnected systems allow outputs from one AI process to feed into others, increasing the reach and persistence of both accurate and erroneous classifications.

Illustrative Harm Scenario Four
Predictive policing systems
Individuals living in a neighbourhood repeatedly flagged by predictive policing systems experience increased police presence and frequent stops, despite no direct involvement in criminal activity. From their perspective, they become subject to ongoing scrutiny or even unjust discrimination, based not on actions but on patterns derived from historical data. As the same data-driven signals are continuously reinforced, residents feel unfairly targeted and stigmatised. This wrongful/harmful pattern amplification and cross-system propagation, where technological outputs reproduce and scale existing biases, results in unfair treatment, sustained exposure to surveillance and diminished trust in institutions.

Technological constraints in identifying harm

Despite their capabilities, AI systems face structural limitations:

- **Context limitations:** AI struggles to capture social, behavioural and situational context, which is often critical for interpreting complex or evolving harmful situations.
- **Signal-to-noise challenges:** Large-scale data processing can generate high volumes of outputs, making it difficult to distinguish meaningful patterns from irrelevant or misleading signals.
- **Dependence on predefined categories:** AI systems typically operate within predefined labels/classes, limiting their ability to detect novel or hybrid forms of harm.

Illustrative Harm Scenario Five
Failure to detect emerging forms of online exploitation
Victims of online exploitation find that harmful content involving them (e.g. manipulated images) remains available across platforms despite reporting efforts in the fully autonomous AI-based reporting system. From their perspective, the AI system fails to recognise the abuse, as it does not fit established categories used for automated detection. This reflects context limitations and reliance on predefined categories, where technological constraints prevent timely identification of evolving harms. The victims experience repeated harm and limited response, as the content continues to circulate without being flagged or removed.

Technological implications for harm detection and interpretation

AI in policing fundamentally reshapes how harmful situations are identified, interpreted and acted upon, by introducing new forms of mediation between data, systems and decision-making. Rather than simply supporting existing practices, AI systems influence what is viewed as harm, as well as how it is classified and prioritised for response. This creates opportunities for earlier detection and broader coverage, but also risks of misrepresentation, omission or distortion of reality.

The effectiveness of AI systems depends not only on their technical performance, but also on their ability to operate reliably across diverse and evolving contexts, where harm may be ambiguous, dynamic or difficult to categorise. As such, technological limitations, in relation to context sensitivity, data quality and interpretability, can directly affect outcomes and experiences.

These dynamics highlight the importance of embedding technical safeguards, such as validation mechanisms, transparency features and human oversight, within system design and deployment to ensure more reliable and accountable outcomes.

Releated References

- Trend Micro (2024). The Top 10 AI Security Risks Every Business Should Know.
- NIST (2024). Artificial Intelligence Risk Management Framework.
- OpenAI (2025). Disrupting malicious uses of AI: an update.
- Raman, D., et al. (2025). Intolerable risk threshold recommendations for artificial intelligence. arXiv preprint arXiv:2503.05812.
- Allaham, M., Kieslich, K., & Diakopoulos, N. (2025). Global Perspectives of AI Risks and Harms: Analyzing the Negative Impacts of AI Technologies as Prioritized by News Media. arXiv preprint arXiv:2501.14040.
- Seghid, N., Iqbal, F., Al-Room, K., & MacDermott, Á. (2026). Emerging Threats in AI: A Detailed Review of Misuses and Risks Across Modern AI Technologies. Frontiers in Communications and Networks.
- EUROPOL (2024). AI and policing - The benefits and challenges of artificial intelligence for law enforcement.
- TRM Labs (2025). The Rise of AI-Enabled Crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises.
- United Nations Interregional Crime and Justice Research Institute (2024). Generative AI: A New Threat for Online Child Sexual Exploitation and Abuse.
- Caballar, R. D. (2025). 10 AI dangers and risks and how to manage them. IBM Think.
- Bozkir, E., et al. (2025). Predictive Policing or Predictive Prejudice? A Study of the Legal, Historical and Ethical Implications of AI in Policing. OxJournal.

MARKET VIEW

12

Market and economic dynamics

The adoption and deployment of AI systems are strongly driven by market forces, including efficiency gains, cost reduction, scalability and competitive advantage. In practice, these dynamics influence not only which technologies are developed and procured, but also how risks and responsibilities are distributed. From a victim-centred perspective, the marketisation of AI introduces structural imbalances where the benefits of innovation are often internalised by providers and deploying organisations, while the risks and harms are externalised onto individuals and communities.

Core AI market dynamics affecting victims

Several economic and market characteristics shape how AI systems impact victims:

- **Commercial incentives over public-interest outcomes:** AI vendors prioritise scalable, marketable solutions that deliver measurable efficiency gains. This can result in systems optimised for performance metrics rather than for accurately capturing complex or sensitive forms of harm.
- **Information asymmetry:** Vendors typically retain greater knowledge about how systems function, including model limitations and data constraints. This creates an imbalance where those affected by AI-driven processes have limited visibility into how outcomes are produced.
- **Externalisation of risk:** The costs associated with errors, misclassification, or unintended consequences are often borne by affected individuals rather than system developers or deployers. Victims may face the burden of identifying, proving and contesting harm.
- **Procurement and deployment pressures:** Organisations may adopt AI tools based on cost-efficiency or perceived innovation value, sometimes without sufficient validation, testing, or contextual adaptation. This can lead to the use of systems that are not fully suited to the environments in which they operate. As a rule, regulatory frameworks related to procurement procedures do not provide specific provisions for these situations.
- **Standardisation and scalability of solutions:** Market-driven AI products are often designed for broad applicability across contexts. This can result in a mismatch between system design and local realities, particularly where victimisation patterns are context-specific.

These dynamics shape not only how harm occurs, but also how it is recognised and addressed:

- **Systemic gaps in accountability:** When multiple actors are involved (developers, vendors, deploying institutions), responsibility for harm becomes fragmented, making it difficult to attribute fault or seek redress.
- **Uneven protection levels:** Access to well-designed and properly validated AI systems may vary across regions and institutions, leading to inconsistent levels of protection and response.
- **Delayed correction of system failures:** Commercial pressures may discourage rapid acknowledgement or correction of system limitations, prolonging exposure to harmful outcomes.
- **Reinforcement of existing inequalities:** Market incentives can prioritise high-demand use cases, potentially neglecting less visible or less profitable forms of victimisation.

Illustrative Harm Scenario Six
Predictive policing systems and commercial AI vendors
Predictive policing systems procured from commercial vendors are often deployed based on promised efficiency gains and cost-effectiveness. In practice, individuals may be subjected to LEA decisions and interventions shaped by proprietary models that cannot be independently scrutinised or explained. When harms occur (e.g., misclassification, disproportionate targeting), affected individuals face difficulties in understanding or contesting outcomes, as access to system logic and data is restricted by commercial protections. This results in information asymmetry and externalisation of harm, where market-driven incentives prioritise scalability and product deployment.

Demand-side mitigation of the impact of AI on victims

While supply-side forces focus mainly on monetary and efficiency gains, as discussed above, demand-side can act as a counterbalancing force to try and mitigate the impact of AI on victims. In tenders, especially in innovation procurement, procurers can build in requirements that ensure systems are designed with victim protection in mind. For example, systems can be designed in conjunction with the procuring entity, ensuring that it is built for the specific purpose and avoids the pitfalls mentioned in the above sections, such as gaps in accountability, black-box designs and burden of proof requirements. Overall, the entities requesting the new AI technologies can play an important role in ensuring that the victims are not disproportionately impacted by the new AI systems being purchased and used in operational environments. However, this will require procurement design that prioritises victim protection over simple monetary and efficiency gains.

Releated References

- Perry, N, (2023). Why 2024 will be pivotal for AI in procurement. Procurement Magazine.
- OECD (2025). Artificial intelligence and competitive dynamics in downstream markets. OECD Roundtables on Competition Policy Papers, No. 331, OECD Publishing, Paris, https://doi.org/10.1787/ccf0624a-en.
- Deloitte (2026). The State of AI in the Enterprise. Deloitte's 2026 AI report tracking adoption and impact.
- European Parliament (2025). Artificial Intelligence and Civil Liability.
- National Telecommunications and Information Administration NTIA (2024). AI Accountability Policy Report.
- AON insights (2026). AI Risk 2026: What Business Leaders Need to Know.
- Research and Markets (2026). AI in Predictive Policing Market Report 2026.
- Evans, B. (2026). Responsible AI in 2026: A 3-step guide for governance that scales. OneTrust blog post.
- Allianz (2026). Allianz Risk Barometer 2026: Cyber and AI as major business risks.

Civil and criminal liability

Although ethical, legal and societal (ELS) aspects are connected to all aforementioned viewpoints – capabilities, technology and market, AI particularly reshapes responsibility, access to justice and the protection of fundamental rights, with extended implications on victims. Moreover, AI systems challenge existing legal and ethical frameworks by introducing new forms of mediation between actions, decisions and outcomes. From a victim's perspective, the key issue is not only that harm occurs, but that the pathways to recognising, attributing and remedying that harm become more complex and fragmented.

Core AI market dynamics affecting victims

AI complicates traditional approaches to civil and criminal liability by altering how causation, responsibility and evidence are established:

- **Attribution of responsibility:** AI systems are developed, deployed and operated by multiple actors, including developers, vendors and institutions. This distributed structure makes it difficult to identify who is legally responsible when harm occurs, particularly where decisions are partially automated.
- **Burden of proof:** Victims may face increased difficulty in demonstrating that harm was caused by an AI system, especially when system logic is not accessible or understandable. Establishing a clear causal link between system outputs and real-world consequences becomes more demanding.
- **Limited access to evidence:** Key information needed to challenge decisions (e.g. training data, model design, decision pathways) is often unavailable to affected individuals. This restricts the ability to contest outcomes or pursue legal remedies.

These challenges can result in delayed or denied access to justice, particularly where legal frameworks have not yet adapted to account for AI-driven processes.

Illustrative Harm Scenario Seven
AI-assisted medical diagnosis and liability gaps
An AI-assisted diagnostic system contributes to an incorrect medical assessment, leading to delayed treatment and worsening health outcomes. From the victim's perspective, it is unclear whether responsibility lies with the healthcare professional, the institution, or the system developer. Access to information about how the system reached its conclusion is limited, making it difficult to establish causation and pursue accountability. Challenges in attribution of responsibility, burden of proof and access to evidence are evident, leading existing legal frameworks to struggle to address harm involving AI-mediated decision-making.

Regulatory developments

Emerging regulatory approaches (e.g., AI Act, especially regarding high-risk systems, GPAI Code of Practice, Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law) alongside more established ones (e.g., Charter of Fundamental Rights of the European Union, European Convention on Human Rights, General Data Protection Regulation) seek to address these challenges by introducing safeguards aimed at improving accountability and reducing barriers for affected individuals. Key directions include:

- **Enhanced transparency requirements:** ensuring that AI systems provide meaningful information about how decisions are made and used, both ex-ante and ex-post.
- **Risk-based regulation:** where higher-risk systems are subject to stricter obligations, including oversight, testing and documentation.
- **Procedural safeguards:** aimed at easing the burden on individuals seeking to challenge automated decisions, including access to explanations and review mechanisms.

Despite these developments, regulatory approaches remain uneven across jurisdictions, creating fragmentation in how protections are applied and enforced. This is particularly challenging considering the cross-border nature of certain AI-systems and their effects.

Ethical considerations and human rights

AI systems raise fundamental ethical concerns that intersect with core human rights principles. These concerns are particularly salient where systems influence decisions affecting safety, dignity, or access to support.

- **Fairness and non-discrimination:** AI systems can produce outcomes that disproportionately affect certain groups, raising concerns about equality before the law and equitable treatment.
- **Transparency and accountability:** Ethical use of AI requires that decisions are explainable and that responsible actors can be identified and held accountable.
- **Respect for human dignity:** Automated processes may fail to recognise the complexity and sensitivity of situations involving harm, leading to responses that are perceived as impersonal or inadequate.
- **Presumption of innocence and fair trial:** especially predictive models may affect the ideas of presumption of innocence and fair trial, considering how certain individuals or groups can be targeted without due process.

From a human rights perspective, AI has the potential both to enhance and to undermine rights such as privacy, equality and access to justice. The impact is uneven and often more pronounced for vulnerable populations. New and already existing instruments related to the protection of human rights must be considered during the whole lifecycle of the AI systems, comprehending how the effects of these tools and systems are connected to real individuals and lives.

Illustrative Harm Scenario Eight

AI in victim support chatbots

AI-powered chatbots are used to provide immediate support and guidance to victims, improving accessibility and availability of assistance. However, when such systems provide inaccurate or incomplete advice, or fail to respond with appropriate sensitivity in distressing situations, individuals may rely on misleading information or feel inadequately supported. Over time, repeated interaction with automated systems can create dependency without clear safeguards or human oversight. This raises ethical concerns related to human dignity, reliability of support and accountability, particularly when automated responses influence decisions in sensitive and high-risk situations.

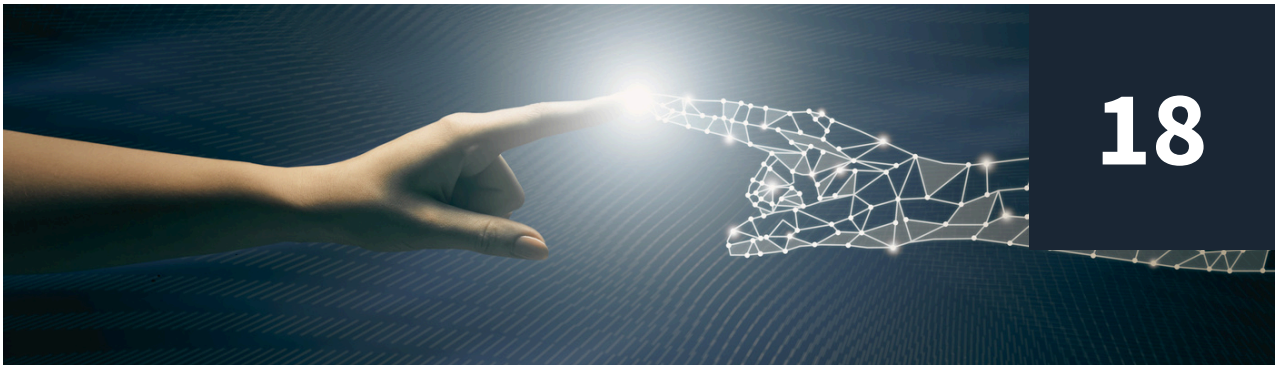
Protection of fundamental rights, freedoms and values as a guiding context

AI systems can create significant efficiencies and support broader societal objectives. However, their deployment may also directly affect fundamental rights, particularly in contexts involving victims or vulnerable groups. From a victim's perspective, the use of AI requires careful consideration of potential risks alongside expected benefits, ensuring that applications do not lead to disproportionate impacts or undermine access to protection and redress.

Fundamental rights and values should therefore be integrated throughout the entire lifecycle of AI systems, supported by transparency and accountability mechanisms, especially where automated decisions are involved. Understanding how these systems operate, how they are applied in practice and how they affect individuals, is essential to ensure that victims can effectively exercise their rights (e.g., to challenge, complain, or seek remedies). In situations where regulatory frameworks remain incomplete or evolving, fundamental rights continue to provide a baseline for assessing the legality and legitimacy of AI use, particularly in cases where individuals may be exposed to harm or unequal treatment.

Releated References

- European Commission (2026). Policies: AI Act.
- European Parliament (2025). Artificial Intelligence and Civil Liability: A European Perspective.
- UNESCO (2021). Recommendations on the Ethics of Artificial Intelligence.
- European Union Agency for Fundamental Rights (2025). Assessing High-risk Artificial Intelligence: Fundamental Rights Risks.
- AI4People (2018). AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations.
- European Parliament (2026). AI liability directive. In "A Europe Fit for the Digital Age".
- Stamford, C. (2026). Gartner Says General Counsel Should Assess AI Insurance to Mitigate AI Risks. Gartner Newsroom.
- European Commission (2024). Directive (EU) 2024/2853 on Product Liability.
- The Council of Europe (2024). The Framework Convention on Artificial Intelligence.
- European Commission (2026). Code of Practice on marking and labelling of AI-generated content.
- Tran, H. C. (2026). Unbounded Harms, Bounded Law: Liability in the Age of Borderless AI. arXiv preprint arXiv:2601.12646.



Highlights

AI is transforming how harm is generated, identified and addressed, introducing both enhanced capabilities and new vulnerabilities for victims. While it enables faster detection, broader reach and improved access to support, it also creates systemic challenges linked to opacity, scalability of harm and fragmented accountability. Overall, the balance between benefits and risks depends on how well technological, market and legal gaps are addressed in practice.

Benefits:

- **Faster identification of harmful activity:** AI enables quicker recognition of incidents, supporting more timely responses.
- **Ability to process large-scale information:** Systems can handle high volumes of data, improving coverage of complex environments.
- **Improved prioritisation of cases:** AI can help surface signals that may require urgent attention.
- **Enhanced detection of less visible harms:** Advanced analytics can uncover patterns not easily identifiable through traditional methods.

Risks:

- **Risk of misclassification and unjust targeting:** Automated assessments can lead to incorrect or disproportionate attention.
- **Limited transparency in decision-making:** Victims may not understand how or why actions affecting them are taken.
- **Reinforcement of existing inequalities:** Data-driven systems can reproduce and scale historical biases.
- **Barriers to contesting outcomes:** Complexity and lack of access to information hinder the ability to challenge decisions.

Overall, achieving a meaningful balance between these benefits and risks requires the consistent integration of safeguards across the lifecycle of AI systems. This includes ensuring transparency and explainability of outputs, embedding human oversight in decision-making processes, strengthening validation and monitoring mechanisms and enabling accessible pathways for affected individuals to challenge outcomes. Without such measures, the advantages of AI may be undermined by systemic gaps that disproportionately impact victims.

ENACT.

European Network Against Crime and Terrorism

@enact-network

enact-eu.net

Scan the QR code to visit ENACT online and read this report in digital format.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.