Vol. 9, Issue 1, Pp 75-86, Winter 2025

Socio-Spatial Studies

Journal of Socio-Spatial Studies

Homepage: https://soc.gpmsh.ac.ir

Original Research Paper

Privacy challenges of Artemis project: A tool to combat online child abuse

Mohammad Mahdi Ghassempour[1*]

[1] *PhD Student in Communication, University of Tehran, Iran*

ARTICLE INFO

Keywords:
Online child abuse
Artemis Project
Microsoft
Privacy
Cyberspace
digital surveillance
online chat

Received: Oct. 2, 2024
Revised: Dec. 3, 2024
Accepted: Dec. 19, 2024

ABSTRACT

Child exploitation in cyberspace is one of the most serious challenges of the digital age, requiring innovative approaches and advanced technologies to combat it. In this regard, Microsoft has developed Project Artemis—an analytical tool that identifies suspected grooming behaviors in online textual conversations and reports them to the relevant authorities. Although introduced as a solution to enhance children's safety, this technology raises concerns regarding user privacy, data collection and processing methods, and potential misuse of surveillance systems. Issues such as extensive monitoring of user communications, possible misidentification of conversations, and challenges related to data governance demand deeper investigation. This paper adopts a descriptive-analytical approach to examine these concerns, aiming to strike a balance between child protection and users' privacy rights. While Artemis proves to be an effective tool in identifying and addressing online child sexual abuse, it also faces limitations such as inefficacy in encrypted chats, lack of comprehensive preventive strategies, and privacy concerns. Full effectiveness requires the development of complementary technologies, legal reinforcement, policy collaboration, and broader engagement from communication platforms. Furthermore, designing localized tools tailored to national cultural contexts should also be considered.

INTRODUCTION

Online child exploitation is a major issue that necessitates a comprehensive approach to counteract. From parents compelling children to showcase themselves on social media, to sexual abuse occurring in virtual environments—all are examples of neglecting children's rights and endangering their mental and physical safety, which has unfortunately become increasingly prevalent. One of the primary platforms facilitating such abuse is online games and the associated chat features, which have often been exploited for predatory purposes.

Elizabeth Jeglic, a psychology professor at John Jay College in New York, highlights how easily child abusers infiltrate online chat rooms to find victims. She notes: "Abusers can reach sexual conversations with a child in 30 minutes—something that is far more difficult in the physical world. In the digital realm, it's accessible, fast, and efficient. If they fail with one target, they move on to another and can track children across multiple platforms to carry out their harmful intent. They often begin by requesting images and videos, escalating the sexual nature of the interaction. When the child feels fear or guilt, the

* Corresponding Author
✉ mm.ghassempour@ut.ac.ir
☎ +989126832143
iD 0009-0008-5346-7508

How to Cite this article:
Ghasempour, M. (2025). Privacy challenges of Artemis project: A tool to combat online child. *Socio-Spatial Studies*, 9(1), 61-73.
https://doi.org/10.22034/soc.2025.224662
URL: https://soc.gpmsh.ac.ir/article_224662.html

abuser may threaten to expose the content to friends and, using information gathered across platforms, coerce them into sending more." (Lyons, 2020)

Given Microsoft's own acknowledgment that ensuring complete security in cyberspace—especially in chatrooms—is impossible, addressing this issue to better protect children's online privacy has become increasingly critical.

However, implementing such technologies raises significant questions about user privacy. On the one hand, protecting children necessitates accurate monitoring and detection of potential threats. On the other hand, interfering with private user conversations can pose challenges related to privacy violations, data collection and processing, and the risk of misuse of surveillance systems. Microsoft itself admits that achieving total safety in the digital realm is unrealistic. Yet the central concern remains: how can we strike a balance between protecting children and preserving user privacy? This paper seeks to analytically examine the privacy-related challenges of Project Artemis and propose strategies to mitigate its negative consequences.

METHODOLOGY

This study adopts a descriptive-analytical methodology, aimed at critically evaluating the privacy implications of Project Artemis, a child protection technology developed by Microsoft. The research is grounded in an interdisciplinary theoretical framework that integrates insights from digital ethics, surveillance studies, and legal discourse surrounding child protection. Primary data sources include technical documentation, public statements by Microsoft and its collaborators (e.g., Thorn, The Meet Group), and relevant academic literature on surveillance technologies and digital privacy. Supplementary sources, such as journalistic investigations and policy papers by organizations like the European Data Protection Supervisor and the United Nations, were used to assess the broader sociopolitical implications of using surveillance tools to combat online child sexual abuse. This approach allows for a contextualized understanding of Artemis's dual role as both a protective mechanism and a potential privacy threat.

The analysis proceeds through a qualitative content analysis of publicly available documents, policy briefs, and academic literature to explore the ethical, technological, and regulatory dimensions of Project Artemis. Drawing upon case studies, such as YouTube's comment-section shutdown and Facebook's use of PhotoDNA, the paper evaluates Artemis's functionality in real-world scenarios. Ethical concerns, including the risk of false positives, user misidentification, and the potential misuse of surveillance systems by authoritarian regimes, are assessed through critical review of data protection frameworks such as the GDPR and COPPA. The methodological choice of a descriptive-analytical lens is particularly appropriate for identifying tensions between child protection imperatives and user privacy rights in digital spaces, facilitating recommendations that align both technological innovation and ethical responsibility.

FINDINGS

1. Privacy and Digital Privacy

The concept of privacy is a foundational principle in legal and ethical discourse, often defined as the right of individuals to control access to their personal information and to be free from unwanted intrusions (Solove, 2008). With the emergence and expansion of digital technologies, this concept has transcended physical boundaries and now encompasses a wide range of online interactions. Digital privacy refers to the protection of personal information that is generated, stored, or shared in digital environments—such as social networks, gaming platforms, and other online spaces (Westin, 1967; Tavani, 2008).

The core challenge in the realm of digital privacy lies in its increasing complexity in the face of hidden data collection mechanisms, algorithmic profiling, and behavioral targeting. Users—especially children and adolescents—are often unaware of the extent and methods of surveillance applied to their online behavior, making them highly vulnerable to various risks (Livingstone & Third, 2017).

2. Surveillance Society

The concept of a surveillance society refers to a social condition in which the collection, analysis, and utilization of personal information are conducted systematically and continuously through digital technologies and institutional structures (Lyon, 2001). French philosopher Michel Foucault, drawing inspiration from Bentham's prison design, proposed the Panopticon as a model of disciplinary power in which individuals, under constant and invisible observation, begin to regulate their own behavior autonomously (Foucault, 1977).

In the digital realm, this model has been structurally reproduced: users are continuously subject to observation, yet they are unaware of who is watching, when, and how. This asymmetry of visibility reflects a fundamental characteristic of algorithmic systems - their capacity to enforce observation while remaining opaque to both subjects and operators, creating power imbalances that research shows require deliberate governance interventions (Tomraee et al., 2024). Lyon (2006) refers to this condition as social sorting—a process whereby individuals are classified and assessed based on data collected about them. In the case of children, such surveillance can result in a paradox where efforts to ensure protection may unintentionally lead to greater violations of their privacy.

The concept of *surveillance capitalism*, introduced by Shoshana Zuboff (2019), offers a powerful theoretical lens through which the evolving relationship between personal data, corporate power, and individual autonomy can be understood. Unlike earlier forms of capitalism that relied on tangible labor or material production, surveillance capitalism extracts value from behavioral data. These data are not limited to what users knowingly share, such as clicks or messages; rather, they include derivative data—metadata, behavioral cues, and inferred preferences—that are often harvested without explicit consent. Zuboff argues that this behavioral surplus is processed using advanced machine learning algorithms to generate predictive products, which are then sold to advertisers, political actors, or other entities seeking to influence user behavior. This commodification of digital behavior turns users into raw material for data-driven economic systems, stripping away agency and rendering personal experience into a site of corporate profit-making.

Crucially, Zuboff situates this process within a broader socio-political context, warning that surveillance capitalism is not merely about profit generation; it also facilitates new architectures of control. As companies gain unprecedented access to the intimate routines, preferences, and vulnerabilities of individuals, they acquire not only the capacity to predict behavior but also to manipulate and shape it—what Zuboff terms *instrumentarian power*. This power, unlike traditional disciplinary power, does not require direct coercion. Instead, it subtly nudges individuals toward specific choices, behaviors, and patterns of engagement, often without their conscious realization. Over time, such systems can normalize surveillance as a default mode of digital existence, eroding expectations of privacy and autonomy. This normalization of digital control is also reinforced through mass media. Crisis-oriented media framing can significantly shape public cognition by reinforcing narratives of institutional failure, mistrust, or threat—thereby influencing how individuals internalize broader sociopolitical realities (Kharazmi & Mohammadi, 2020).

Within this conceptual framework, initiatives like Project Artemis, launched by Microsoft in 2020, warrant close scrutiny. Project Artemis is an automated tool designed to detect potential child grooming in text-based communications on online platforms, particularly gaming environments. By analyzing textual patterns and flagging suspicious interactions, the tool aims to identify predators before harm occurs. On the surface, this appears to be a noble application of artificial intelligence for the purpose of child protection—a task with clear ethical urgency. However, when viewed through the lens of surveillance capitalism, Artemis can also be understood as part of a growing trend in which private corporations leverage public safety imperatives to expand their surveillance infrastructures.

Indeed, Artemis exemplifies what Zuboff calls the "double logic" of surveillance technologies: while they offer clear benefits—such as enhanced security, crime prevention, and child protection—they simultaneously deepen corporate access to personal communication and behavioral data. Although Microsoft has emphasized that Artemis is used only in contexts where grooming is suspected, the operational opacity of the algorithm, the absence of independent oversight, and the potential for false positives raise substantial concerns. This reflects a broader tension in AI deployment—where protective tools are often entangled with data exploitation, prompting calls for transparent, publicly accountable governance (Sharifi Poor Bgheshmi & Sharajsharifi, 2025a). Such tools may inadvertently expand the surveillance frontier by embedding algorithmic monitoring into everyday communications, including those of innocent users. The risks become more pronounced in jurisdictions lacking robust data protection regulations or independent audit mechanisms, where such systems could be repurposed for broader social monitoring or political control.

Moreover, the deployment of Artemis in commercial digital spaces raises questions about privatized governance. As tech companies increasingly take on quasi-regulatory roles—deciding which content is acceptable, which behaviors are flagged, and which users are sanctioned—they also consolidate moral and legal authority over public digital spaces. This blurs the lines between corporate interest and public good, creating what scholars like Cohen (2019) have described as a "datafied public sphere" in which surveillance and commodification are structurally entangled. The ethical paradox becomes apparent: tools designed to safeguard children may also habituate society to pervasive monitoring, leading to a normalization of surveillance that outpaces democratic deliberation.

In this light, Artemis is not an isolated case but rather part of a broader shift in the digital ecosystem toward preventative surveillance—technologies that preemptively intervene based on predictive analytics. While such systems may reduce harm in specific contexts, they also reinforce the asymmetry of informational power between corporations and users. Without transparent accountability structures, the expansion of these technologies risks institutionalizing a form of behavioral governance that circumvents traditional legal and ethical checks. This constitutes not progress but predation - a fundamental reconfiguration of social power disguised as protection, where surveillance infrastructures expand under the moral cover of child safety while systematically evading democratic constraints (Toosi et al., 2025). As Zuboff (2019) cautions, what begins as a tool for protection can evolve into an architecture of influence, control, and quiet coercion, especially when embedded in environments where users have little understanding of or control over how their data is processed.

3. Balancing Protection and Privacy

The aforementioned theories illustrate that in the effort to protect children in cyberspace, there is always an inherent tension between security and privacy. Similar ethical tensions have been observed in other AI-driven systems, where predictive technologies, despite offering functional benefits, raise complex concerns around privacy, bias, and accountability—necessitating strong implementation safeguards and policy oversight (Nosraty et al., 2025). Tools like Artemis, which leverage artificial intelligence technologies, enable the rapid identification of potential predators. However, they simultaneously reproduce structures of pervasive surveillance and invisible control. This dual effect stems from a fundamental imbalance in protective AI systems - while their technical capabilities for threat detection advance rapidly, the necessary frameworks for ethical oversight and operational accountability consistently lag behind (Hosseini et al., 2021). Parallel research in organizational psychology emphasizes that perceived institutional support—shaped by age, education, and social background—significantly affects individuals' trust, well-being, and resilience within structured systems (Toosi, 2025). As Foucault (1977) notes, the most effective forms of power are those that become invisible and are internalized within individuals. In this context, media literacy has been proposed as a strategic competency that equips individuals to navigate digital environments more critically—helping users recognize subtle forms of influence, assess information validity, and resist the internalization of invisible control mechanisms (Arsalani et al., 2025).

Therefore, the theoretical and ethical assessment of Project Artemis must not only focus on whether the tool is capable of protecting children, but also critically examine how this protection is implemented and what impact it has on the freedom and autonomy of users. This aligns with broader information technology risk paradigms, where systems designed to mitigate one category of risk (e.g., child exploitation) often generate secondary risks (e.g., normalized surveillance) that demand equally strategic mitigation (Soroori Sarabi et al., 2023).

4. Combating Online Child Abuse

Many perpetrators of child abuse exploit the communication channels available on online chat platforms to harass and manipulate children. Typically, they begin by establishing a seemingly friendly relationship with their victims and gradually gain access to all aspects of their online and offline lives. This issue is widely observed on major gaming platforms such as Xbox (Lyons, 2020).

However, chat functionalities in online gaming are not limited to Xbox; similar risks exist on other platforms like Roblox, Minecraft, PlayStation, and others. The methods and the severity of the harm caused vary depending on the structure and features of each platform and must be evaluated accordingly (Jenson, 2017).

With the expansion of such platforms and the rise in incidents of child exploitation, there is a critical need for robust, reliable mechanisms to counter these threats. These include prompt detection and reporting of abuse by online gaming companies, swift removal of inappropriate content, and suspension of accounts belonging to perpetrators (Rogers & Woodhouse, 2016). Complementing these efforts, scholars have emphasized that media literacy can serve a preventive role by helping users—particularly young audiences—identify and resist media content that glamorizes high-risk or exploitative behaviors (Soroori Sarabi et al., 2020). In response to such incidents, legal frameworks have been introduced to combat and penalize these behaviors. One of the key approaches is holding platforms accountable and requiring them to implement monitoring and restriction mechanisms. For example, a 2017 report by *The Times* described a man being sentenced to 15 years in prison for threatening a child with rape and murder via Xbox chat.

Many platform providers are already aware of abuse occurring through their chat systems and have implemented various control strategies. For instance, Facebook claims to use machine learning to combat child exploitation. In 2018, it announced a system that tracks suspicious contact between adults and minors, using its extensive access to user data—greater than most other platforms (Lyons, 2020).

In response to widespread reports of child exploitation, YouTube suspended hundreds of channels and disabled the comment sections on tens of millions of videos featuring children. Andy Burrows, from a children's charity, confirmed that abusers were creating YouTube pages to connect with like-minded individuals and were using the comments section to pursue their malicious goals (Alexander, 2019).

In parallel with other platform initiatives, Microsoft, in collaboration with other companies, has launched Project Artemis—a technological effort to detect and prevent online child sexual exploitation. The following section discusses the details and mechanisms of this project.

5. Project Artemis and How It Works

Several technology companies have made dedicated investments to develop tools and technologies aimed at creating a safer digital environment for children. These efforts are designed to detect and control online child exploitation whenever such incidents occur. One such tool is Project Artemis, which appears to be effective in identifying child abuse in online chat environments. Originally based on Microsoft's experience with Xbox, this tool is now also available to other online communication services such as Skype. The motivation behind the development of Artemis was the alarming rise in online grooming and child sexual exploitation (Gregoire, 2020).

Project Artemis is designed to detect, track, and control instances of child exploitation in online chats. It was launched in November 2018 during the Microsoft 360 Cross-Industry Hackathon, with support from the WePROTECT Global Alliance and Child Dignity Alliance, two major child protection organizations. The project involved collaboration between Microsoft and companies such as The Meet Group, Roblox, Kik, and Thorn, with Hany Farid—creator of the *PhotoDNA* tool for detecting and reporting abusive child images—serving as the project lead (Lyons, 2020).

Hany Farid, a renowned academic, had already collaborated with Microsoft and Dartmouth College in 2009 to develop PhotoDNA, a free tool that has helped detect, disrupt, and report millions of images of child sexual abuse. Today, it is used by over 150 organizations worldwide (Gregoire, 2020).

The operational model of Artemis starts with detecting suspicious words and conversational patterns within online chats. These flagged conversations assist human moderators in identifying potential threats. The tool evaluates chat content based on historical dialogue structures and assigns risk scores to conversations. These scores guide companies in determining which chats need further review and intervention.

After this initial screening, flagged chats are reviewed by a human moderator who assesses the severity and legitimacy of the threat. Depending on the outcome, appropriate reports are sent to law enforcement agencies or child protection organizations. In some cases, if the AI detects a high-risk, urgent situation, the data may be forwarded directly to authorities without the need for human intervention.

Julie Cordua, CEO of Thorn, a nonprofit focused on preventing child exploitation, emphasizes: “Abuse happens on nearly every chat-enabled platform. Instead of blaming platforms, we must encourage them to use proactive tools. If someone claims their platform is abuse-free, we should question their perception” (Lyons, 2020).

Microsoft has made Artemis available to all companies offering chat services, with the goal of standardizing monitoring practices and developing preventive tools against future abuse. The company claims that Artemis contributes to a safer online world for children by creating cleaner environments free from sexual exploitation and other dangers. Yet technological safeguards alone are not sufficient. Recent research emphasizes that equipping children with media literacy skills—alongside the active involvement of parents—can significantly enhance their ability to critically engage with digital content and resist harmful online influences. Such integrative educational approaches are vital complements to algorithmic detection systems, particularly in addressing the social and psychological dimensions of online child safety (Hosseini et al., 2025).

However, moderating chats in the context of child abuse is challenging due to the subtle nuances and variations in conversational tone. Platforms differ significantly—game chat rooms are not the same as social media messaging apps—so Artemis must be adaptable. The system allows customization to match the specific communication patterns of different platforms.

Currently, Thorn is responsible for managing Artemis licensing and providing technical support. From January 10, the project began distributing licenses to interested companies. Service providers who wish to test or implement the technology are instructed to contact Thorn directly via their website or email (Gregoire, 2020).

6. Privacy Challenges in Project Artemis

One of the most significant concerns surrounding Project Artemis is the extensive surveillance it imposes on user conversations across online platforms. The Child Rights International Network (CRIN) has highlighted the challenges and limitations associated with technologies designed to detect online child sexual abuse, emphasizing growing concerns over their implications (EU Draft Regulation, 2023).

Such surveillance can potentially lead to violations of user privacy, as detection systems analyze the content of private messages, which may indirectly result in access to individuals’ personal information. Research on AI-driven surveillance underscores these concerns, demonstrating how automated content analysis—even when designed for protective purposes—can inadvertently normalize intrusive data practices and weaken privacy safeguards (Toosi et al., 2024). While the primary goal of such monitoring is to identify and prevent child abuse, it remains crucial to ensure that individual privacy rights are upheld, and that data is collected and processed only when absolutely necessary and with a high degree of accuracy (Fotiadis, 2025).

Prior to the development of Project Artemis, Microsoft had introduced PhotoDNA, a technology designed to identify illegal images. This tool has been adopted by many companies, including Facebook and Twitter, for detecting and reporting abusive visual content (Microsoft, 2011). However, some critics argue that such technologies—if not subject to strict oversight—could evolve into tools for broad surveillance over user communications (European Data Protection Supervisor, 2021).

Similarly, content analysis systems used by Google and YouTube's automated content moderation algorithms have a history of erroneously blocking legitimate, non-threatening content, raising concerns about the accuracy and reliability of these detection tools (Google Support, 2023). These types of errors are seen as significant operational failures.

Furthermore, the collection of user conversation data for the purpose of identifying potential child abuse poses additional complexities (Deldari et al., 2023). These conversations may contain sensitive personal information, which, if misused, could become a serious privacy threat (Chou et al., 2025). In addition, there are significant security challenges associated with protecting this data. If the surveillance systems are not fully secure, unauthorized access by hackers or malicious actors could expose user information (Novikava, 2024). In such cases, users' private data may be put at serious risk. A recurring challenge in AI implementation is the gap between institutional enthusiasm for innovation and the readiness to manage ethical, educational, and governance-related risks—an imbalance evident across domains, from business education to child protection technologies (Rahmatian & Sharajsharifi, 2021).

There have been past incidents where surveillance technologies have led to data leaks involving sensitive user information. For example, in 2019, a security flaw in Facebook's content detection system exposed the private data of millions of users (Wired, 2021). These cases demonstrate that, even when surveillance tools are developed with good intentions, inadequate data security can lead to unintended and harmful consequences.

It is also important to recognize that surveillance technologies used in projects like Artemis can, if misused, become tools for controlling and monitoring user behavior by governments or corporations (United Nations Office of the High Commissioner for Human Rights, 2022). In such cases, concerns arise regarding the preservation of individual freedoms and the right to privacy. This is particularly critical in countries with weak legal and regulatory safeguards, where such technologies might be turned into instruments of oppression and user control (Federal Register, 2025). To prevent such misuse, there must be clear, enforceable laws and regulations governing the deployment of these technologies. Research in broader regulatory contexts shows that fragmented or ambiguously defined legal frameworks—like those seen in economic crime legislation—often result in enforcement inefficiencies and strategic blind spots. This underscores the need for clear, preventive, and well-structured policies in the realm of AI governance as well, where long-term effectiveness depends not only on technical capacity but on legal coherence and design (Taheri et al., 2022).

A key issue in evaluating Project Artemis is its compliance with international privacy standards, such as the General Data Protection Regulation (GDPR) in the European Union and the Children's Online Privacy Protection Act (COPPA) in the United States (Stephens, 2018). These regulations emphasize that user data—especially children's data—must be stored and processed according to the highest security standards. If Artemis fails to properly align with these frameworks, it could face legal and regulatory challenges (Stephens, 2018). This concern reflects a broader legal pattern: when enforcement systems are built on ambiguous or inconsistently applied rules, they often fail to gain societal compliance—no matter how coercive or technologically advanced. As legal scholarship in other domains shows, effectiveness depends on whether laws are grounded in formal legislative principles that enhance their legitimacy, clarity, and preventive capacity (Aghigh et al., 2022).

Another concern is the potential for error in AI-based detection technologies. Artemis systems may misidentify benign conversations as suspicious, flagging messages that pose no actual threat to children (Leschanowsky et al., 2024; European Parliament, 2023). These algorithmic failures become particularly problematic when users lack the technical literacy to understand or challenge the system's decisions - a well-documented gap in public understanding of AI surveillance tools (Khodabin et al., 2022). Such false positives can lead to unjust consequences, such as erroneous reports to law enforcement or unwarranted restrictions on users' accounts. For this reason, ensuring high precision in identifying and evaluating conversations, and minimizing diagnostic errors, is of critical importance.

Some cybersecurity experts argue that despite these challenges, technologies like Artemis are essential tools for protecting children in the digital space. They maintain that the benefits of such tools in reducing child abuse outweigh the privacy concerns. For instance, the RAINN organization, which collaborates with the Artemis project, considers it a key component of the social platforms' toolkit for combating child exploitation (Connelly, 2022).

Conversely, digital rights activists have raised strong warnings that without rigorous oversight, such systems could result in widespread privacy violations. An article in *The Guardian* highlighted growing concerns about a European Union proposal to scan digital messages in order to combat child sexual abuse, warning that this initiative could open the door to massive, unchecked surveillance (Viner, 2025). A similar tension is visible in global healthcare debates, where AI is seen as a vital innovation despite serious concerns about privacy, bias, and governance—underscoring the need for balanced frameworks that enable impact without compromising rights (Toosi et al., 2025).

7. Tools for Combating Online Child Abuse in Iran

The global architecture of the internet is deeply marked by structural imbalances in governance and infrastructure control, with the United States playing a dominant role in setting regulatory standards and hosting the critical services that underpin the digital ecosystem. These structural imbalances have enabled what scholars term 'algorithmic colonization' - where tools like Project Artemis, developed by transnational corporations under U.S. jurisdiction, effectively dictate child protection standards worldwide while bypassing national sovereignty frameworks (Sharifi Poor Bgheshmi & Sharajsharifi, 2025b). This asymmetry presents a profound challenge for many countries, particularly those in the Global South, in managing their cyberspace autonomously. Authoritarian or monopolistic dynamics in internet governance—characterized by the concentration of domain name systems, cloud services, and security protocols in the hands of a few Western corporations—have created a de facto dependency that limits national sovereignty in digital policymaking and infrastructure development. One of the most acute consequences of this imbalance is the constrained ability of nations to develop, implement, and enforce cybersecurity policies tailored to local needs, especially in highly sensitive areas such as online child protection.

Online child abuse, including grooming, exploitation, and trafficking, has emerged as one of the gravest digital threats in recent years. The increased online presence of children and adolescents—accelerated by remote education, social media use, and mobile technology—has created new vulnerabilities that are exploited by offenders operating across national boundaries. While international cooperation and technological tools such as Microsoft's Project Artemis are designed to detect and respond to online grooming behavior, these tools are often bound by legal, infrastructural, and political constraints that limit their applicability in non-Western contexts. For instance, Artemis is embedded within U.S. legal frameworks, relies on English-language communication patterns, and is integrated into platforms subject to U.S. law enforcement cooperation protocols. As a result, its effectiveness in countries like Iran is significantly diminished, both technically and jurisdictionally. This pattern reflects broader global concerns about AI as a driver of technological dependency, where countries lacking regulatory parity or infrastructural independence are often forced to adopt tools embedded in foreign legal and governance frameworks. Research on AI's geopolitical impacts shows that such asymmetries can undermine sovereignty and entrench strategic disadvantages for emerging economies (Rahmatian, 2025).

Iran, like many other nations outside the Western digital sphere, faces systemic limitations in adopting such technologies. Sanctions, restricted access to U.S.-based services, and legal incompatibilities with American platforms hinder seamless integration. Furthermore, tools like Artemis are developed within specific sociocultural and linguistic assumptions that do not translate easily across diverse digital environments. Consequently, relying on foreign technologies for the detection and prevention of online child abuse can not only produce suboptimal outcomes but also compromise data security and national legal consistency. This underscores the urgent need for technological self-sufficiency, particularly in areas as ethically and politically sensitive as child protection.

In response to these constraints, the concept of a National Information Network (NIN) has gained traction as a strategic initiative aimed at ensuring digital sovereignty. The NIN seeks to build a domestic infrastructure that supports secure access to national data services, enables independent oversight, and reduces dependency on foreign technology providers. Beyond mere infrastructural autonomy, the NIN envisions fostering a culture of technological localization, whereby digital tools are developed with close alignment to local legal systems, cultural norms, and linguistic diversity. In the context of child protection, this would entail the creation of indigenous software capable of detecting contextually relevant patterns of online abuse, informed by localized definitions of harm, family norms, and judicial frameworks.

For countries like Iran, where specific cultural and religious norms intersect with cybersecurity concerns, localized approaches are not merely preferable but necessary. Western tools often fail to account for culturally specific behaviors, online discourse patterns, and reporting protocols. A domestically developed detection system could incorporate Persian-language nuances, local idioms, and national content regulation criteria—making it both more effective and more aligned with national values. This cultural mismatch underscores a fundamental design flaw in one-size-fits-all detection systems: they consistently underestimate how digital literacy gaps and local communication norms shape online interactions—a pattern well-documented across diverse user groups (Sakhaei et al., 2024). Moreover, the governance of such systems would remain within national jurisdiction, thereby enhancing accountability, legal coherence, and public trust.

The development of indigenous child protection technologies within a national network model also enables strategic innovation. By involving local experts, policymakers, educators, and civil society in the design and implementation process, such systems can be calibrated to address not only online grooming and exploitation, but also broader issues such as cyberbullying, exposure to harmful content, and digital literacy deficits among youth and parents. These integrated strategies foster a more resilient digital environment, reducing long-term dependency on external actors and empowering nations to define their own standards of digital ethics and child safety.

Ultimately, the challenge of online child abuse cannot be meaningfully addressed in isolation from the structural conditions of global internet governance. As long as critical technologies remain under the control of a few dominant actors, and as long as digital infrastructures are unequally distributed, many countries will struggle to implement protective measures that are both effective and sovereign. The pursuit of localized, culturally sensitive, and legally consistent tools—embedded within broader strategies for national digital autonomy—is therefore not only desirable but imperative for ensuring the safety and dignity of children in an increasingly networked world.

8. Challenges and Barriers in Developing Indigenous Tools

One of the major obstacles in the development of indigenous tools is the presence of legal and regulatory barriers, particularly at the international level and especially in countries with strict oversight frameworks. Currently, many large tech companies, due to their dominance over cyberspace, formulate their services and policies in accordance with the laws of their home countries. As a result, other nations—especially those with different legal and regulatory systems—often find it difficult to utilize these tools effectively.

This issue is particularly critical in the context of online child abuse, where effective intervention requires access to data and constant monitoring of online interactions. Relying on foreign tools that are governed by external legal frameworks could pose risks to national interests and sovereignty.

Therefore, developing homegrown technologies that are tailored to a country's specific cultural, legal, and social context is not only a step toward addressing digital harms, but also a strategic move to enhance national sovereignty and cybersecurity. Such tools, if designed in alignment with domestic laws and regulations, can provide safer online environments for children within each country.

9. The Importance of Addressing Online Child Abuse in Iran

Among the essential prerequisites for achieving this goal is increased attention to the issue of online child abuse within the country. This matter has been largely overlooked in Iran, receiving insufficient consideration. Online child abuse stands as one of the most significant threats in cyberspace, necessitating dedicated attention from governmental institutions, non-governmental organizations (NGOs), and other social stakeholders. Many online risks and threats that may endanger children are neglected due to a lack of public awareness, inadequate societal education, and insufficient infrastructural safeguards.

Immediate and coordinated action by governments and relevant organizations can pave the way for the development of technological and regulatory solutions. Such measures would not only mitigate existing harms but also establish a foundation for designing and implementing indigenous tools to combat these threats. In this regard, formulating precise policies to counter online child abuse and supporting domestic initiatives in this field could mark a turning point in enhancing the safety and well-being of children in Iran's digital space.

CONCLUSION

It should be noted that while Project Artemis represents a significant step forward, as previously discussed, it cannot be considered a flawless or all-encompassing solution. One key area often overlooked is critical AI literacy—developing user-centered educational frameworks that enable individuals, especially young users, to interrogate algorithmic systems, recognize bias, and navigate digital environments ethically and autonomously (Khodabin et al., 2024). Despite addressing certain gaps, its limitations must be acknowledged—especially those stemming from structural, institutional, or educational readiness. Studies on AI in education emphasize that without institutional modernization—both curricular and generational—technological integration is likely to falter, regardless of perceived potential or user enthusiasm (Tomraee et al., 2025). For instance, it remains unclear whether the tool is compatible with end-to-end encrypted chat platforms. Additionally, further measures to prevent observer stress disorder in this context have yet to be defined. Another limitation of Artemis is its language support—currently restricted to English—which necessitates expansion given the global scope of online chat platforms. However, published reports indicate that serious efforts are underway to address these challenges, including the integration of additional languages and supplementary tools to enhance the technology.

Microsoft has described this project as a useful yet preliminary step in combating online child exploitation. Cordova, emphasizing the need for platforms to revise their self-regulatory approaches, argues that prevention must remain a core focus. With Artemis, the first step toward better detection has been taken, and all chat and video service providers across platforms must adapt accordingly—failure to do so may result in severe consequences, including unchecked child exploitation on their platforms (Lyons, 2020).

Overall, sexual exploitation—particularly online child sexual abuse—and ensuring a safe digital environment for children remain complex and formidable challenges. Similar to other high-risk domains such as epidemic response, the successful application of AI to child protection challenges hinges not only on technological sophistication, but on the presence of robust regulatory frameworks, data reliability, ethical oversight, and interdisciplinary collaboration. Studies show that public trust and system effectiveness grow when AI is embedded within structured, accountable ecosystems that balance innovation with governance (Sakhaei et al., 2024). While tools like Artemis mitigate some issues, achieving the defined objectives requires broader collaboration among technology companies and organizations, alongside sustained refinement of existing measures. Given the gravity of combating online child exploitation, a global call to action is imperative. Among the most foundational elements of such a global response is sustained investment in education—both technical and ethical—that fosters critical awareness, institutional resilience, and a culture of responsibility across organizations and platforms. As research has shown, education not only equips personnel to manage technological complexity, but also strengthens their capacity to uphold values, respond adaptively to evolving risks, and align organizational practice with broader societal obligations (Zamani et al., 2024).

Furthermore, beyond the technological aspects of this project, legal, policy, and enforcement dimensions must be prioritized. This point is echoed in other high-stakes domains such as healthcare, where the integration of AI has revealed deep concerns around explainability, data privacy, and professional accountability. Studies show that even well-intentioned AI systems face resistance when transparency is lacking or ethical frameworks are underdeveloped—highlighting the need for targeted training, regulatory clarity, and human-centered design as prerequisites for responsible AI deployment (Tomraee et al., 2022). Reports suggest that over the past 14 months, the project has made progress in identifying, tracking, and controlling cases, demonstrating its operational efficacy (Gregor, 2020). Similar insights have emerged in broader assessments of AI integration, where transformative potential is consistently accompanied by infrastructural, ethical, and institutional challenges—requiring adaptive, human-centered strategies that align innovation with societal values and inclusive governance (Rahmatian & Sharajsharifi, 2022).

However, since Microsoft's exclusive development of this tool raises concerns about centralized control—particularly given Iran's internet penetration rates—greater attention must be paid to domestic solutions for combating online child exploitation. This underscores the urgent need to advance Iran's National Information Network (NIN) by leveraging gaps in the current global internet governance monopoly. Strengthening the NIN's technical, infrastructural, and content frameworks—alongside developing indigenous tools to counter online child abuse—should be prioritized to accelerate the breakdown of this digital hegemony.

CONFLICT OF INTEREST

No conflict of Interest declared by the author(s).

REFERENCES

Aghigh, S. R., Salehi, K., & Barkhordari, A. (2022). Ignoring the Legal Requirements of Criminal Law in the Field of Crimes Against Security. *Public Law Knowledge Quarterly, 11*(38), 104-140. doi: 10.22034/qjplk.2022.1490.1393

Alexander, J. (2019). YouTube is disabling comments on almost all videos featuring children. Available at: theverge.com

Arsalani, A. , Rahmatian, F. and Hosseini, S. H. (2025). Media literacy for business personnel: A strategic approach for better efficiency. *Code, Cognition & Society, 1*(1), 1-28. https://doi.org/10.22034/ccsr.2025.526844.1000

Chou, K.H., et. al. (2025). Bots can Snoop: Uncovering and Mitigating Privacy Risks of Bots in Group Chats. ARTIFACT EVALUATED

Connelly, C. (2022). Tech Innovator on the Fight to Create a Safer Internet and Prevent Child Predators. Available at: rainn.org

Deldari, E., et al. (2023). Users' Perceptions of Online Child Abuse Detection Mechanisms. ACM Journals. Proceedings of the ACM on Human-Computer InteractionVol. 8, No. CSCW1.

Child Rights International Network (CRIN). (2023). *Explaining the technology for detecting child sexual abuse online*. Retrieved from https://home.crin.org/readlistenwatch/stories/explainer-detection-technologies-child-sexual-abuse-online

Europarl (2023). Proposal for a regulation laying down the rules to prevent and combat child sexual abuse. Available at: europarl.europa.eu

European Data Protection Supervisor (2021). ePrivacy new developments. Available at: edps.europa.eu

Federalregister (2025). Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons. Available at: federalregister.gov

Fotiadis, A. (2025). The EU wants to scan every message sent in Europe. Will that really make us safer? Available at: theguardian.com

Foucault, M. (1977). Discipline and punish: The birth of the prison (A. Sheridan, Trans.). Vintage.

Google Support (2023). Google Search Console reports false positive harmful content issue deceptive pages. Available at: support.google.com

Gregoire, C. (2020). Microsoft shares new technique to address online grooming of children for sexual purposes. Available at: blogs.microsoft.com

Hosseini, S. H., Khodabin, M. , Soroori Sarabi, A., & Sharifi Poor Bgheshmi, M.S. (2021). Artificial intelligence and disaster risk management: A need for continuous education. *Socio-Spatial Studies, 5*(1), 13-29. https://doi.org/10.22034/soc.2021.219422

Hosseini, S. H. , Nosraty, N. and Tomraee, S. (2025). Children, Healthy Lifestyle and Media Literacy. *Journal of Cyberspace Studies, 9*(1), 1-23. https://doi.org/10.22059/jcss.2024.387609.1120

Jenson, K. (2017). Pedophiles Hunt Kids in Popular Gaming Chat Rooms. Available at: protectyoungminds.org

Kharazmi, Z., & Mohammadi, S. (2020). Persian-Language Media Overseas as the Western Tools of Public Diplomacy: Framing COVID-19 Pandemics in Iran by VOA and BBC. *Journal of World Sociopolitical Studies, 4*(1), 1-36. https://doi.org/10.22059/wsps.2020.308749.1171

Khodabin, M., Sharifi Poor Bgheshmi, M.S., & Movahedzadeh, F. (2024). Critical AI Literacy: Preparing Learners for Algorithmic Societies. *Journal of Cyberspace Studies, 8*(2), 371-397. https://doi.org/10.22059/jcss.2024.102582

Khodabin, M., Sharifi Poor Bgheshmi, M. S., Piriyaei, F., & Zibaei, F. (2022). Mapping the landscape of AI literacy: An integrative review. *Socio-Spatial Studies, 6*(1). https://doi.org/10.22034/soc.2022.223715

Leschanowsky, A., et al. (2024). Evaluating privacy, security, and trust perceptions in conversational AI: A systematic review. Computers in Human Behavior. Volume 159, 108344

Livingstone, S., & Third, A. (2017). Children and young people's rights in the digital age: An emerging agenda. *New Media & Society*, 19(5), 657–670. https://doi.org/10.1177/1461444816686318

Lyon, D. (2001). *Surveillance society: Monitoring everyday life*. Open University Press.

Lyon, D. (2006). 9/11, synopticon, and scopophilia: Watching and being watched. The Sociological Review, 54(2_suppl), 221–230. https://doi.org/10.1111/j.1467-954X.2006.00676.x

Lyons, K. (2020). Microsoft tries to improve child abuse detection by opening its Xbox chat tool to other companies. Available at: theverge.com

Microsoft (2011). Facebook to Use Microsoft's PhotoDNA Technology to Combat Child Exploitation. Available at: blogs.microsoft.com

Nosraty, N., Soroori Sarabi, A., Arsalani, A., Toosi, R., & Sharajsharifi, M. (2025). Artificial intelligence for disaster risk management in the beauty industry. *International Journal of Advanced Multidisciplinary Research and Studies, 5*(3), 1076-1086. https://doi.org/10.62225/2583049X.2025.5.3.4422

Novikava, A. (2024). How to prevent unauthorized access: 10 best practices. Available at: nordlayer.com

Ohchr (2022). Spyware and surveillance: Threats to privacy and human rights growing, UN report warns. Available at: ohchr.org

Rahmatian, F. (2025). From silicon to sovereignty: MBA students' views on AI's disruption of global power dynamics. *Journal of World Sociopolitical Studies, 9*(4).

Rahmatian, F., & Sharajsharifi, M. (2021). Artificial intelligence in MBA education: Perceptions, ethics, and readiness among Iranian graduates. *Socio-Spatial Studies, 5*(1). https://doi.org/10.22034/soc.2021.223600

Rahmatian, F., & Sharajsharifi, M. (2022). Reimagining MBA education in the age of artificial intelligence: A meta-synthesis. *Socio-Spatial Studies, 6*(1). https://doi.org/10.22034/soc.2022.223610

Rogers, D., & Woodhouse, J. (2016). *Prevention of online child abuse* (CDP-2016-0146). House of Commons Library. https://researchbriefings.files.parliament.uk/documents/CDP-2016-0146/CDP-2016-0146.pdf

Solove, D. J. (2008). *Understanding privacy*. Harvard University Press..

Sakhaei, S., Soroori Sarabi, A., Tomraee, S., Khodabin, M., & Sharajsharifi, M. (2024). Disaster risk management and AI: A grounded theory approach to epidemic response. *International Journal of Advanced Multidisciplinary Research and Studies, 4*(3), 1699-1708. https://doi.org/10.62225/2583049X.2024.4.3.4420

Sakhaei, S., Soroori Sarabi, A., & Alinouri, S. (2024). Teaching IT Use to Elderly: A Media Literacy Solution. *Journal of Cyberspace Studies, 8*(2), 295-316. https://doi.org/10.22059/jcss.2024.101608

Sharifi Poor Bgheshmi, M.S., & Sharajsharifi, M. (2025a). Between exploitation and resilience: Reconciling AI's role in surveillance capitalism and disaster risk management. *Journal of Cyberspace Studies, 9*(2).

Sharifi Poor Bgheshmi, M. S., & Sharajsharifi, M. (2025b). Managing the crisis: AI and the demise of national sovereignty? *Journal of World Sociopolitical Studies, 9*(4).

Soroori Sarabi, A., Arsalani, A., & Toosi, R. (2020). Risk management at hazardous jobs: A new media literacy?. *Socio-Spatial Studies, 4*(1), 13-25. https://doi.org/10.22034/soc.2020.212126

Soroori Sarabi, A. , Zamani, M. , Ranjbar, S., & Rahmatian, F. (2023). Innovation – But with Risk: The Strategic Role of IT in Business Risk Management. *Journal of Cyberspace Studies, 7*(2), 253-275. https://doi.org/10.22059/jcss.2023.101605

Staff, T. (2020). Meet the new anti-grooming tool from Microsoft, Thorn, and our partners. Available at: thorn.org

Stephens, A. (2018). The relationship between COPPA and GDPR: getting it right for your business. Available at: privacycompliancehub.com

Taheri, M., Milani, A. R., & Salehi, K. (2022). Studying the Legal Criminal Policy of Iran and England Regarding Economic Crimes. *Medical Law Journal, 16*, 1022-1035. http://ijmedicallaw.ir/article-1-1729-en.html

Tavani, H. T. (2008). Informational privacy: Concepts, theories, and controversies. In K. E. Himma & H. T. Tavani (Eds.), The Handbook of Information and Computer Ethics (pp. 131–164). Wiley.

Tomraee, S. , Hosseini, S. H., & Toosi, R. (2022). Doctors for AI? A systematic review. *Socio-Spatial Studies, 6*(1), 13-26. https://doi.org/10.22034/soc.2022.219431

Tomraee, S., Hosseini, S. H., Zamani, M., & Sakhaei, S. (2025). Perceptions of Iranian medical students on artificial intelligence in healthcare and curricular integration. *International Journal of Advanced Multidisciplinary Research and Studies, 5*(3), 1107-1117. https://doi.org/10.62225/2583049X.2025.5.3.4425

Tomraee, S. , Toosi, R. and Arsalani, A. (2024). Perspectives of Iranian Clinical Interns on the Future of AI in Healthcare. *Journal of Cyberspace Studies, 8*(2), 347-370. https://doi.org/10.22059/jcss.2024.101610

Toosi, R. (2025). A survey examination of psychological support in the workplace. *Journal of Cyberspace Studies*, 9(2).

Toosi, R., Hosseini, S. H., Nosraty, N., & Rahmatian, F. (2024). Artificial intelligence, health, and the beauty industry. *International Journal of Advanced Multidisciplinary Research and Studies, 4*(3), 1689-1698. https://doi.org/10.62225/2583049X.2024.4.3.4419

Toosi, R., Nosraty, N., & Tomraee, S. (2025). Using AI to enhance health: A global perspective. *Journal of World Sociopolitical Studies, 9*(4).

Toosi, R. , Tomraee, S. , Khodabin, M. and Sakhaei, S. (2025). Telemedicine: An AI solution, at last?. *Code, Cognition & Society, 1*(1), 59-87. https://doi.org/10.22034/ccsr.2025.526987.1002

Viner, Katharine. 2025. The EU wants to scan every message sent in Europe. Will that really make us safer? Available at: theguardian.com

Westin, A. F. (1967). Privacy and freedom. Atheneum.

Wired. 2021. Facebook Had Years to Fix the Flaw That Leaked 500M Users' Data. Available at: wired.com

Zamani, M. , Hosseini, S. H. and Rahmatian, F. (2024). The Role of Education in Successful Business Management. *Journal of Cyberspace Studies, 8*(2), 317-346. https://doi.org/10.22059/jcss.2024.101609

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.