https://doi.org/10.48047/AFJBS.6.Si4.2024.908-918

African Journal of Biological Sciences

AI Privacy Policies and Security

Madhuri Nallam[1*], Rishith Murala[2], S V V Sai Vara Praveen Maganti[3], Bhargavi Chinni[4],Murali Mohan Vutukuru[5], Dr.M.Madhusudhana Subramanyam[6]

[1*]Department of Computer science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email: -madhuri.nallam123@gmail.com
[2]Department of Computer science and Engineering Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email: - rishithmurala25@gmail.com
[3]Department of Computer science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email: - praveenpj1414@gmail.com
[4]Department of Computer science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email: -bhargavichinni2019@gmail.com
[5]Department of Computer science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email: - muralimohan.klu@gmail.com
[6]Department of Computer science and Engineering Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email: - mmsnaidu@yahoo.com

*Corresponding Author:Madhuri Nallam
Department of Computer science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram, Andhra Pradesh, India Email:-madhuri.nallam123@gmail.com

Article History
Volume 6, Issue Si4 2024
Received: 23 May 2024
Accepted: 08 June 2024
Published: 06 July 2024
Doi: 10.48047/AFJBS.6.Si4.2024.908-918

ABSTRACT

AI technologies often trigger concerns about privacy, a concept whose meaning can be challenging to grasp. Consequently, people's worries about privacy tend to be unclear, complicating efforts to address these concerns and clarify how AI either poses or doesn't pose threats to individuals. This article highlights overlooked distinctions and explored their impact on concerns about how AI technology affects privacy threats. It implies that security, rather than the fundamentals of privacy, is frequently brought up when individuals voice concerns about privacy with regard to artificial intelligence. However, the focus on security overlooks the significance of privacy in fostering autonomy and shaping our identities. Enhancing understanding of these nuances could assist AI developers in explaining to users which interests are affected and which are not using AI systems.

Keywords - Artificial intelligence (AI), Anonymization Techniques, Firewalls, Multi-factor authentication, Data Encryption, Blockchain Technology, Privacy, Security, Autonomy, Data Security.

I. INTRODUCTION

Privacy stands out as one of the key worries surrounding AI technologies. However, understanding the concept of "privacy"has been elusive, evident in the absence of a definitive definition within philosophical or legal realms. Philosophical discussions often portray it as a 'concept in disarray' (Solove, 2008) or an entrance into an 'unknown swamp' (Inness, 1992), with some arguing that privacy lacks a singular, coherent meaning (Thomson, 1975; Solove, 2008). Consequently,

people's concerns about privacy are often vague, making it challenging to address these worries and articulate how AI technologies either endanger or don't endanger individuals' interests. This article aims to highlight overlooked distinctions and elucidate their impact on concerns regarding how AIrelated technology threatens privacy. The prevalent understanding of privacy commonly centres on the idea of control, especially concerning personal information. This perspective asserts that privacy predominantly involves the authority to block unauthorized access or usage of one's data without explicit permission. Despite its dominance, this viewpoint has encountered substantial critique from scholars like Judith Thomson and Tom Macnish (Thomson, 1975; Macnish, 2018). We won't go into great detail about this debate here, but we argue that it's oversimplified to equate privacy with control extensively delve into this discourse, we contend that equating a lack of control with a lack of privacy oversimplifies the concept. To truly comprehend why this occurs, exploring the subtleties of privacy, notably the pivotal role played by sentient entities capable of semantic comprehension, is imperative. Discussions concerning artificial intelligence (AI) and privacy frequently grapple with balancing the advantages of granting companies unrestricted access to user data, such as tailored services and corporate profits, against the "right to privacy". This essay delves into a pivotal yet often disregarded facet of this discourse: the intrinsic value of user control over personal information. The typical argument advocating for user control primarily revolves around shielding individuals from potential harm stemming from unauthorized data access. This encompasses concerns like identity theft, loan rejection, and other security breaches. While these "security interests" undoubtedly hold weight, we posit that another distinct and equally significant interest is often overlooked: an interest in privacy itself. This lessaddressed interest in privacy is the cornerstone of traditional privacy concepts and holds immense value due to its link to human autonomy. It bestows individuals with the authority to dictate how their personal information is utilized, shaping their identity and interactions with the world. Despite its significance, the interaction between AI and privacy, viewed from this perspective, has largely been disregarded. This essay aims to rectify this neglect by examining the interplay between AI and privacy as an essential human entitlement. while acknowledging the validity of security concerns surrounding personal data, this essay sheds light on the frequently overlooked interest in privacy itself, intrinsically linked to human autonomy. Recognizing this distinction is pivotal in shaping AI technologies that honour and uphold individual rights to privacy.

Type of Privacy Policy	Description
Data Collection Policy	Outlines what types of data are collected, how they are collected, and for what purposes.
Data Usage Policy	Describes how collected data is used by the organization, including any sharing with third parties.
Data Retention Policy	Specifies the duration for which data is retained and the methods used for secure storage and eventual deletion.
Data Protection Policy	Details the security measures in place to protect data from unauthorized access, breaches, and other threats.
User Rights Policy	Explains the rights users have regarding their data, including access, correction, deletion, and portability.
Consent Policy	Describes how user consent is obtained, recorded, and managed, especially for sensitive data and marketing purposes.
Cookie Policy	Provides information on the use of cookies and similar technologies on the organization's website and services.
Third-Party Sharing Policy	Details the conditions under which data is shared with third parties, including partners, affiliates, and vendors.
Compliance Policy	Describes adherence to relevant data protection laws and regulations (e.g., GDPR, CCPA).
Privacy by Design Policy	Outlines how privacy considerations are integrated into the development and operation of products and services.

Fig1: Types of Privacy Policy

II. LITERATURE REVIEW

This study offers a comprehensive review of the literature on AI security and privacy regulations. First, we can discuss the importance of AI privacy policy. AI systems often collect and process large amounts of data, so a strong privacy policy is crucial to protect users' personal information. It should outline how data is collected, used, and stored, as well as provide transparency and control over data sharing. And AI systems need to have robust security measures in place to prevent unauthorized access or data breaches. This includes encryption, secure data storage, regular security audits, and implementing best practices to safeguard user data. Now, we can touch upon the challenges and concerns surrounding AI privacy and security. As AI becomes more advanced, there are concerns about potential misuse of personal data, algorithmic biases, and the need for regulations to ensure ethical and responsible AI practices. Then we can discuss the future of AI privacy policy and security. With the rapid advancements in AI technology, there is a growing need for continuous evaluation and improvement of privacy policies and security measures to keep up with emerging threats and protect user privacy

III. CONCEPTUAL STUDY

Privacy is a multifaceted concept that extends beyond the mere acquisition of personal information, encompassing a wider domain of personal boundaries and the desire for freedom from unwelcome intrusions. In our exploration within this article, our primary emphasis will be on "informational privacy." This aspect concerns the intricate dynamics of gathering, utilizing, and revealing personal data. While ongoing discussions debate the exact categories of information that fall within the realm of privacy, our focus in this article doesn't venture into the specifics of those debates. In this piece, our primary focus revolves around "privacy interests", denoting the rights or expectations individuals uphold concerning the gathering, utilization, and disclosure of their personal data. Our concern isn't merely the knowledge or possession of personal information by others; rather, it's the intentional acquisition of such information that encroaches upon an individual's privacy interests. It's crucial to differentiate between the "loss" and the "violation" of privacy. Privacy is lost when personal information is inadvertently disclosed or stumbled upon, such as when confidential data is carried away by the wind or during unforeseen incidents like individuals being exposed during an emergency evacuation. However, privacy is violated when someone intentionally obtains personal information, infringing upon an individual's privacy interests, as seen in unauthorized surveillance or data breaches. By cantering our attention on privacy interests, we underscore the significance of safeguarding individuals' rights to manage their personal data's circulation and shield themselves from unwarranted intrusions into their private spheres. When individuals talk about "privacy interests", they often refer to their wish to regulate the circulation of information about themselves and prevent unauthorized access to this data. While ongoing discussions persist regarding the most effective conceptualization of privacy (Menges, forthcoming), it's pivotal to acknowledge that one of the primary motives behind seeking control over personal information is to shield oneself from potential harm. For instance, imagine banking details being exploited by criminals to steal funds or an individual using your home address for malicious purposes.

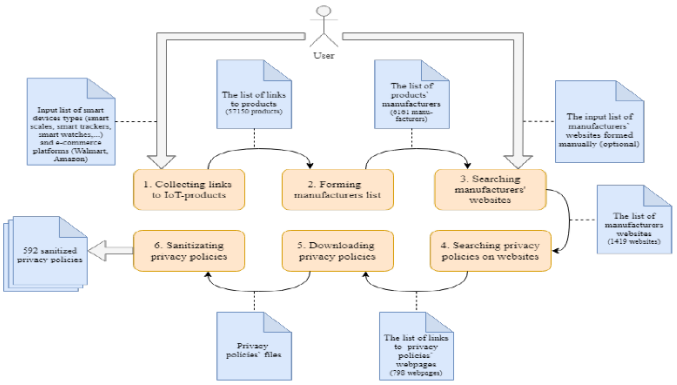

Fig2: The scheme of the dataset generation technique.

In such instances, personal information is utilized in a manner that causes harm or distress, giving rise to what's termed" security interests." Preserving these security interests is a key driver prompting individuals to manage the exposure of their personal data. privacy interests encompass the desire to regulate the use of personal information and shield oneself from potential harm resulting from its misuse. Consequently, security interests represent a vital facet within the broader realm of privacy. The rise of modern technology has fundamentally altered how individuals shape and handle their public identities. Among these technological advancements, social media platforms stand out for revolutionizing the art of constructing personas, offering individuals the unprecedented ability to engage and communicate with expansive audiences. A significant advantage of social media in persona construction is its capacity to amplify one's self-presentation. Unlike traditional methods confined within closeknit social circles, these platforms allow individuals to craft and share their personas with a global audience, greatly broadening the scope of persona-building endeavours. An illustrative case can be found in the political sphere, notably with Barack Obama's groundbreaking use of social media during his presidential campaigns. This demonstrated how these platforms can sway public perceptions and even influence voter behaviour. Subsequently, figures like Donald Trump further leveraged social media's impact, reinforcing its role in shaping political narratives and melding public sentiment. Beyond its expansive reach, the digital realm of communication offers distinctive opportunities for persona construction. The absence of face-to-face interaction grants individuals' greater control over their online personas, enabling them to selectively present or conceal facets of their identities.

Organization	Challenge	AI-powered cybersecurity to the rescue
Google	Android malware bypassing security and infecting multiple applications on Google Play Store	Google's Bouncer, an automated system, scans apps for malicious codes, gathers app data, feeds it into deep neural networks, and identifies harmful behavior
ED&F MAN	Minor server (non-data) breach highlighting the firm's low defense against crypto-mining scams	Modern security operation center implements network and endpoint detection and response platform with privileged access analytics that drills down the crypto miner threat actor in time
Apollo Hospitals	No real-time quantified view of risk posture and breach likelihood of their critical assets that store patient data	Enterprise-wide, unified, and real-time cybersecurity & digital business risk quantification platform to quantify infrastructure security risks and measure adherence to international compliance

Fig 3: AI in cybersecurity: Implementation challenges that cannot be overlooked.

This control over information flow facilitates the creation of personas that diverge significantly from an individual's offline identity. The emergence of online role-playing games (RPGs) exemplifies this phenomenon, where participants adopt virtual identities that deviate from their real-world selves. Similarly, instances of catfishing showcase how social media enables personabuilding that lacks authenticity, allowing for the creation of fabricated online identities. While social media undeniably enhances persona building capabilities, it's essential to discern that not all forms of persona construction align with ethical standards. While it can be a natural part of personal growth, persona building should never compromise authenticity or moral integrity. The potential for deceit and manipulation inherent in social media underscores the necessity for ethical considerations when shaping online identities. Modern technology, especially social media, has profoundly expanded the possibilities and scope of persona construction. The capacity to engage with vast audiences and the malleability of digital communication has introduced novel ways of self-presentation and identity management. Nevertheless, maintaining ethical standards and a commitment to authenticity should underpin any endeavours in personabuilding.

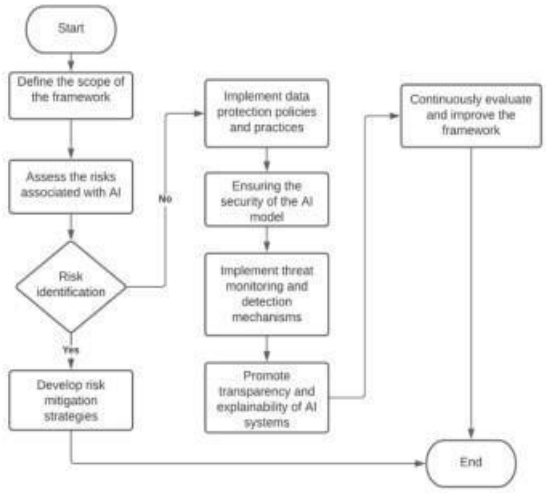

Fig. 4: Diagram showing the steps involved in creating a framework that ensures AI security and privacy.

Epistemic privilege, the inherent knowledge advantage individuals possess about themselves, forms the bedrock of privacy by allowing control over personal information and shaping others' perceptions. However, the advent of artificial intelligence (AI) and extensive surveillance practices presents a significant challenge to this privilege. Mass surveillance, described by Kevin Macnish as the automated collection and processing of people's data irrespective of their surveillance liability, is a pervasive practice conducted by both governmental bodies and private entities. Often unnoticed, this comprehensive data gathering encompasses various facets of individuals' lives. AIdriven mass surveillance far surpasses traditional surveillance methods in both scale and sophistication. Leveraging data from diverse sources like online activities, location tracking, and social media interactions, AI constructs intricate profiles that could potentially surpass individuals' self-awareness, granting these systems an informational edge over those being monitored. This imbalance in knowledge has profound implications. AI systems, equipped with extensive personal data, wield influence over how individuals are perceived and treated. They customize advertisements, predict behaviour, and even influence decisions regarding employment, financial assessments, and access to services. Consequently, individuals might struggle to retain control over their narratives and safeguard their privacy. The erosion of epistemic privilege raises

questions about privacy violations. While context matters, the undeniable fact is that AI-powered mass surveillance raises significant privacy concerns.

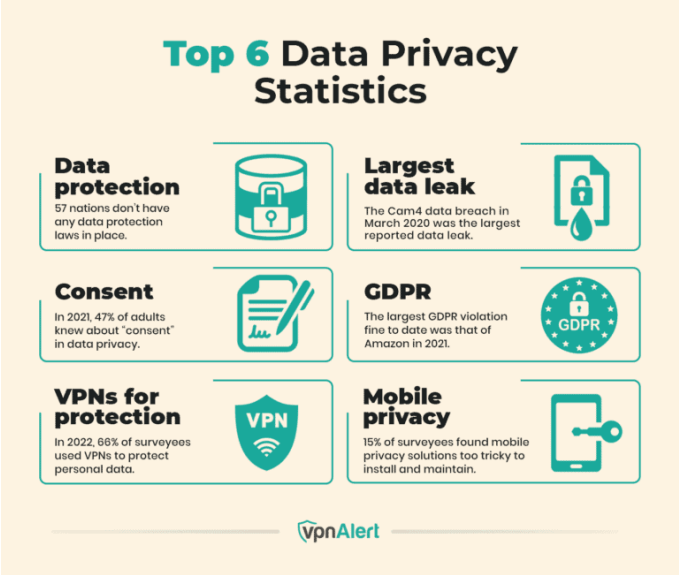

Fig 5: Data Privacy Statistics

The immense capacity of AI systems to collect and analyse personal data, often without individuals' awareness or consent, challenges the fundamental notion of personal information control. Addressing these concerns necessitates a multifaceted strategy. Implementing robust data governance frameworks, transparent data collection practices, and granting individuals meaningful control over their data are vital steps to safeguard privacy in the AI era. To enable people to make knowledgeable decisions about their online presence, it is also crucial to raise public awareness and educate the public about AI-based monitoring. The erosion of epistemic privilege due to AI-powered mass surveillance threatens individual privacy. This situation demands a comprehensive approach, emphasizing data governance, transparency, individual control, and public education to safeguard privacy in the face of evolving technological capabilities. In the sphere of personal interactions, individuals wield the power to craft their self-image by selectively disclosing or withholding personal details—a process termed "persona-building". This practice adheres to social norms that discourage unauthorized acquisition of personal information, akin to peering into someone's diary or observing them through a window. However, automated data-gathering systems seem indifferent to these norms, functioning autonomously and formulating opinions about individuals without considering their desires or choices. Consequently, these systems dominate an individual's profile construction, rendering the individual a passive participant in this process. The widespread prevalence of these automated systems often leaves individuals oblivious to the extent of data collection and analysis. They typically lack control over this data-gathering process and may even remain unaware that they are under scrutiny or being profiled. This starkly contrasts with regular social interactions where individuals possess the agency to determine when and how they present themselves, influencing others' perceptions. Although individuals retain some control over theirself-presentation in everyday interactions, it isn't absolute. Certain situations may make individuals feel unable to escape public scrutiny, despite their desire for privacy. Additionally, maintaining anonymity might be their intent, yet their presence in specific public domains could expose their identity. Moreover, individuals possess limited control over others' interpretations of the information they reveal. While they can carefully curate their self-image, they cannot guarantee how their words or actions will be perceived by others. The ascent of

automated data gathering systems has substantially reshaped the landscape of consent and control in managing personal information.

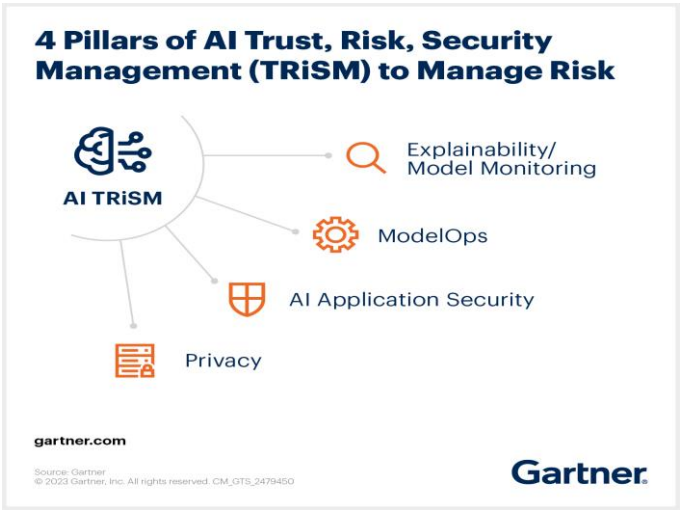

Fig 6: AI TRiSM: Tackling Trust, Risk and Security in AI Models

These systems challenge individuals' capacity to shape their narratives and safeguard privacy due to their pervasive and independent nature. Addressing these challenges necessitates a holistic strategy that emphasizes transparency, individual control, and ethical guidelines governing data collection and utilization. The ability to gather information about others has always been limited. Continuous observation and access to an individual's thoughts and feelings are practically impossible. This inherent constraint results in certain aspects of one's actions, beliefs, and inclinations remaining inaccessible to others. In essence, individuals have limitations in their ability to fully observe and understand others. Individuals have a unique advantage when it comes to self-awareness. They have direct access to their own thoughts, feelings, and motivations, granting them a deeper understanding of themselves than any external observer can possess. This epistemic privilege, the knowledge advantage individuals hold regarding themselves, is a defining characteristic of human experience. Despite this general trend, exceptions exist. Parents, for instance, may have a more profound understanding of their young children's inner lives than the children themselves. Similarly, individuals may develop close friendships where one friend possesses a deeper understanding of the other than the individual has of themselves. Nevertheless, the prevailing scenario is one where individuals hold a degree of epistemic privilege regarding their own information. This epistemic privilege empowers individuals to influence how others perceive them. By managing the information they disclose, individuals can shape perceptions in a way that aligns with their desired self-presentation. This process, known as "persona-building", entails carefully selecting and revealing information to different individuals to cultivate desired perceptions. Generally, individuals strive to present different facets of their personalities to different people. Rather than constructing a single, monolithic persona, individuals tend to build multiple personae, tailored to specific social contexts and relationships. This act of persona-building, we argue, lies at the heart of privacy per system. The ability to construct and maintain multiple personae hinges crucially on epistemic privilege. If individuals were to assume that others possess complete knowledge about them, the very concept of persona-building would become untenable. In such a scenario, individuals would lack the control necessary to manage how others perceive them. This absence of control, in our view, implies the absence of privacy as well. Epistemic privilege plays a fundamental role in shaping individual privacy. The ability to control

the flow of information about oneself and influence how others perceive them is a core aspect of what it means to have privacy. This interplay between epistemic privilege and persona building highlights the complex and multifaceted nature of privacy in human society.

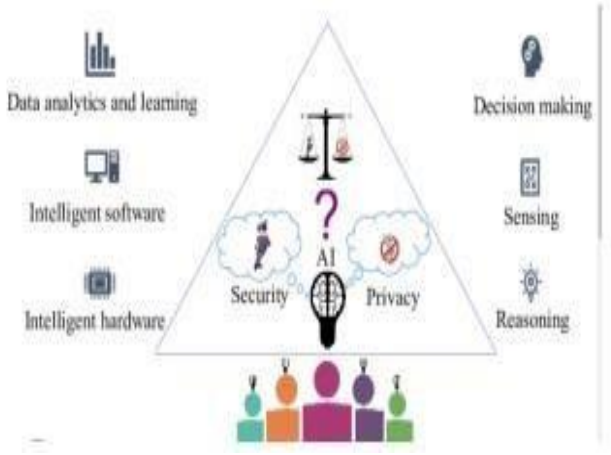

Fig. 7: An exemplary illustration of data and privacy.

IV. BENEFITS OF CHOOSING AI PRIVACY AND SECURITY

Protects Personal Information: AI privacy and security measures safeguard sensitive data, ensuring that personal information is kept confidential and protected from unauthorized access. Builds Trust: By prioritizing privacy and security, users' confidence that their data is being managed properly and securely is fostered by AI systems. Reduces Risks: Strong privacy and security protocols assist lower the likelihood of harm coming to people or organizations by minimizing the risks of data breaches, cyberattacks, and possible abuse of personal information. Safeguards Intellectual Property: AI privacy and security features protects against theft or illegal access to priceless models, algorithms, and confidential data. Maintains Reputation: People and organizations with a strong reputation for being responsible data guardians and upholding privacy rights are those that place a high priority on privacy and security. Future-Proofs Systems: By incorporating privacy and security from the start, AI systems are better prepared to adapt to evolving threats and regulatory requirements, ensuring long-term viability.

V. STATISTICAL ANALYSIS BETWEEN PRIVACY AND SECURITY

Privacy and security in AI are closely intertwined. Statistical analysis can help identify patterns and trends in data breaches, privacy violations, and security vulnerabilities within AI systems. By analysing data related to privacy incidents and security breaches, statistical techniques can provide insights into the frequency, severity, and impact of these incidents in AI applications. Statistical analysis can also be used to assess the effectiveness of privacy-enhancing technologies and security measures implemented in AI systems, helping to identify areas for improvement. Through statistical analysis, researchers can quantify the relationship between privacy and security in AI, exploring how changes in one aspect may impact the other and informing the development of robust privacy and security frameworks. statistical analysis can assist in evaluating the compliance of AI systems with privacy regulations and security standards, providing quantitative evidence to support policymaking and governance efforts. Although AI systems might seem self-contained, there are instances where individuals can perceive how their actions affect the profiles generated by these systems. For example, people can track changes in their credit scores to assess how their financial behaviours are reflected. This loop of feedback indicates some level of interaction between individuals and AI systems, granting individuals a degree of influence over their profiles.

However, the question lingers regarding whether this feedback loop substantially impacts the development of independent selves. Some argue that mass surveillance, by subtly shaping behaviour, can hold individuals into forms desired by those controlling the technology (Cohen, 2013). If valid, this implies that AI does contribute to shaping moral agency, albeit in a manner that doesn't necessarily foster autonomy. It's crucial to distinguish that while AI might impact autonomy, not every instance of autonomy necessarily aligns with privacy concerns.

Although privacy is vital for independent growth, instances where actions are driven by personal convenience, rather than a desire for a particular perception, may not fall within the realm of privacy, as we define it. Additionally, the feedback loop facilitated by AI likely focuses on specific facets of an individual's identity, often less significant ones. For instance, someone may argue that their fundamental identity remains unchanged regardless of their credit score. This limited influence implies that AI's impact on autonomous development is confined compared to the profound effect of human interactions in shaping our understanding of ourselves and our values. To sum up, while AI systems may feature a feedback loop, their effect on autonomous development seems limited and distinct from privacy concerns. AI's influence appears to be restrained to certain identity aspects and might not significantly shape an individual's core self or moral agency. Human interactions remain the primary influencers in fostering self-understanding and autonomous growth.

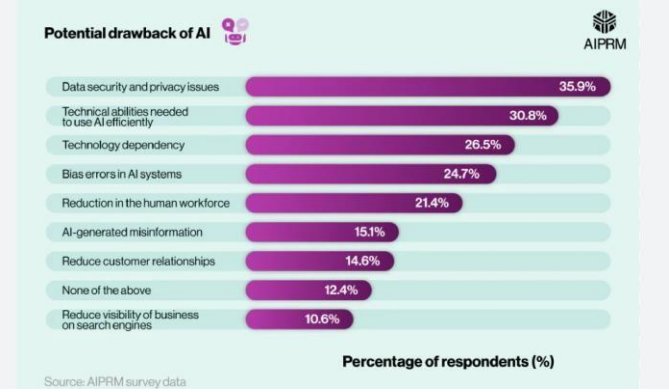

Fig 7: AI Statistics 2024

VI. CONCLUSION

The emergence of Artificial Intelligence (AI) has triggered ongoing debates about its potential influence on privacy. Although AI undoubtedly affects privacy in diverse ways, the actual nature of this influence may not always align with common perceptions. One of AI's fundamental challenges involves its potential to disrupt the 'epistemic privilege,' typically enjoyed by individuals concerning their own information. This privilege assumes that individuals have the most direct and accurate access to their personal data. However, AI systems, capable of gathering, analysing, and inferring insights from extensive data, can disrupt this traditional balance of knowledge. Another challenge arises from AI's ability to challenge the control individuals usually wield over their personal information during interpersonal interactions. Without their awareness or express approval, AI systems are able to gather and analyse personal data about people, which may limit their ability to control how this data is used and perceived. It's critical to distinguish security problems from privacy issues. While security interests are concerned with protecting people from harm caused by illegal access to or exploitation of their personal data, privacy concerns are related to an individual's right to manage how others see them. AI systems don't fundamentally threaten privacy, even if they could present threats to security interests. This is due to the fact that AI

systems cannot, by themselves, develop the types of perceptions that can tamper with a person's sense of who they are or how others see them. ting their capacity to manage the use and perception of this information. It's crucial to differentiate between privacy itself and security concerns. Privacy pertains to an individual's right to control how others perceive them, while security interests focus on shielding individuals from harm due to unauthorized access or misuse of their personal data. While AI systems may pose risks to security interests, they don't inherently challenge privacy itself. This is because AI systems, in isolation, cannot form the kinds of perceptions that could interfere with an individual's self-perception or how others view them. The abundance of personal data within AI systems does heighten the risk of privacy breaches, especially if this data is accessed by entities capable of forming perceptions based on that information. In such cases, AI systems can indirectly impact privacy by enabling or facilitating privacy violations. Recognizing that AI systems pose risks to both privacy itself and security interests is crucial. Even though they might not directly threaten privacy itself, the interconnected nature of these concepts necessitates considering both when assessing AI's impact on individual privacy. The capacity to attribute meaning and form perceptions is not a prerequisite for posing a threat to security interests. Entities lacking semantic understanding can access and misuse personal data gathered by AI systems, potentially causing harm or exploitation. while privacy itself might not be directly threatened by AI systems, safeguarding security interests and preventing unauthorized access or misuse of personal data remains crucial. Achieving this requires robust data governance frameworks, adherence to privacy by design principles, and responsible development practices in AI. AI's impact on privacy is intricate and multifaceted. Although AI systems can pose threats to security interests, they do not directly challenge privacy itself. However, understanding the intricacies of these concepts is pivotal for creating effective policies and designing AI technologies that respect and safeguard individual privacy rights

VII. REFERENCES

[1] . H. Saif, T. Dickinson, L. Kastler, M. ernandez, and H. Alani, "A semantic graph-based approach for radicalisation detection on social media," in Proc. Extended Semantic Web Conf. (ESWC 2017), pp. 571- 587

[2] X. Luo, R. Shen, J. Hu, J. Deng, L. Hu, and Q. Guan, "A deep convolution neural network model for vehicle recognition and face recognition," Procedia Comput. Sci., vol. 107, no. C, pp. 715-720, Apr. 2017. doi: 10.1016/j .procs.2017.03.153.

[3] S. Ghosh, A. Das, P. Porras, V. Yegneswaran, and A. Gehani, "Automated categorization of onion sites for analyzing the dark web ecosystem," in Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, New York, 2017, pp. 1793-1802.

[4] P. Vitorino, S. Avila, M. Perez, and A. Rocha, "Leveraging deep neural networks to fight child pornography in the age of social media," J. Vis. Commun. Image Represent., vol. 50, pp. 303-313, Jan. 2018. doi: 10.1016/j. jvcir.2017.12.005.

[5] H. Bouma et al., "Automatic analysis of online image data for law enforcement agencies by concept detection and instance search," in Proc. SPIE 10441, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies, 2017. doi: 10.1117/12.2277970.

[6] Z. Lipton, "The mythos of model interpretability," Commun. ACM, vol. 61, no. 10, pp. 36-43, 2018.

[7] S. Raaijmakers, M. Sappelli, and W. Kraaij, "Investigating the interpretability of hidden layers in deep text mining," in Proc. 13th Int. Conf. Semantic Systems, Amsterdam, The Netherland s, 2017, pp. 177-180 .

[8] N. T. Lee, “Detecting racial bias in algorithms and machine learning,” J. Inform., Commun. Ethics Soc., vol. 16, no. 3, pp. 252–260, 2018.

[9] S. Sabour, N. Frosst, and G. Hinton, “Dynamic routing between capsules,” in Proc. 31st Conf. Neural Information Processing Systems (NIPS), 2017, pp. 3856– 3866. [10]S. Corbett-Davies and S. Goel, The measure and mismeasure of fairness: A critical review of fair machine learning. 2018. [Online].

[11] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A survey of methods for explaining black box models,” ACM Comput. Surv., vol. 51, no. 5, 2018. doi: 10.1145/3236009. 12. F. C. Keil, “Explanation and understanding,” Annu. Rev. Psychol., vol. 57, no. 57, pp. 227– 254, 2006.

[12] https://images.app.goo.gl/drnAwwBhRtzPjHBY6

[13] https://images.app.goo.gl/muhsHcwc8HhToX387

[14] https://images.app.goo.gl/YxbTunNvsZnLxZo68

[15] https://images.app.goo.gl/1WiJTcWYdRdEU6JP7