Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier

Preprint · March 2025

DOI: 10.13140/RG.2.2.19510.61760

5 authors, including:

Deepak Puthal
Newcastle University
221 PUBLICATIONS **7,847** CITATIONS
SEE PROFILE

Amit kumar Mishra
Banaras Hindu University
204 PUBLICATIONS **816** CITATIONS
SEE PROFILE

Saraju P. Mohanty
University of North Texas
652 PUBLICATIONS **13,462** CITATIONS
SEE PROFILE

Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier

Deepak Puthal[1], Amit Kumar Mishra[2], Saraju P. Mohanty[3], Antonella Longo[4], and Chan Yeob Yeun[5]

[1] Indian Institute of Management Bodh Gaya, India
[2] Aberystwyth University, UK
[3] University of North Texas, Denton, USA
[4] University of Salento, Italy
[5] Khalifa University, Abu Dhabi, UAE
Email: deepak.puthal@ieee.org, amm89@aber.ac.uk, Saraju.Mohanty@unt.edu, antonella.longo@unisalento.it, chan.yeun@ku.ac.ae

Abstract

The progression of artificial intelligence (AI) technologies has reached a level that greatly enhances the different organizational sectors by facilitating them with the means to advance and improve systems and processes. Shadow AI implies the usage of AI tools and systems by individuals within an entity, respectively, without permission thereby implying that these tools were not directly monitored or controlled by the centralized IT or security department. It also contributes to significant cyber risks such as data and security breaches, abuse of compliance, and, in general, an increased threat landscape. This paper highlights into the emerging global security trends and Shadow AI while also covering the unique positioning within the threat landscape concerning unauthorized computation of sensitive data, safety vulnerabilities of the unmonitored AI models, and model poisoning alongside data leakage-marked out. Moreover, this paper covers how Shadow AI distracts the attack landscape while increasing the level of security problem for the organization. Shadow AI, however, can be employed to increase the ability to respond to threats, locate irregularities, and increase the range of options available for cyber solutions even with all its risks.

Keywords: Artificial Intelligence, Shadow AI, Cyber security, Federated Learning, Explainable AI

1. Introduction

The explosive expansion of artificial intelligence (AI) technologies across different sectors has led to a major shift in workflows, decision-making, and innovativeness. However, this extensive penetration has also resulted in the development of “Shadow AI” [1]. Such deployment and usage of AI tools, models, or systems by employees of an organization without explicit endorsement, supervision, or management by IT and cyber security departments is defined as Shadow AI [2]. These unregulated implementations of AI systems often occur as a result of employees or teams exploiting easy to acquire AI solutions to tackle particular problems, automate processes, or enhance their market standing without permission. Despite these endeavors being made in good faith, they are outside of the normal governing frameworks, making gaps and complexities within the organization’s structure. Shadow AI spans across multiple domains, such as marketers who use AI-based analytics tools and

recruiters who use AI-based recruitment tools [3]. However, this so-called shadow or informal use is often undertaken in the absence of adequate security, accountability, and transparency measures, exposing organizations to the risks of data breaches, regulatory non-compliance, and malicious threats.

When it comes to cyber security, Shadow AI is responsible for a numerous security concerns. The reliance on interconnected systems as a digital infrastructure for any organization that uses automation, data processing and predictive analytics even if a single AI is on grossly increases the number of interfaces for adversaries to exploit [3]. Shadow AI uses sensitive data without permission and operates from a grey area, raising ethical concerns. Looking from a purely technical standpoint, Shadow AI only exasperates the compliance landscape, making it incredibly difficult to uphold data protection laws and SAS. Shadow AI, being unsanctioned also greatly complicates the regulation of the data being used [4].

Analyzing Shadow AI as a security problem in relation to the protection of digital infrastructures and cyber security implies overcoming some problems that are faced in modern cyber landscapes [5]. As enterprise networks and AI systems grow more complex, failing to account for Shadow AI endangers the overall security architecture of the organizations. By looking at its sources, dangers, and possibilities, businesses can devise plans that prioritize having a strong security governance framework while also reaping the advantages of innovation. Involves building synergies between IT, security, and the users, so the organization can leverage Shadow AI without exposing it to undue risk [6]. Besides, assessing the implications of Shadow AI helps understand the architecture of future oriented secure AI frameworks considering compliance, trust and resilience within the digital infrastructure.

This paper aims to illuminate the dual nature of Shadow AI, i.e., its capacity to drive innovation and its potential to introduce vulnerabilities, and provide actionable recommendations for organizations to address the challenges it poses.

2. The Rise of Shadow AI

The rise of Shadow AI highlights both the opportunities and risks posed by decentralized innovation in the realm of AI.

2.1 Reasons for the Shadow AI Explosion

A number of factors can explain the rapid growth of Shadow AI. First, there is the enabling policy: AI tools have been made widely available to the general public. With the standardization of AI technologies and the availability of platforms, frameworks, and API's, any knowledgeable employee can deploy and utilize AI solutions on their own [4]. The presence of low-code and no-code platforms further dissipates the entry maelstrom, permitting non-technical staff to test and apply AI into their workflows without the benefits of a central IT department [6]. Similarly, cloud-based AI services and open-source machine learning frameworks like TensorFlow and PyTorch have enabled people to quickly design and implement AI solutions without the required governance.

One more critical metric is the emergence of employee-driven innovation. There are numerous ways through which employees have a strong desire to perform, including attempting to make use of AI tools to increase productivity, or automate everyday tasks [7]. Furthermore, this innovation generates within organizations improves how efficient and creative they are. However, this shadow volition

may contribute to the emergence of Shadow AI when undertaken without regard to the organization's policies and security [6].

In many organizations, employees and departments are under pressure to meet tight deadlines and compete. If IT departments are unable to fulfil these expectations because of a lack of people or because of internal politics, employees have no choice but to seek out AI tools that are available [8]. However, this flexibility comes at the price of control since these technologies are frequently adopted without sufficient guidance, which results in risks.

2.2 Instances of Shadow AI Usage in Different Sectors

Shadow AI refers to the use of informal shadow artificial intelligence applications and this rapidly climbing trend has been observed in almost all of industry sectors due to the specific challenges and opportunities existing in every type of business. For instance, in the healthcare sector [9], unauthorized AI tools are employed by clinicians and researchers alike to aid in evaluating patient data, predicting potential diseases, or assisting in patient diagnosis. AI application tools themselves also raise issues with data privacy, because compliance with regulations, such as HIPAA, and blindly relying on unregulated algorithms cannot be practical or up to date.

Within the financial industry, shadow AI is more and more often used in data processing and risk management. Employees may adopt unapproved AI-powered tools for fraud detection, credit scoring, or investment analysis [8]. The underlying problem, however, is that the incorporation of these features has a downside: the usage of AI tools does not adhere to any organizational or regulatory standards and provides various risks, such as inaccurate predictions or even breaches to the security system of the organization.

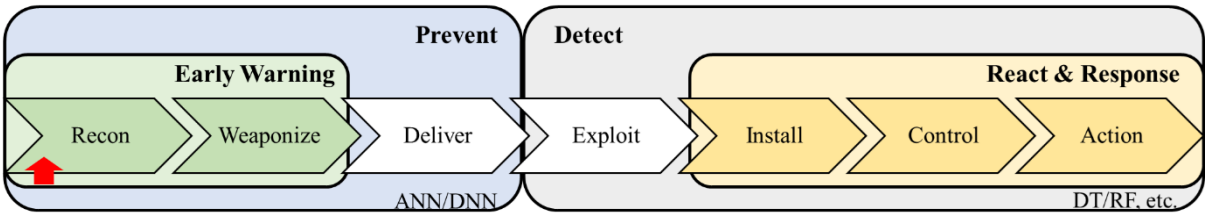
Moreover, AI tools such as grading tools and student engagement applications are being used by teachers and admins in colleges and many educational institutions without any permissions from the IT departments. If integrated with care, these applications can greatly aid in improving one’s learning experience, but they come at the cost of potentially having access to personal information or having a faulty decision making system.

2.3 Overview of Its Dual Nature: Innovation Vs Risks

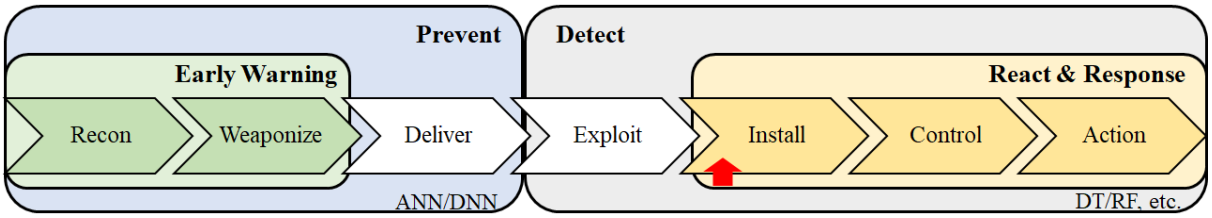
The dual nature of Shadow AI’s is the most striking feature, both the dark and light sides. Shadow AI allows employees to develop and use specialised AI solutions for particular problems quickly and easily. It strengthens motivation, promotes performance and provides new avenues for business development. When employees are encouraged to take initiative, it results in diversity in their approaches to solving the problem.

In contrast, Shadow AI tools also have the downfall of having no surveillance as a result they pose great danger. Their existence in the organization makes the organization vulnerable to breaches of privacy, security and compliance as they are unauthorized to supervise tools [8]. A case in point is the unregulated mass of sensitive data that organizations use because they rely on AI tools that lack necessary safety measures, which makes the mass of data appealing to hackers or even thieves. Moreover, because concerns about accountability and oversight were not raised before deployment, the algorithms may produce incorrect outputs that may even be biased because of the lack of quality checks implemented, which affects the decision-making process and taints one’s reputation [10].

Furthermore, the lack of transparency with Shadow AI makes it difficult to manage other risks. The IT departments cannot locate the tools and since there is no prior knowledge of their presence trying to address the vulnerabilities becomes fruitless [11]. This gap breeds the waste of resources and dulls one's defense to future attacks. A system diagram and the comparison of normal AI tools and Shadow AI use in an organization is as shown in Figure 1 [12]. In the Figure 1(b), the concept of Shadow AI entertain the user to use the software without the knowledge of internal process and data usage. Where the red arrow mark is highlighting involvement of user process. Whereas, as shown in Figure 1 (a), the user is aware of the processes from the beginning and this will not lead to the missus of the data.



(a) Using an authorised AI module in an organization; the processes are aware to the organization from the arrow mark.



(b) Using a Shadow AI module in an organization; the processes are aware to the organization from the arrow mark.

Figure 1. Process of using AI and Shadow AI use in an organization.

3. Cyber Security Implications of Shadow AI

The fact that applications of Shadow AI are on the increase in different sectors of the economy creates a new and major cyber risk. While Shadow AI allows employees and departments to come up with creative ideas, it equally works out of the control of organizations, which exposes them to potential exploitation by cybercriminals. Therefore, it is vital for organizations to grasp these cyber security implications in attempts to ensure that they encourage innovation but build robust security defences [11]. The threat landscape introduced by Shadow AI is more thoroughly explained below.

3.1 Unauthorized AI Tools Processing Sensitive Data

The use of Shadow AI without authorization, which also happens to be badly controlled by IT measures poses one of the greatest threats when it comes to cyber security management. For one think, when employees go ahead and implement AI tools that are not approved, they frequently do so without having strict security measures that are essential for the protection of the sensitive organizational data [7]. Therefore, to illustrate, for instance, employees that do not pay attention to the policies and practices of data storage and handling that a particular platform comes with might begin using the tools that allow them to analyze customer data and financial records. Such a situation without properly being transparent and controlled worsens the potential of leaking data, or gaining access without authorization or even data manipulation by malicious individuals [11].

The use of unregulated AI tools could render organizations vulnerable to legal and financial repercussions as such tools may break data protection laws like the General Data Protection

Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), or Digital Personal Data Protection (DPDP Act), 2023 that have not been approved yet.

3.2 Potential Entry Points for Cyberattacks

Shadow AI tools tend to broaden the cyber attack base for organizations. Also, unauthorized AI tools rarely have security patches and updates put in place, making them easier computer systems to breach. For instance, Shadow AI tools have security vulnerabilities such as malware deployment, unauthorized access and sensitive information compromises [10]. Moreover, the use of employees' personal devices for Shadow AI as well as the use of unsecure cloud environments can serve as platforms for phishing, man-in-the-middle, and ransomware attacks [13]. Cybercriminals can capitalize on these vulnerabilities that can be used to infiltrate the organization's broader IT structure, causing serious disruption in the organization.

3.3 Breaches in Compliance and Regulatory Standard

The use of Shadow AI tools usually creates breaches of compliance and regulatory standards which is a matter of great concern for many organizations. The regulatory policies including GDPR, HIPAA, DPDP Act, and the California Consumer Privacy Act (CCPA), expect a level of control on the manner in which personal data is gathered, processed, and stored [14][15]. Whenever Such AI tools are deployed outside the approved and authorized limits and authority, these controls are normally disregarded which contravenes the set standards. An example is an unauthorized employee who employs a data analytics AI powered tool and transfers sensitive or proprietary information to servers in regions that do not have sufficient data protection laws, thereby trespassing the rule. Such compliance failures usually result in heavy penalties, damage to the organization's image and a loss of confidence from clients and business partners.

3.4 Examples of Shadow AI Security Incidents

Real-world examples illustrate the tangible cyber security risks associated with Shadow AI.

1. **Healthcare Data Breach via Unauthorized AI Tools:** Let us assume an instance of a hospital employee who engaged in conducting research by applying an unapproved AI based programme to sift through medical records. The application had poor encryption measures in place along with absence of developed access controls. Applying several countermeasures the cybercriminals were able to compromise with the sensitive patient's information such as their personal details and medical history saved in the third party cloud based platform [16]. The incident is a violation of the HIPAA law and mechanism and moreover it brings about legal repercussions for the hospital management as well as loss of trust from the patients. The medical history of a person is exceedingly sensitive information [17].
2. **Financial Sector Vulnerabilities:** An analysis of a financial services provider rendered that a group of analysts had incorporated their use of a third-party AI powered fraud detection system unlocking it to the active segments of the IT department. The transactions that were identified as interceded frauds were highly effective during these interventions. However, the data collected was transferred to a cloud based system later, which was then attacked cybernetically by fraudsters. The fraudsters were able to access sensitive customer-related information and proprietary algorithms which resulted in the business incurring losses [17]. Reputation and finances were the two main areas in which the company suffered enormous loss. This incidence helped Identify the significant impacts and vulnerabilities faced by the company.

3. **Shadow AI in Retail Supply Chain:** In the region for selling goods and services, a supply chain manager used an AI-based inventory optimization tool with a view of overcoming inefficiencies that had been evident in inventory management. Unfortunately, that tool required the organization to provide access to its supply chain management system and it worked over an unsecured pathway. Cybercriminals used this flaw as an avenue to introduce malicious code into the system which corrupted certain functionalities thereby temporarily crippling the entire company's logistics network [18]. This incident highlights how Shadow AI can open critical gaps for an organization's operational infrastructure.

4. Shadow AI and the attack surface expansion

Shadow AI, for instance, refers to the use of unregulated or unapproved AI tools and systems into a specific company, in most cases, by employees or teams who are outside the coverage zone of the IT departments. Although these tools and programs are intended to increase efficiency, help solve certain problems or stimulate creativity, the fact that they are unsanctioned makes it possible for one reason or another for the adversaries to be able to exploit these weaknesses. As Shadow AI grows, it is necessary to investigate its role in the expansion of the attack surface and its risks to the organizational assets.

4.1 Complexity of the Attack Surface

Within the scope of cyber security, Shadow AI tends to add a layer of novel complications, particularly concerning the attack surface of an organization. The attack consists of all the possible vulnerabilities alongside entry points that an attacker can make use of shadow AI and such precise vulnerabilities are already difficult enough to deal with using the traditional methods [5]. These vulnerabilities include the use of endpoints, servers, networks, and applications. Shadow AI tools, however, make it harder since they add a fresh level of unrevealed systems to the mix, ones that the IT team aren't even aware of. Most employees using AI based tools while analyzing data, automating processes or while making business decisions tend to overlook the security standard process resulting in offering casual exposure to key data and information without the required protocols being utilized [10].

Furthermore, the other dimension of Shadow AI exacerbates the situation owing to its diverse and distributed nature. These kinds of tools often get added to a cloud service, edge devices, and in other personal devices without being looked into properly, blindfolding what the peripheral security of the organization can handle [6]. It is for this reason, that implanting structures to monitor device changes proves to be of no use since there are endless possibilities for such devices to alter. And due to this altered reality, aiming to assess the plethora of vulnerabilities in a timely manner crosses the fine line of difficulty into the impossible realm [5]. With the help of Shadow AI, mechanisms on how to deliver protection on critical structures alongside cyber systems are drawn out and made even harder previously protected mechanisms.

4.2 Unmonitored AI Tools and Vulnerabilities

4.2.1 Model Poisoning

Finally yet importantly, model poisoning is also a distinct possibility while using unmonitored AI tools. Undoubtedly, unmonitored tools are prone to all manner of weaknesses. For example, an AI can be taught using data that has been strategically corrupted in what is called a model poisoning attack. To illustrate, consider the case of a Shadow AI, in which an employee uses an external AI

model for predictive analytics without checking its security credentials [19]. In such cases, hackers may have gained access to the model and altered it so that the predictions are not accurate or may alter the functionality of underlying mechanisms, such as granting unauthorized resources or classifying data incorrectly. The straightforward explanation is that since there is no meaningful oversight and auditing of Shadow AI applications, the tools are easy to abuse.

4.2.2 Data Leakage

While Shadow AI has many benefits, it also poses serious challenges that need to be addressed. One of the most serious issues with Shadow AI is data leakage. A typical Shadow AI tool tends to make use of external cloud services for data storage and processing. Any employees using these tools may inadvertently transfer sensitive organization data to third-party servers with no regard for security protocols. For instance, a worker might employ a productivity enhancing AI system that transfers sensitive customer data to a cloud application [19]. If the application is breached, or is not compliant with data security rules and regulations, then unauthorized access, breaches, and possible damage to reputation and legal suits follow.

Additionally, Shadow AI tools tend to be more prone to interception due to their vulnerability of not having encryption or secure APIs, and sometimes have no authentication measures in place at all. The Table 1 gives a multidimensional taxonomy of the attacks, impacts and vulnerabilities on Shadow AI.

Table 1: Taxonomy of Attacks on Shadow AI

Category	Attack Type	Impact	Vulnerability
Data Integrity Attacks	Data Poisoning	Model produces incorrect predictions	Lack of data validation in Shadow AI
	Adversarial Attacks	AI misclassifies malicious inputs	Weak adversarial robustness
Privacy Attacks	Model Inversion	Extraction of sensitive training data	Lack of encryption in AI deployment
	Membership Inference	Determining if data was used for training	Shadow AI model lacks privacy safeguards
Evasion Attacks	Input Manipulation	Bypassing AI-based security defenses	Lack of secure input validation
Access Exploits	Unauthorized API Usage	Uncontrolled AI model interactions	Shadow AI models calling unprotected APIs
	Privilege Escalation	Gaining unauthorized access to AI features	Lack of authentication in Shadow AI
Regulatory and Compliance Risks	Data Exfiltration	Leakage of sensitive business data	AI tools transferring data to external servers
	Compliance Violation	Legal consequences and penalties	Shadow AI bypassing security policies

4.3 Examples of Vulnerabilities Introduced by Shadow AI

1. **Unsecured API Endpoints:** Shadow AI tools for example most of the times use API's for transferring data, but these API's endpoints sometimes do not have security. Attackers take advantage of these weaknesses to insert harmful codes, disrupt the systems, and collect private and sensitive data.

2. **Overprivileged Access:** In order to complete a certain task, a Shadow AI system is usually granted wide access to its organizational data or network. Such excessive permissions could act as an entry point to the entire IT infrastructure and could be abused by attackers for wrongdoing.
3. **Insecure Third-Party Integrations:** Several Shadow AI tools require the interfacing with third party applications, distribution systems or even SaaS systems. Most of these integrations operate outside the purview external security screenings, thereby creating gaps in the security of the organization [13].

5. Opportunities for Cyber Security with Shadow AI

The presence of Shadow AI essentially creates risks, however being an inaccurate AI there is a plethora of opportunities to strengthen the cyber security strategies. Organizations can maximize the effectiveness of Shadow AIs by incorporating it into their structure. Through this, the detection and response to potential threats enhances. As a potential risk, Shadow AI must be managed in such a way that it alters the insight of the organization's security systems.

5.1 Shadow AI for Faster Threat Detection and Response

The detection of threats and their subsequent response is one of the areas that Shadow AI has the ability to optimally exploit. As for newer Shadow AI applications, they can address this issue more efficiently since they function independently and deploy pattern recognition, anomaly detection, and other algorithms as well in real time [19]. By using this already available technology, the organizations can utilize Shadow AI to outpace the obsolete systems in deploying responses [3].

As an illustration, Shadow AI tools can greatly improve anomaly detection by spotting even slight deviations in user behavior or network traffic – this is useful due to its link with underground activities like advanced persistent attacks and insider threats. On the other hand, there are pre-validated AI systems, and due to their bureaucratic regulations, they take time to be deployed; unlike Shadow AI tools which are at the forefront of technology development and allow an organization to play with new threat detection methods [3]. In addition, Shadow AI tools can also assist in certain parts of the incident response process like flagging suspicious activities for investigation, or isolating vulnerable or compromised systems, which decreases the time of exposure and increases the toughness of the system against cyber threats.

5.2 Employee Enabled AI Enhancements for Security System Management.

Shadow AI in most cases is sourced from employee facilitated innovation whereby in an enterprise, individual or teams come up with AI tools to fix specific engagement. While these tools are generally developed to enhance productivity or to address business needs, there are also ways of utilizing them to develop security systems for the organization [5]. Employees, for instance, are much closer to the operational needs of the institution and posses unique understanding of its problems and hence the development of specialized AI-powered tools for close cyber security needs can also be achieved.

An employee could for example use an AI tool to streamline workflows with respect to data but such tools could be used to flag history of data access breaches. Employee-initiated AI models which are built to assume operational inefficiencies can also be used to assume flaws in system configurations [2]. If enterprise cares to surround employees with sufficient encouragement and focus strategy of developing security related embedded goals, then security systems which are agile and context aware could be readily available.

AI security solutions based on Shadow AI can also take advantage of self-learning. As many of the Shadow AI tools are built with flexibility in mind, their retraining on new data in an effort to meet new threats means that security measures remain competent in fast-changing spaces.

5.3. Decentralized Secure AI Governance Frameworks

In such a decentralized governance model, the challenge would be to design and put in place a governance framework that allows Shadow AI tools to be registered, monitored as well as validated without hindering innovation. For example, a safe "AI sandbox" space could be instituted by organizations where employees may experiment with their tools and channels in a safe space that is still controlled. These environments could provide some degree of automated vetting to the tools for checking for security holes, regulatory compliance and company policies [7].

Furthermore, Shadow AI activity audit trails can be provided in a more transparent form using blockchain technology in decentralized governance systems. Once an organization is enlisted to use Shadow AI tools, it will use and modify these tools on an appropriate blockchain to record their use, which increases accountability of the organization while reducing the chances of misuse [20]. In the same way, secure governance frameworks can incorporate federated learning techniques allowing the Shadow AI tools to interact and exchange information across decentralized networks without disclosing sensitive information [8].

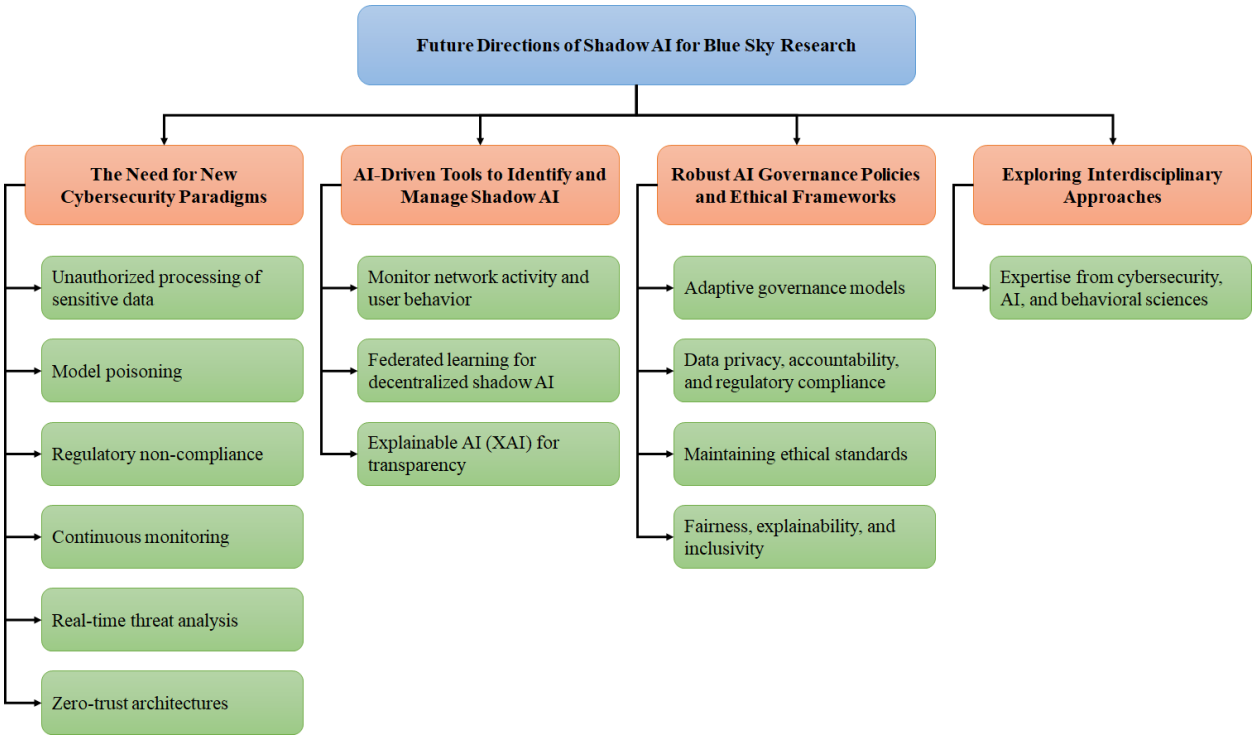


Figure 2: Future Directions of Shadow AI for Blue Sky Research

6. Future Directions

Every organization is trending towards the incorporation of AI, and with the rise of unsupervised and unregulated Shadow AI tools the need for a multi-faced strategy to encourage innovation while mitigating risk, is prevalent. This insurmountable snow of unregulated AI forces us in this direction. Given this, developing a supply of Shadow AI strategies. This section outlines the future directions for blue-sky research in addressing Shadow AI, outlining the governance policies, technology

integration, and interdisciplinary cooperation that are essential to address these challenges. The summary of future directions of Shadow AI as shown in the Figure 2.

6.1 The Need for New Cyber Security Paradigms

Given the growing importance of shadow AI across a variety of sectors, paradigms that have served the cyber security environment previously are being pressured to be reviewed. Considering how decentralized Shadow AI is, it makes it impossible to always actively looking for pre-set threats and most contemporary frameworks do place empires on a sole location. Instead, novel approaches for the future should seek to accommodate the wide range of potential non compliancy risks shadow AI brings with it – such as data processing without authority, model pollution, and non-disconformity to rules and regulations and still empower the core vision of growth [10].

An appealing area for consideration is a construction of adaptive cyber security systems enabling to cope with the various and changing risks relating to Shadow AI. An adaptive system would utilize Shadow AI tools in continuous supervised monitoring while forming real-time intrusion analytics functions. In addition, organizations will have to configure proactive resilience solutions, such as zero-trust architectures that are focused on detecting and eliminating the threats that originate from the use of unregistered AI tools. Furthermore, this type of research should also look into the extent to which insecurities – especially the brittle ones that could potentially discourage employee's creation, should be imposed. Several applications, including autonomous vehicles and drones, the users are directly using the AI tools [21][22], those required the user privacy.

6.2 AI-Driven Tools to Identify and Manage Shadow AI

Over the years, organizations have had to contend with multiple tools they either build or purchase. These tools that fall outside of a user's control are called shadow AI. As such, organizations need to be able to identify those tools and stop their usage. Inherent technology fluctuations have made it necessary to find these shadow AIs unnoticed. Furthermore, there is a need for strong insights in regards to the potential risk of providing a safe joint interface or finding a safe way to restrain the fusion. Such a heavy shift more than likely needs an AI shift that involves observing user behavior and tracking network activity [8].

Additional machine learning techniques for AI detection can review metadata attached to internal AI tool use and identify unauthorized shadow AI's use, which can help reduce gaps in security. Additionally, AI detection is not limited to metadata, for instance, NLP algorithms could scan any AI related references in documents or other communication limits that indicate employee talks about sharing shadow AIs or other tools. However, these AI detection tools need to be integrated with automated risk assessment tools that help determine the risk associated with the compliance and security of working with an AI shadow tool.

Additional AI solutions of the future will require taking better precautions while utilizing Shadow AI tools as well. As an example, federated learning could assist organizations by not needing to centralize all their data [23], enabling them to set and impose security policies over decentralized Shadow AI tools, thus aiding in maintaining security alongside operational capabilities. Additionally, explainable artificial intelligence (XAI) methods [24] can be utilized for showcasing how Shadow AI tools are functioning, thus aiding to set more informed strategies regarding their adoption or mitigation.

6.3 Creating Robust AI Governance Policies and Ethical Frameworks

The uncontrolled use of Shadow AI, in turn, demonstrates how specific policies and ethical standards for governance of these tools are to be developed. Future direction of research should be to develop such governance models that are not limited to IT compliance but also takes into consideration AI performance, data ethics, and risk management as well.

The governance framework of the Shadow AI includes as top issues: data privacy, accountability, and compliance with regulatory requirements. This means setting policies on what is considered the acceptable application of AI in companies and introducing appropriate checks to enforce compliance. For example, decentralized policy enforcement frameworks can be used in organizations where employees are allowed to invent while complying with certain conditions for the AI tools developed.

Strategies for preventing the use of Shadow AI, whether due to bias in AI decision making tech or unauthorized use of surveillance systems, also need inclusion within the ethical frameworks put in place. Future research will elaborate on the methods that can facilitate fairness, explainability, and inclusiveness in the evolving sphere of AI as a whole. Such multi-faceted frameworks will assist in risk minimization that comes with Shodow AI while assisting organizations better become more innovative with minimum responsible risk tied with it.

6.4 Exploring Interdisciplinary Approaches

Reconstructing technical frameworks to managing the use Shadow AI cannot solely be used as a strategy for addressing the distinct challenges posed by it. New researches that aim to address the issue of Shadow AI need to incorporate perspectives from a variety of fields such as AI, Cyber Security and Behavioral Sciences [25].

As an example, behavioral sciences can shed some light on why employees resort to Shadow AI tools and how this behavior in influenced by the culture within the organization. These behavioral aspects can be tackled in organizations through targeted approaches such as providing tools which discourage the use of Shadow AI tools or promoting the use of solely AI authorized ones.

In the same manner, interdisciplinary specialists look into any potential solutions that cyber security and AI could present while looking into the problems posed by Shadow AI. For example, machine learning algorithms can be blended with cyber security to create an AI model that is not vulnerable or easily tampered with. In addition, working with policy and legal experts can help ensure that solutions conform to changing regulations allowing for the unified approach to Shadow AI management within any industry.

7. Conclusion

The development of Shadow AI has both risks and rewards; it transforms the novel landscape of AI integration, development, and cyber protection in organizations. Through this paper, we have investigated into the complex repercussions of Shadow AI, focusing on its effects on information security, attack surface expansion, as well as management and governance opportunities. It also features one of the most prominent problems regarding the Shadow AI phenomenon: it has this two-sided character. Firstly, it allows coping with issues faster and tie-up the employees to the AI, which makes sophisticated AI tools available to them. While on the other hand, it is also noted that Shadow AI poses serious threats to cyber security, including unauthorized use of information, exposed models and loss of compliance. The anarchic nature of Shadow AI allows for a much wider attack vector in

the organization which allows for networks and data protection solutions to be exploited by advanced and targeted attacks like model poisoning, data leakage and sophisticated adversary attacks.

8. Declarations

- **Author's contribution.** D. Puthal conceptualized, designed the idea, and wrote the manuscript. The other authors, A. K. Mishra, S. P. Mohanty, A. Longo, and C. Y. Yeun, contributed to the conceptualization and writing. All authors reviewed and improved the readability of the paper.
- **AI tools.** While preparing this work, the authors used the ChatGPT and Google Gemini Tools cautiously to improve grammar, language and readability. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.
- **Competing Interests.** Not Applicable
- **Funding Information.** Not Applicable
- **Data Availability Statement.** Not Applicable
- **Research Involving Humans or Animals.** Not Applicable
- **Informed Consent.** Not Applicable

References

1. T. Chin, Q. Li, F. Mirone, and A. Papa. "Conflicting impacts of shadow AI usage on knowledge leakage in metaverse-based business models: A Yin-Yang paradox framing." Technology in Society (2024): 102793.
2. E. Yilmaz, and O. Can. "Unveiling Shadows: Harnessing Artificial Intelligence for Insider Threat Detection." Engineering, Technology & Applied Science Research 14, no. 2 (2024): 13341-13346.
3. K. Michael, R. Abbas, and G. Roussos. "AI in cybersecurity: The paradox." IEEE Transactions on Technology and Society 4, no. 2 (2023): 104-109.
4. T. O. Abrahams, S. K. Ewuga, S. Kaggwa, P. U. Uwaoma, A. O. Hassan, and S. O. Dawodu. "Mastering compliance: a comprehensive review of regulatory frameworks in accounting and cybersecurity." Computer Science & IT Research Journal 5, no. 1 (2024): 120-140.
5. R. Kaur, D. Gabrijelčič, and T. Klobučar. "Artificial intelligence for cybersecurity: Literature review and future research directions." Information Fusion 97 (2023): 101804.
6. N. Poehlmann, K. M. Caramancion, I. Tatar, Y. Li, M. Barati, and T. Merz. "The organizational cybersecurity success factors: an exhaustive literature review." Advances in Security, Networks, and Internet of Things: Proceedings from SAM'20, ICWN'20, ICOMP'20, and ESCS'20 (2021): 377-395.
7. R. Tiwari, and S. Shanmugam. "Compliance Management, Compliance and Technical Documentation Management." In 2024 IEEE Symposium on Product Compliance Engineering-(SPCE Bloomington), pp. 1-5. IEEE, 2024.
8. A. Habbal, M. K. Ali, and M. A. Abuzaraida. "Artificial Intelligence Trust, risk and security management (AI trism): Frameworks, applications, challenges and future research directions." Expert Systems with Applications 240 (2024): 122442.
9. A. Musamih, K. Salah, R. Jayaraman, I. Yaqoob, D. Puthal, and S. Ellahham. "NFTs in healthcare: vision, opportunities, and challenges." IEEE consumer electronics magazine 12, no. 4 (2022): 21-32.
10. S. Yusif, and A. Hafeez-Baig. "Cybersecurity policy compliance in higher education: a theoretical framework." Journal of Applied Security Research 18, no. 2 (2023): 267-288.
11. N. Ameen, A. Tarhini, M. H. Shah, N. Madichie, J. Paul, and J. Choudrie. "Keeping customers' data secure: A cross-cultural study of cybersecurity compliance among the Gen-Mobile workforce." Computers in Human Behavior 114 (2021): 106531.

12. N. Wirkuttis, and H. Klein. "Artificial intelligence in cybersecurity." Cyber, Intelligence, and Security 1, no. 1 (2017): 103-119.
13. D. Puthal, S. P. Mohanty, S. A. Bhavake, G. Morgan, and R. Ranjan. "Fog computing security challenges and future directions [energy and security]." IEEE Consumer Electronics Magazine 8, no. 3 (2019): 92-96.
14. M. I. Ali, and K. A. Hussain. "Unveiling the tapestry: a comparative investigation into data-protection legislation in India and Pakistan." Socrates. Rīga Stradiņš University Faculty of Law Electronic Scientific Journal of Law. 2024, no. 1-28: 1-8.
15. A. Said, A. Yahyaoui, and T. Abdellatif. "HIPAA and GDPR Compliance in IoT Healthcare Systems." In International Conference on Model and Data Engineering, pp. 198-209. Cham: Springer Nature Switzerland, 2023.
16. N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, and J. Qadir. "Privacy-preserving artificial intelligence in healthcare: Techniques and applications." Computers in Biology and Medicine 158 (2023): 106848.
17. A. O. Ugwu, X. Gao, J. O. Ugwu, and V. Chang. "Ethical implications of AI in healthcare data: a case study using healthcare data breaches from the US department of health and human services breach portal between 2009-2021." In 2022 International Conference on Industrial IoT, Big Data and Supply Chain (IIoTBDSC), pp. 343-349. IEEE, 2022.
18. M. E. Lokanan, and V. Maddhesia. "Supply chain fraud prediction with machine learning and artificial intelligence." International Journal of Production Research (2024): 1-28.
19. D. Puthal, and S. P. Mohanty. "Cybersecurity issues in AI." IEEE Consumer Electronics Magazine 10, no. 4 (2021): 33-35.
20. D. Puthal, S. P. Mohanty, E. Kougianos, and G. Das. "When do we need the blockchain?." IEEE Consumer Electronics Magazine 10, no. 2 (2020): 53-56.
21. H. El-Sayed, M. Chaqfa, S. Zeadally, and D. Puthal. "A traffic-aware approach for enabling unmanned aerial vehicles (UAVs) in smart city scenarios." IEEE Access 7 (2019): 86297-86305.
22. D. Puthal, Z. H. Mir, F. Filali, and H. Menouar. "Cross-layer architecture for congestion control in Vehicular Ad-hoc Networks." In 2013 International Conference on Connected Vehicles and Expo (ICCVE), pp. 887-892. IEEE, 2013.
23. J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang. "A survey on federated learning: challenges and applications." International Journal of Machine Learning and Cybernetics 14, no. 2 (2023): 513-535.
24. R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron. "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities." Energy and AI 9 (2022): 100169.
25. S. L. Pfleeger, and D. D. Caputo. "Leveraging behavioral science to mitigate cyber security risk." Computers & security 31, no. 4 (2012): 597-611.