

Full Name – Chris Phillips

Student ID – S2429902

Module Title – Research Methods and Professional Development (ACFA7004
COMPUTING)

Date - 1st May 2026

Assignment – Research Proposal

Research Title

*An integrated Socio-Technical Framework for AI
Resilience: Addressing Shadow AI and
underused offensive security within the UK
Transport Sector.*

Table of Contents

Research Title	1
Introduction	3
Context and Background.....	3
Research Problem	4
Research Aim and Objectives.....	5
Significance of the research.....	6
Literature Review.....	7
Governance and Human Drivers	7
Technical Resilience and Offensive AI Conflicts.....	8
Socio-Technical Alignment and Barriers to Implementation	8
Research Design and Methodology.....	11
Research Design	11
Research Methods and Materials	11
Methods.....	11
Materials	12
Expected Outcome	14
Timeline	15
References.....	16

Introduction

Context and Background

The UK public sector is currently situated at a critical juncture regarding its technological evolution. Central to this transition is the Smart Government transformation, a strategic vision aimed at modernising public service delivery through the integration of innovative, data-driven technologies. Within the UK transport sector, this ambition is paramount; digital resilience is no longer a minor concern but a fundamental requirement for maintaining national infrastructure and safety. However, as identified by the UK Government [1], this strategic ambition frequently clashes with a restrictive operational reality. The sector is currently hindered by significant implementation challenges, most notably an extensive lack of specialised technical expertise and severely restricted budgetary allocations.

These barriers are further complicated by what Prihatmanto et al. [2] describe as "technical debt", the burden of legacy infrastructure that lacks the interoperability required for modern AI integration. In the transport sector, where systems are often safety-critical and demand high-availability, the difficulty of transitioning from fragmented legacy to unified, intelligent architectures creates a state of organisational stall. This environment has fostered a disparate landscape where individual departments manage risks in isolation, leading to a state of strategic inconsistency rather than a unified socio-technical defence.

This inconsistency has given rise to two distinct crises that threaten the integrity of the transport infrastructure. Internally, a significant governance gap has emerged. In the absence of accessible, sanctioned AI tools, employees are increasingly turning to unauthorised Shadow AI to meet rising productivity demands. A study by Microsoft [3] reports that upwards of 71% of staff are effectively taking the guardrails off, creating what Puthal et al. [4] define as an "unseen frontier" of security and compliance risks, including potential data leakage and model poisoning.

Simultaneously, an external defensive gap has formed. Whilst the threat landscape rapidly evolves through AI-driven automated attacks, AI enhanced defensive postures within the transport sector remain largely passive. There is a noted failure to adopt the proactive, offensive AI security mechanisms that researchers like Jagli et al. [5] prove are necessary to maintain a technical advantage over modern, real-time threats. Instead, resources remain focused on the reactive maintenance of legacy systems, leaving the sector vulnerable to sophisticated bot-driven incursions.

Research Problem

The central research problem addressed is the profound strategic inconsistency that currently defines AI governance within the UK transport sector. While high-level strategic visions from central Government promote a fast-moving transition toward digital innovation, the operational reality at the departmental level is defined by a contradiction where internal risks remain unmanaged and external defensive capabilities are underused. This inconsistency is not simply a technical oversight; but rather a fundamental symptom of a fragmented implementation strategy that is severely constrained by legacy architecture, inadequate budgets, and a critical deficit in specialised departmental resources. Technical debt has meant the cost of retrofitting security to aging infrastructure now adds layers of complexity and cost, turning the issue from a simple oversight to a concerning economic challenge.

Internally, rigid or completely absent official policies have directly led to a significant and persistent governance gap. As established by Puthal et al. [4] and Microsoft UK [3], the lack of sanctioned, accessible and user-friendly AI tools has forced upwards of 71% of the workforce into using unauthorised Shadow AI as a workaround to maintain productivity. This widespread behavioural shift creates unknown risk regarding sensitive data usage and potential breaches. These are risks that current, static monitoring protocols fail to identify or mitigate. This governance gap is exacerbated by the consumerisation of AI; when official departmental infrastructure lags behind the capabilities of publicly available tools, the resulting Shadow AI usage represents not just a security risk, but an organic attempt by the workforce to solve operational inefficiencies that official policy fails to address.

Simultaneously, the sector suffers from an external defensive gap due to a failure to adopt proactive, automated security mechanisms. Research by Bin Akhtar and Rawol [6] demonstrates that traditional, reactive, human-led defences are increasingly inadequate against sophisticated, AI-driven threats and bot-led attacks that require real-time, autonomous responses.

The critical research gap identified is that existing academic literature frequently treats Shadow AI and AI Security as separate technical silos. Technical solutions ignore the human-centric risk whilst policy only solutions ignore the technical advancements of AI-driven bot attacks. Consequently, there is a significant lack of research regarding an integrated socio-technical framework tailored to the specific constraints of the UK transport sector. Without such a framework, the sector remains without a unified strategy to transform these disparate risks into a single, manageable resilient challenge, thereby compromising national infrastructure security.

Research Aim and Objectives

The aim of this research is to evaluate the strategic inconsistency of Artificial Intelligence adoption within the UK transport sector. By appraising the internal governance gap created by unmanaged shadow AI and measuring the technical resilience gain offered by proactive offensive security mechanisms, this study seeks to bridge the disconnect between high-level strategic vision and fragmented operational reality. The final objective of this study is the development of an integrated socio-technical framework designed to align departmental productivity with national security mandates within the specific constraints of the UK transport infrastructure.

To achieve this aim, the following three research objectives have been established:

- Objective 1: To evaluate the efficacy of current data governance frameworks within the UK transport sector in identifying and mitigating the security risks associated with unsanctioned shadow AI usage.
- Objective 2: To benchmark the technical feasibility and resilience gain of using automated offensive AI mechanisms, such as proactive bot detection, as a defensive tool for transport infrastructure.
- Objective 3: To develop an integrated socio-technical active guardrail framework—the project artefact—that addresses institutional barriers such as technical debt to align user productivity with agency security mandates.

The first objective focuses on the governance gap, moving beyond a simple policy check to a critical analysis of security necessity. This is informed by the warnings of Puthal et al. [4], who identify shadow AI as an unseen frontier of potential data leakage that current frameworks often fail to address. By investigating the behavioural shift of the 71% of staff using unsanctioned tools, the research identifies the human-centric drivers that must be managed to secure the sector.

The second objective shifts the focus to defensive capability, specifically measuring how offensive AI techniques provide a tangible resilience gain over traditional reactive postures. This assessment incorporates Yampolskiy's [7] research regarding controllability limits to ensure that the proposed application of AI is not merely assessed for speed, but for safety and verification within a public sector context.

The final objective is the combination of the study, addressing the implementation challenges identified by Prihatmanto et al. [2], specifically the technical debt and legacy barriers that cause digital transformations to fail. By using the socio-technical framing advocated by Michael et al. [8], the research ensures the final active guardrail framework serves as a practical bridge between technical resilience and the reality of limited UK transport sector resources. This streamlined three-objective approach ensures the research remains feasible within the allocated timeframe while directly addressing the integration of a socio-technical framework.

Significance of the research

The significance of this research is grounded in the possibility of resolving the strategic inconsistency that compromises the security of the UK transport sector. By providing a data-driven justification for the integration of proactive AI defences, the study moves beyond theoretical debate to offer a practical design for institutional resilience, as advocated by Bin Akhtar and Rawol [6]. Societally, the research is critical as it addresses the vulnerabilities of national infrastructure; ensuring that transport systems are resilient against automated threats as a matter of public safety and economic stability.

Technically, the development of the active guardrail framework introduces a novel method for balancing human productivity with effective security protocols. Rather than relying on ineffective restrictive measures, which Haag and Eckhardt [9] suggest often fail to reduce shadow IT usage, this study proposes a socio-technical alignment that acknowledges the reality of user behaviour while maintaining active and progressive monitoring. The practical applications of the research include a technical decision matrix that can be used by departmental leads to justify AI investment and a governance protocol that ensures innovation does not outpace controllability. Ultimately, this research provides the necessary bridge between the ambition of smart government and the operational security requirements of a modern digital state.

Literature Review

The purpose of this literature review is to critically appraise the existing body of knowledge regarding AI governance and automated security, specifically identifying the trends and gaps that inform the UK transport sector's current strategic inconsistency. The scope of this review considers peer-reviewed research from 2022 to 2025, alongside critical governmental and industrial reports, to ensure the study is grounded in both theoretical depth and current operational reality.

The review is organised thematically into three core areas: the governance of human drivers regarding AI use, the technical paradigms of offensive AI resilience, and the socio-technical barriers to implementation. By grouping these studies, the review aims to move beyond isolated technical silos to evaluate the connections between human behaviour and intelligent defensive systems. Ultimately, this section will summarise major contributions and theoretical models—such as socio-technical framing and AI controllability—to demonstrate how this research will broaden current findings by proposing a unique, integrated framework for the transport sector.

Governance and Human Drivers

The emergence of unauthorised Artificial Intelligence usage, frequently termed Shadow AI, represents a significant shift in the internal threat landscape of the public sector. Puthal et al. [4] describe this phenomenon as an unseen frontier, characterised by unauthorised AI computation and high risks of sensitive data leakage. Their research highlights that this lack of visibility within traditional monitoring frameworks directly facilitates model poisoning and significant compliance breaches. In the context of the UK transport sector, this theoretical risk is given empirical weight by the Microsoft [3] findings, which indicate that approximately 71% of the workforce uses unsanctioned AI tools to maintain productivity in the absence of accessible, official and secure alternatives.

Critically, the literature suggests that traditional, restrictive cybersecurity measures are increasingly inadequate for addressing this behavioural shift. Haag and Eckhardt [9] argue that Shadow IT, the predecessor to Shadow AI, is primarily driven by unmet user needs within the workplace. Their findings demonstrate that a strategy of blocking access to unauthorised tools is significantly less effective than a management approach that seeks to understand and integrate user requirements. This suggests that the governance gap in transport organisations is not merely a technical failure of software blocking, but a failure of policy to align with the operational realities of a modern workforce.

The combination of these works establishes that Shadow AI is a multi-faceted challenge. While Puthal et al. [4] justify the technical and security necessity for enhanced internal monitoring, Haag and Eckhardt [9] provide the sociological justification for why these

security measures must be flexible and user centric. Within the transport sector, where regulatory compliance and data integrity are most important, this literature highlights a critical need for governance frameworks that move beyond inaction. The current strategic inconsistency identified in the sector stems from a failure to recognise that the human desire for productivity will consistently bypass rigid IT guardrails unless a managed, socio-technical middle ground is established.

Technical Resilience and Offensive AI Conflicts

The external threat landscape facing the transport sector is increasingly defined by automated, AI-driven attacks that target traditional, passive security architectures. Bin Akhtar and Rawol [6] provide the technical foundation for a shift in defensive strategy, arguing that AI transforms cybersecurity into a proactive, intelligent, and adaptable standard. Their work illustrates that by using AI-powered security mechanisms, organisations can move beyond static firewalls and manually intervened security operations processes to a system that organically evolves alongside the threat. In the context of the UK transport sector, this resilience gain is critical. As infrastructure becomes more interconnected, the inactivity involved in reactive, human-led responses becomes a primary vulnerability.

Supporting this shift, Jagli et al. [5] evaluate the specific role of machine learning and behavioural analysis in real-time threat detection. Their key finding suggests that these technical tools allow defenders to stay one step ahead of attackers by identifying patterns in bot traffic that remain invisible to conventional signature-based detection. This represents a significant advancement over the reactive methods currently used within many public sector departments. This literature provides the mathematical and technical justification for the offensive AI component of the framework, proving that automated detection is not just an efficiency gain but a necessity for modern infrastructure resilience.

However, a critical gap exists between this technical potential and operational deployment. While Bin Akhtar and Rawol [6] and Jagli et al. [5] prove the effectiveness of these tools in controlled or theoretical environments, they offer limited insight into how these automated systems should be governed when integrated with legacy transport systems. This technical literature justifies the offensive considerations of the proposed framework, however simultaneously highlights the need for a socio-technical association. Without a governance layer to manage these intelligent systems, the transport sector risks trading one form of unmanaged risk (Shadow AI) for another (unchecked automated defence). The combination of these technical papers establishes the benchmark that the proposed framework seeks to reach, while confirming that technology without strategic alignment, remains an incomplete solution.

Socio-Technical Alignment and Barriers to Implementation

The final thematic area addresses the significant disconnect between the availability of advanced security technologies and their successful integration within the public sector. Michael et al. [8] provide the primary theoretical foundation for this study by arguing that

cybersecurity must be conceptualised through a socio-technical framing. The work suggests that AI threats cannot be managed through technical tools alone; instead, they must be understood as a combination of social, technical, and environmental dimensions. For the transport sector, this means acknowledging that the governance of shadow AI and the deployment of offensive security are two aspects of existing deficiencies which require simultaneous and harmonised resolution.

However, the practical application of this socio-technical approach is often hindered by implementation challenges. Prihatmanto et al. [2] identify that "Smart Government" transformations frequently fail not due to a lack of vision, but because of legacy technical debt and poor interoperability between fragmented systems. Within the UK transport sector, this is corroborated by the UK Government [1], which highlights that budget variances and a lack of departmental resources often force organisations to maintain reactive legacy postures rather than adopting proactive AI resilience.

To bridge this gap, the literature stresses the need for ethical and safety-critical guardrails. Leo and Archie [10] mandate that AI-driven security must balance operational efficiency with meticulous ethical standards to remain valid in a public sector context. This is a critical requirement because, as Yampolskiy [7] demonstrates, advanced AI systems face formal controllability barriers that make safety verification difficult. Consequently, the proposed active guardrail framework acts as the blend of these works. It adopts the socio-technical framing of Michael et al. [8] to address the technical debt identified by Prihatmanto et al. [2], while ensuring that any offensive capability respects the safety boundaries established by Yampolskiy [7]. This combination proves that an integrated framework is the only suitable path forward for securing the transport sector against the dual threats of internal governance gaps and external automated attacks.

In summary, the reviewed literature reveals a significant state of strategic inconsistency within the UK Government digital landscape. Internally, a profound governance gap exists where approximately 71% of staff [3] use unauthorised shadow AI to meet productivity demands, creating an unseen frontier of risk that traditional restrictive measures fail to allay. Simultaneously, an external defensive gap is evident. Whilst research proves that offensive AI mechanisms, such as automated behavioural analysis can effectively out scale bot-driven attacks, these capabilities remain unused in favour of reactive legacy maintenance.

The overarching gap identified across these themes is the lack of an integrated approach. Current academic works frequently treat human-centric governance and technical defensive resilience as separate issues. Furthermore, the technical debt and institutional barriers prevalent in the UK transport sector often prevent the theoretical benefits of AI security from being realised in practice. This research addresses these deficiencies by proposing the new framework. This is not a single tool intended to fix a singular issue, but a dual-layered socio-technical strategy designed to manage human risk and technical resilience simultaneously. By integrating the ethical requirements of AI controllability with the technical necessity for offensive security, the framework

provides a practical link for the transport sector to move from a posture of reactivity to one of a proactive, managed resilience.

Research Design and Methodology

Research Design

The research design for this study is a single, fixed case study focusing on the UK transport sector. This design is selected to provide a deep, contextualised understanding of the strategic inconsistency between high-level innovation mandates and operational security realities. A case study approach is uniquely appropriate for this research as it allows for the investigation of current experiences within real-life context, specifically the juxtaposition of legacy infrastructure with the explosion of artificial intelligence.

To ensure a comprehensive appraisal and to satisfy the requirement for an integrated research artefact, the study adopts a mixed methods approach. This approach focuses on two primary methodological streams: a quantitative assessment of technical resilience and a strategic gap analysis of existing governance. This dual-pillar design ensures that the final active guardrail framework is grounded in both practical performance data and an understanding of current policy shortcomings.

The overall research philosophy is grounded in realism. This stance prioritises practical solutions to complex problems, acknowledging that neither a purely technical nor a purely social approach is sufficient to secure a satisfactory solution. Consequently, the research moves beyond the theoretical, to produce a framework that acts as a functional research artefact for the sector. By using a deductive approach, the research tests the hypothesis that current transport sector strategies are mismatched with operational reality, providing the evidence required to propose a realigned, socio-technical alternative.

Research Methods and Materials

Methods

The research uses two primary methods to achieve the desired objectives and inform the development of the framework artefact. Secondary quantitative comparative analysis and strategic gap analysis.

Firstly, a secondary quantitative comparative analysis is employed to establish the resilience gain provided by offensive AI. This method involves extracting performance benchmarks, specifically accuracy and detection latency, from established literature and applying them to an industry-standard Bot-IoT dataset. By mathematically comparing the response times of automated AI models against the established performance benchmarks for manual detection found within broader cybersecurity industry literature, the research physically demonstrates the technical necessity for proactive defence. This provides the empirical engine for the proposed framework, proving that AI can scale at the speed of modern automated threats where traditional,

human-dependent processes fail. This comparative approach relies entirely on publicly available technical data and published performance metrics, ensuring the study remains feasible without the requirement for restricted departmental access.

Secondly, a strategic gap analysis is conducted to map the disconnect between user behaviour and organisational policy. This involves the evaluation of high-fidelity secondary data from Microsoft [3] report to quantify the volume and risk exposure of shadow AI usage within the professional environment. By contrasting these behavioural realities with the mandates found by the UK Government [1], the research identifies the specific institutional barriers, such as technical debt and rigid policy that lead to strategic inconsistency.

This two-factor methodology ensures that the final active guardrail framework is not merely a theoretical proposal. The technical comparative analysis defines the guardrail's defensive capabilities, while the strategic gap analysis defines the active governance protocols required to manage human risk. By combining these findings, the research produces an integrated socio-technical artefact that is both technically robust and tailored to the specific resource constraints of the UK transport sector.

Materials

The successful completion of this case study relies on the integration of two distinct data streams, each requiring specific materials and analytical tools to ensure academic rigour.

Research data and sources for the strategic gap analysis will use secondary survey data [3]. This material provides the necessary frequency counts and user-reported motivations to establish a behavioural baseline for shadow AI usage. To contrast this with official mandates, the research also incorporates strategic policy documents, specifically the UK Government [1]. These materials are essential for identifying the institutional barriers, such as technical debt and budget variances, that necessitate a new governance protocol.

The quantitative comparative analysis of offensive security requires an industry-standard, publicly available Bot-IoT dataset to serve as a proxy for transport network traffic. This dataset is required to apply the accuracy, precision, and latency benchmarks established in the works of Bin Akhtar and Rawol [6]. These technical benchmarks serve as the primary materials for demonstrating the resilience gain of automated systems over manual baselines.

To perform the comparative and descriptive statistical analysis, the research will use software tools such as Python for technical simulation and SPSS for the strategic gap analysis of behavioural data. Python (with Pandas and NumPy libraries) is required for processing the technical benchmarks within the Bot-IoT dataset, while SPSS will be used to conduct descriptive statistical evaluations of user-reported data from the Microsoft

2025 Frontier Firm report. In addition Microsoft excel will be used as an analytical tool for relevant data calculations.

Hardware requirements are limited to standard computing resources with sufficient processing power to execute simulation scripts in a controlled environment. Access to academic repositories such as IEEE Xplore and the university library is required for continuous literature support and to ensure all technical metrics remain current.

As this study relies exclusively on publicly available secondary datasets and published governmental reports, there is no ethical consideration for direct human participant intervention. Consequently, issues regarding informed consent, primary data anonymity, and participant harm are mitigated. However, ethical integrity is maintained through the transparent reporting of the secondary data's original context and the objective application of performance metrics. All data will be handled in accordance with UK university ethical guidelines, ensuring that the interpretation of these materials into the final framework remains unbiased, verifiable, and strictly focused on the UK transport sector's strategic needs.

Expected Outcome

The primary expected outcome of this research is the consideration for an active guardrail framework, a validated socio-technical artefact designed to resolve the strategic inconsistency within the UK transport sector. By combining the human drivers identified leading to shadow AI usage with the technical necessity of offensive security, the research expects to provide a mathematical justification for strategic alignment. The impact of this work is focussed on shifting the conversation from restrictive policing to the development of a scalable, resilient defence that balances operational efficiency with UK Government mandates.

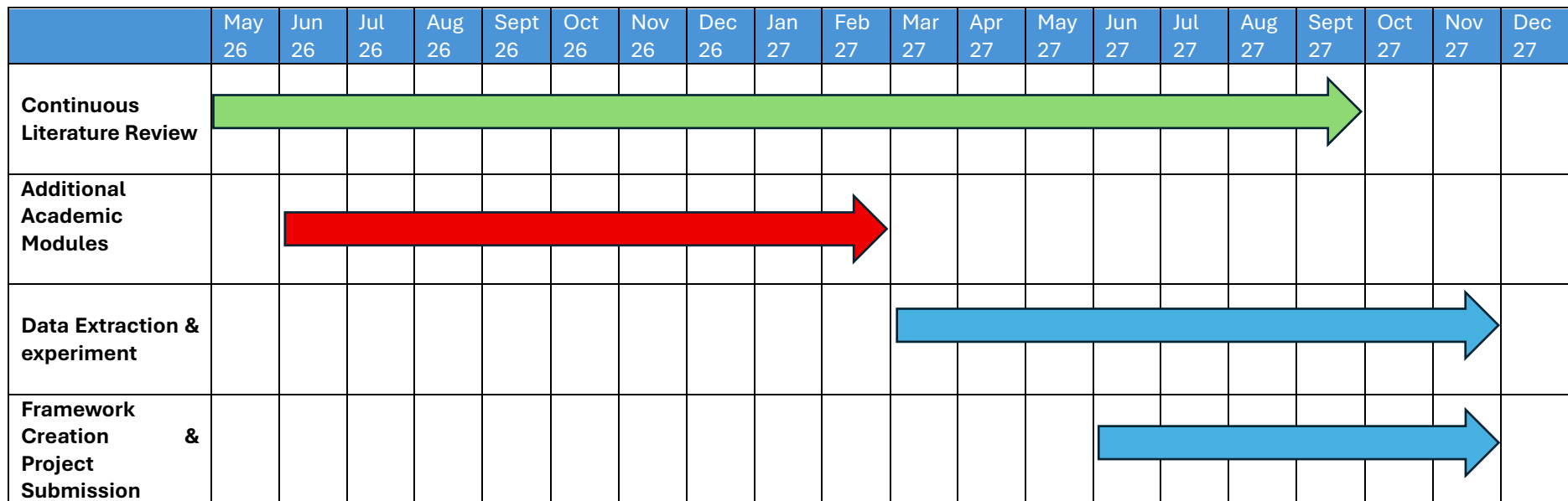
Technically, the research expects to demonstrate a quantifiable resilience gain, proving that automated AI-driven detection can significantly outperform traditional manual response intervals. This provides a practical bridge for departmental leads, offering a decision matrix that acknowledges the reality of technical debt while providing a roadmap for future-proof digital advancements. Societally, the framework aims to bolster the reliability of UK transport sector services, ensuring that public services remain secure against the increasing velocity of automated threats.

However, certain research limitations are acknowledged. To maintain the security integrity of the UK civil service, the study uses industry-standard secondary datasets as a proxy for sensitive government traffic. While these datasets are relevant to UK industry, it is recognised that a simulation environment may not capture every nuance of a complex public sector architecture. Furthermore, while the framework addresses the institutional barriers created by legacy systems, it cannot physically remediate existing technical debt. Finally, given the rapid evolution of the field, it is anticipated that the technical benchmarks established during this study will require periodic review to remain valid as the artificial intelligence landscape matures toward late 2027.

Timeline

The following Gantt chart outlines the 21-month execution plan for this research, spanning from the current proposal phase in early 2026 through to the final MSc Project submission in late 2027. This timeline is structured to accommodate the requirements of part-time study while ensuring sufficient periods for thorough data extraction and iterative framework development.

Key milestones include the completion of additional academic modules in late 2026/early 2027, followed by a concentrated six-month phase for the simulation experiment and comparative analysis. The final quarter of 2027 is dedicated to the creation of the active guardrail framework and the formal documentation of the research findings. This staggered approach allows for a continuous literature review process, ensuring the framework remains aligned with the rapidly evolving artificial intelligence landscape throughout the duration of the study.



References

- [1] Gov.uk, “State of digital government review A review of technology and data in the public sector: successes, challenges and root causes.” Jan. 01, 2025. [Online]. Available: <https://www.gov.uk/government/publications/state-of-digital-government-review/state-of-digital-government-review>
- [2] A. S. Prihatmanto, R. Andrian, W. D. Sunindyo, and R. Sutriadi, “Transforming Public Services: A Systematic Review of Smart Government Frameworks, Architectures, and Implementation Challenges,” *IEEE Access*, vol. 12. pp. 135799–135810, 2024. doi: 10.1109/access.2024.3450907
- [3] Microsoft, “2025: The Year the Frontier firm is born.” Apr. 23, 2025. [Online]. Available: <https://www.microsoft.com/en-us/worklab/work-trend-index/2025-the-year-the-frontier-firm-is-born>
- [4] D. Puthal, A. K. Mishra, S. P. Mohanty, A. Longo, and C. Y. Yeun, “Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier,” *SN Computer Science*, vol. 6, no. 5, Apr. 2025. doi: 10.1007/s42979-025-03962-x
- [5] D. Jagli, R. Temkar, L. Nakirekanti, and A. Bhatt, “The Role of Artificial Intelligence in Cyber Security.” *Journal of Electrical Systems*. vol 20. pp. 5283-5291 April 2024. doi: 10.52783/jes.6327
- [6] Z. Bin Akhtar and A. T. Rawol, “Enhancing Cybersecurity through Artificial Intelligence (AI) -Powered Security Mechanisms,” *Information Technology and Journal of Research and Development*, vol. 9, no. 1. pp. 50–67, Aug. 2024. doi: 10.25299/itjrd.2022.16852

- [7] R. V. Yampolskiy, "On the Controllability of Artificial Intelligence: An Analysis of Limitations." *Journal of Cyber Security and Mobility*. May 2022. doi: 10.13052/jcsm2245-1439.1132
- [8] K. Michael, K. M. Vogel, J. Pitt, and M. Zafeirakopoulos, "Artificial Intelligence in Cybersecurity: A Socio-Technical Framing," *IEEE Transactions on Technology and Society*, vol. 6, no. 1, p. 15, Mar 2025. doi: 10.1109/tts.2024.3460740
- [9] S. Haag and A. Eckhardt, "Dealing Effectively with Shadow IT by Managing Both Cybersecurity and User Needs." *MIS Quarterly Executive*. p. 399, Dec. 2024. doi: 10.17705/2msqe.00104
- [10] N. Leo and O. Archie, "AI and Cybersecurity for SMEs: Balancing Ethical Considerations and Operational Efficiency." *ResearchGate* Sept. 29, 2024. doi: 10.13140/RG.2.2.33120.49923