

AI and Machine Learning

by Naod Habtom

Submission date: 04-Jan-2026 06:11PM (UTC+0000)

Submission ID: 272012940

File name: 50226_Naod_Habtom_AI_and_Machine_Learning_712345_783474894.docx (89.09K)

Word count: 6003

Character count: 36327



Research and case scenario in Data Security and Compliance

Data Security and Compliance

Computing Portfolio



NOVEMBER 14, 2025
NAOD HABTOM (2306565)
University of Wales Trinity Saint David (UWTSD)

Contents

<i>Abstract</i>	2
Introduction.....	2
Task 1.....	2
1 Overview	2
2. Risk Landscape for AI/ML Systems.	3
3. Governance Frameworks and International Standards.	3
4. Data Protection and Regulatory Compliance.	4
5. Security Architecture Implications.....	5
6. Monitoring and Automated Response for AI Systems.....	5
7. Critical Evaluation and Real-World Examples.	6
8. Summary Task 1.....	6
Task 2.....	7
1. Introduction and Organisational Context.	7
3. Organisational Structure and Key Roles.	8
4. Security Policies and Procedures.	8
5. DORA, NIS2 and GDPR are all Compliance.....	9
6. Governance And Strategy In NIST CSF 2.0.	9
7. Risk Assessment and Management.	10
8. Contingency and Disaster Recovery Planning.	10
9. Protection Mechanisms and Control Implementation.	11
10. Programme Roadmap and Assurance.....	11
11. Summary Task 2.....	11
Conclusion	12
References.....	13

GOVERNANCE OF AI AND MACHINE LEARNING SYSTEMS: A ZERO-TRUST APPROACH TO SECURITY AND COMPLIANCE

Naod Habtom

Department of Computing and Information Systems

University of Wales Trinity Saint David

Email: 2306565@student.uwtsd.ac.uk

Abstract

The use of Artificial Intelligence (AI) and Machine Learning (ML) in business systems may have increased efficiency, but has also led to more sophisticated security and ethical challenges. Governance of AI and ML models has become essential to ensure transparency, fairness and resilience against emerging threats. This essay critically examines governance frameworks, including the NIST AI Risk Management Framework, ISO/IEC 42001 and the upcoming EU AI Act. The essay assesses their implications for accountability and compliance. As per the European Commission (2024), AI governance must be in coherence with Cyber Security Paradigms (for alternate perspective Zero Trust, GDPR and DORA)^[1]. To conclude, effective AI-ML governance requires an integrated, risk-based approach to technology, ethics, and regulation.

Keywords: Artificial Intelligence (AI), Machine Learning (ML), AI Governance, Zero Trust, ISO/IEC 42001, NIST AI RMF, DORA, NIS2, GDPR, Cybersecurity Compliance

Introduction

The quick development of Artificial Intelligence, Machine Learning and hybrid workforce models has transformed business operations in organizations, yet this digital acceleration has made cybersecurity threats more complex. Today's businesses should not just focus on network intrusions, but also regulate algorithmic risks, data invocations and compliance deviations through AI-enabled systems^[2]. The European Commission in 2024 has written regulatory frameworks and their pressure on organizations, which include the General Data Protection Regulation, digital operational resilience, and the Network and Information Systems Directive. The security of each organisation is very critical for the European Commission.

This report, therefore, comprises two integrated tasks. According to Task 1, the governance of AI systems and ML systems will be examined through the lens of a number of frameworks. The NIST AI Risk Management Framework, ISO/IEC 42001, and the proposed EU AI Act will be examined. Best practices for risk mitigation and ensuring ethical use will be reported. Task 2 will take these insights and apply them to a case study regarding Resilient Intellect Ltd, a security firm that is expanding into Europe. Specifically, a security and compliance program which aligns with Zero-Trust will be designed. When combined, these parts show how a structured governance framework, along with the flexible security architecture, collectively enhance the resilience, transparency and regulatory compliance within an organization.

Task 1

1 Overview

Today, every sector's digital transformation, innovation, and work efficiency rely on Artificial Intelligence (AI) and Machine Learning (ML). Furthermore, industries use AI and ML to enable predictive capabilities. However, this has now created a broader and more complex threat landscape that exerts immense pressure on governance and cybersecurity frameworks. As AI systems become more widespread, concerns over algorithmic bias, data privacy, model manipulation, and transparency have emerged with significant regulatory and ethical implications^[3]. As organisations begin to implement AI and ML in their core operational processes and decision-making, governance frameworks must evolve effectively to manage emerging risks and ensure accountability.

The governance of AI and ML systems and the need for developing effective risk-based governance regimes that

can suit both technical and regulatory requirements will be dealt with in this paper. The paper evaluates internationally recognised frameworks such as the NIST AI Risk Management Framework, ISO/IEC 42001, and the forthcoming EU AI Act, their applicability and limitations in today's business environment (ISO, 2024). It also examines how these frameworks relate to broader cybersecurity strategies – particularly the use of Zero Trust to make models and data secure^[3].

The analysis findings feed directly into Task 2, in which the governance principles are applied to the case of ResilientIntellect Ltd. This paper will demonstrate how strong AI/ML governance promotes resilience, transparency, and compliance. It also explores the growing legal scrutiny of operational interdependence in a digitalised world through evaluating theoretical and practical perspectives^[4].

2. Risk Landscape for AI/ML Systems.

The potential dangers arising from AI and machine learning systems cover much more than the usual cybersecurity stuff; they also involve the quality of data, the reliability of the algorithm, and the kinds of systemic dependencies that change over time. AI systems differ from traditional software because they create their own rules as they learn, which gives rise to uncertainty^[5]. We can group the most serious risks into three major areas, such as threat to the data itself, model risk, operational risk or supply chain risk.

Threats to datasets may arise during collection, labelling or training, which drags the data. Data poisoning is when an adversary tries to maliciously manipulate the training process by injecting adversarial data through direct input or through the dataset. This data can heavily distort the learning process and result in wrong predictions or biased predictions^[6]. It is equally worrying to use someone's personal or proprietary data, as this is against ethical principles and legal frameworks like the General Data Protection Regulation (GDPR)^[7]. Inadequate data lineage tracking weakens accountability and makes it hard to trace data origin and verify compliance through complex AI pipelines (Mouzakitis et al., 2023).

Risk of Model-specific attack refers to adversarial attack, model inversion or model extraction attacks. Here, malefactors tamper with the output of an AI model or

reverse-engineer the used model to generate access to protected data, such as sensitive personal data or intellectual property of the model, as depicted by Papernot et al. (2018). When datasets are unbalanced, such as lacking sufficient representative samples or including errors, harm is caused. In high-stakes areas like finance and health, that's especially true. Bias and discrimination can lead to harm to one's reputation, regulatory fines and ethical breaches. The aforementioned problems indicate the need for governance mechanisms embedding fairness, explanation and validation across the AI lifecycle^[8].

Risk often arises from something that causes harm, depending on its impact. A lot of organisations are choosing to use pre-trained or open-source models, which introduce hard-to-assess or hard-to-mitigate external vulnerabilities^[9]. In addition, deploying models continuously via MLOps pipelines introduces attack vectors that are not detected due to insufficient monitoring of automated updates. Regulatory compliance is especially difficult for multinational operations with differing data localisation laws and risk classification frameworks^[10].

The risks call for a governance approach and not a technical fix in isolation. By combining risk analysis with oversight of impact, accountability and ethics, AI and ML can boost reliable and sustainable digital transformation. Governance must evolve into an adaptive system, balancing innovation with control, to maintain trust in AI-based decision making in a world that is becoming more regulated^[11].

3. Governance Frameworks and International Standards.

AI and ML technology is shaping the future economy, and the governing frameworks and international standards are being designed to make AI systems more trustworthy, accountable and controllable. There are many structured ways of assessing, identifying, and reducing risk in data, model, and operational layers. Some of the most influential documents when it comes to AI governance are the NIST AI Risk Management Framework (AI RMF 2023), the ISO/IEC 42001 (2024), and the EU Artificial Intelligence Act (2024). These three documents relate to governance from institutional, technical, and regulatory perspectives, respectively. The combination of these documents should

somehow contribute to a strong governance system for AI^[12].

The NIST AI RMF is a comprehensive, voluntary framework that helps organizations Govern, Map, Measure, and manage artificial intelligence. They stress the importance of risk management integration into the AI lifecycle with special focus on “trustworthiness” attributes, including transparency, accountability and robustness, NIST 2023. Its lack of binding nature may limit enforcement measures, while its strength lies in the different sectors it can apply to.

Poor AI systems can result from the upskilling of users through AI-enabled information and communication technologies. In contrast, ISO/IEC 42001 sets up a management system for AI. It follows the continuous-improvement cycle of ISO/IEC 27001. Documented policies are needed for leadership. Also, performance evaluation is mandated, and periodical audits are mandated. This framework puts governance on a formal footing by embedding ethics in operational requirements that are measurable. While it imposes consistency in procedures, it may not allow the flexibility organisations require to work with an AI architecture that evolves rapidly.

The EU AI Act is a regulatory mechanism based on a risk classification system that classifies AI applications into unacceptable, high-risk, limited-risk and minimal or no-risk classification^[13]. High-risk systems like credit scoring systems, healthcare diagnosis, and employment screening must meet stringent requirements, including data governance, human oversight and requests for information. According to the Act, compliance assessment and post-deployment monitoring create a legally binding accountability framework for AI providers. Although this facilitates coordination throughout the EU, critics claim it might limit innovation because of compliance weights, especially for smaller businesses^[14].

Moreover, non-binding principles like the OECD AI Principles (2019) and UNESCO’s Recommendation on the Ethics of AI (2021) support compliance with formal standards to reinforce fairness, human-centred design and sustainability. AI governance should be inclusive, equitable and collaborative. This means drawing decisions on social choice rather than technology.

The NIST AI RMF emphasises adaptability and improved processes, while ISO/IEC 42001 enforces

structured accountability. The EU AI Act sets compliance with enforceable accountability to provide better AI governance. The best governance approach might thus bring together all three of NIST’s risk functions, ISO’s audit discipline and the regulatory obligations of the AI Act^[15]. This hybrid model offers organisations privacy controls while allowing for the flexibility of remote and in-person work.

Governance cannot remain still, but rather, must keep up with AI. Supervised or controlled models will be necessary as the models become autonomous and self-learning. Providing oversight will simplify governance. The fusion of technical and legal frameworks underpins responsible AI governance. This is a dynamic process to ensure safety, reliability and public trust in intelligent systems (Brundage et al., 2023).

4. Data Protection and Regulatory Compliance

Keeping data safe and following the laws are important parts of making rules for AI and ML. This is because using algorithms to make decisions usually requires a lot of data. And this could be personal, behavioural, or financial in nature. As AI technologies continue to grow, regulators the world over have strengthened compliance obligations to ensure organisations remain accountable and transparent throughout the AI lifecycle^[16]. GDPR is the backbone of European data governance, which requires legal processing, data minimisation, and the right to explanation in case of automated decision-making. It directly relates to AI systems that make or assist in autonomous high-impact decisions.

Still, GDPR covers the full spectrum of digital resilience that AI-driven enterprises and applications demand. A number of European legislative measures have passed in recent times that govern the cybersecurity resilience of organisations, including financial institutions, telecom operators and other essential parties. DORA focuses on ongoing risk assessments, testing and monitoring of ICT providers. On the other hand, NIS2 requires network and information systems security for operators of essential and digital services. In short, all of them push compliance beyond privacy protection toward holistic operational governance.

We must harmonise the various overlapping frameworks. AI systems that consume and process big data for predictive

purposes must implement measures for explainability to comply with the transparency principles of the GDPR, as well as measure their resilience and data integrity under DORA^[17]. Additionally, when you work across borders, it is harder to keep your business in compliance. As a result, global companies need standardised data-governance schemes (Floridi, 2024), which would incorporate automated compliance monitoring, access controls, and audit trails, among other things.

To put it another way, good AI compliance is more than just complying with the law. Moreover, it involves implanting data-protection principles in the design. Most people will know this as privacy by design. AI pipelines must use encryption, anonymization and minimisation from the first stages of their development. Governance models should be equipped with the processes for ethical review, bias audits, and algorithmic transparency for accountability and fairness. Strong data protection is needed for the warranty of trust in AI. It could change compliance from a legal requirement to a foundation of sustainable digital transformations^[18].

5. Security Architecture Implications.

AI and ML governance can only work well on a resilient security architecture that embeds continuous validation, least privilege and a proactive defence mechanism. ZTA or Zero Trust architecture has emerged as a central paradigm in managing AI-related risks that end implicit trust and verification at every layer of interaction (Rose et al., 2020). Through dynamic risk signals, users, devices, and applications are always authenticated in Zero Trust. The governance principles of AI applications are closely aligned with this architecture, establishing an auditable access control and accountability in the data pipeline.

AI systems rely on cloud platforms, APIs, and third-party tools, all of which increase the attack surface. Governance thus necessitates the integration of Secure Access Service Edge (SASE) and Zero Trust Network Access (ZTNA) to control connectivity. According to Shah and Kumar (2023), these models provide identity-centric access, endpoint verification and micro-segmentation, isolating sensitive model-training environments from the larger IT environment. Furthermore, organisations use data loss prevention (DLP) tools and/or Cloud Security Posture

Management (CSPM) to follow the flow of sensitive data as they deploy in multiple clouds.

Embedding Zero Trust in AI systems ensures that they comply with the security-by-design and resilience-by-default principles outlined in both ISO/IEC 42001 and the EU AI Act (ISO, 2024; European Commission, 2024). For example, DLP mechanisms help comply with GDPR's data minimisation requirement while strong authentication safeguards address DORA's ICT-risk requirements. It turns compliance frameworks from mere theoretical proposals into enforceable technical reality.

However, zero-trust implementation for AI presents challenges. Monitoring and access checks require many computing resources and identity infrastructures. Excessive dependence on machine learning for decision-making can skyrocket false positives, lowering efficiency. Governance relies on a combination of both (semi)automated and human mechanisms for adaptive control and proportionate enforcement^[19].

All in all, AI governance operations are built on security architecture. When enterprises adopt Zero Trust and related cloud-native protection mechanisms, they build an ecosystem that not only defends against attacks but also demonstrates compliance with relevant regulations through measurable controls and visibility through auditability.

6. Monitoring and Automated Response for AI Systems.

The guarantee layer is monitoring and automated response. As the models change and engage in a new situation, continuous monitoring will ensure detection of anomalies, data integrity, and ethical outcomes. Good monitoring is a combination of technical telemetry and behavioural analytics, together with governance dashboards that are aligned with the risk appetite and compliance of the organisation (NIST, 2023).

Monitoring AI-driven systems requires more than network-level metrics. The models require monitoring of model behaviour, data drift, and fairness metrics to identify deviations that could compromise accuracy or compliance.

Connecting with SIEM software makes it easier to find patterns between operational and security anomalies^[20]. With the help of MLOps security, models at the training and deployment phases get their sequencing validated in real time. This helps reduce model drift and unapproved updates^[21].

With automated response mechanisms for monitoring, systems can also respond to detected threats. For instance, when a model is trying to be poisoned or attacked, automation pipelines could cause a rollback, isolation, or retraining with verified datasets. DORA's guidelines focus heavily on resilience and ensuring a steady operational capacity.

Governance should mark the limits of automation. When machines fix problems on their own, it can be hard to tell who is responsible when something goes wrong. So, a human must ultimately approve automated actions to ensure they are ethical and legal^[22]. Ultimately, the governance of monitoring and response systems is an amalgamation of technical control and ethical accountability, which grounds intelligent automation in trust and transparency.

7. Critical Evaluation and Real-World Examples.

Seeing how an AI governance framework works shows you where they are useful and where they fail. For example, Microsoft's (2023) Responsible AI Standard shows how structured governance can be embedded in a product through fairness, accountability, transparency and safety principles. The cross-functional review boards and ongoing assessment of models will effectively operationalise ethical AI (Brundage et al., 2023). On the other hand, certain public-sector projects, such as predictive policing, have come under scrutiny for algorithmic bias and poor transparency, revealing the reputational and legal risks of weak governance.

In the healthcare sector, 2023 was marked by an occurrence of data poisoning in diagnostic imaging models. As a result, risk assessment and dataset validation became more important than ever. According to NIST (2023), NIST AI RMF adoption enhances traceability and accountability during post-incident analysis. The European financial institutions implementing ISO/IEC 42001 frameworks see

better alignment with compliance under DORA and NIS2, as well as better scores on audit readiness and stakeholder confidence^[23].

These cases show that no framework is secure or fair in all cases. In order to have effective governance, one must adapt to the context, that is, combine standards with organisational culture and follow through. There is some evidence from practice which shows that AI governance works best when integrated into decision-making and not tacked on to a compliance exercise.

8. Summary Task 1

Governance of AI and ML is transformational in nature for organisational risk management. The shift from isolated data-protection policies to integrated governance systems reflects the idea that such technical innovation should be matched by accountability, transparency, and resilience (NIST 2023). The study has shown that frameworks such as the NIST AI RMF, ISO/IEC 42001, and EU AI Act provide essential building blocks for risk management, continuous improvement, and enforcement, respectively. These standards enhance Zero Trust architectures and operational frameworks (DORA, NIS2) with a view to establishing a comprehensive framework for AI risk.

Critically, governance must remain adaptive. Regulatory power cannot chase the speed of AI technologies. Therefore, dynamic monitoring, automated assurance, and ethical oversight need to match the pace of technology in order to prevent system failure. Going forward, governance should be marked by explainability, fairness, auditing and alignment of human-algorithmic accountability. When organisations make these methods a part of their corporate strategy, they will remain in compliance and will not stifle innovation.

In the end, good AI/ML governance is not just a defensive tool. But is also a potential strategic asset that builds trust, reduces uncertainty and ensures long-term digital resilience. Combining regulatory compliance, ethical design, security and Zero Trust creates a foundation for responsible innovation, so that the benefits of AI can be realised safely, equitably and transparently across the modern enterprise.

Task 2.

1. Introduction and Organizational Context.

Resilient Intellect Ltd is a mid-sized security firm that harnesses AI to detect cyber threats and attacks. The firm works with a blend of local and offshore teams. They use cloud-native technology and advanced analytical models to provide intelligence to their clients. With the organization planning to enter the European Union, it will experience increased compliance obligations for the Digital Operational Resilience Act (DORA), the Network and Information Systems Directive 2 (NIS2), and the General Data Protection Regulation (GDPR) (European Commission, 2024). Data flows, operational resilience and dependencies on third parties must be demonstrably controlled through legislation.

The company risks having its algorithms tampered with and malicious use of its data, as well as other attacks to foil cyber-threat prediction. Further, in a hybrid work environment, securing the device that is being used and enabling access to secure different cloud services is difficult (NIST, 2023). Resilient Intellect Ltd must implement an integrated governance and security framework that enables continuous assurance, least-privilege access and resilience by design to stay competitive and compliant.

This chapter shows the organisation's hybrid AI environment security and compliance programme shaped by the Zero Trust principle. The framework brings together governance and technical architecture to improve usability and efficiency, empowering enterprise-wide risk management, compliance, and operational assurance. By using this plan, ResilientIntellect Ltd can show they are following the rules, they will not get a cyber incident, and the digital economy that changes fast will trust them.

2. Governance Principles and Design Framework.

ResilientIntellect Ltd will take a Zero Trust, defence-in-depth, risk-based governance approach, so nothing that is a user, a device, or a system component can be trusted by default. The concept of Zero Trust believes that the safety of the network cannot be assumed or taken for granted (Rose et al, 2020). All layers, from the ingestion of data to the deployment of models, must adhere to the principles of security by design and transparency^[24].

The structure design framework goes according to the CSF 2.0, which is from NIST and directly maps to the Identify, Protect, Detect, Respond and Recover Functions. Governance becomes operationalised through review and monitoring, measurement of risk metrics, and automated control enforcement. Monitoring of AI pipelines for signals of data drift and prevention of unauthorised access will occur. Likewise, compliance testing will be integrated into DevSecOps pipelines in the build and deploy stages.

The governance cycle integrates formal processes for mapping risks, evaluating ethics, and documenting accountability to align with ISO/IEC 42001 and the NIST AI Risk Management Framework. Every stage of the cycle is governed by policies that identify, escalate, and collect evidence. Regular reports to the board ensure transparency, while audit trails enable traceability in compliance domains such as GDPR and DORA.^[25]

Also, one is to share the responsibility through the design. The technical teams, compliance officers and officers manage risks through a federated mechanism. The control catalogue maps the organisation's assets to relevant regulatory and security requirements. When governance meets architecture, the company achieves compliance and operational assurance – turning security from a cost into a measurable function that enables growth and innovation.

3. Organizational Structure and Key Roles.

For effective regulation and governance, one needs a proper organisational structure of an enterprise. Resilient Intellect Ltd does not have a model that combines strategy, techniques and compliance. A Chief Information Security Officer (CISO) is an executive-level employee who also supervises the strategy, policy formulation, and regulatory engagement. The Head of Governance covers the DORA and NIS2 regulations and ensures the operational teams complete their audits on time, incident reports, and resilience assessments that are reported to the CISO.^[26]

A cloud security architect is responsible for designing secure configurations across cloud service providers, deploying Zero-Trust policies and retaining visibility of data flows across hybrid networks (NIST, 2023). A DevSecOps Lead supports a proactive approach to security, focusing on embedding security controls into the development lifecycle. This integration of security in build pipelines, especially invaluable in the training and deployment of AI models, prevents vulnerabilities from slipping through into production^[27].

The Operations Management at SOC is under the domain of the SOC Manager, who operates the teams engaged in endpoint detection, incident response, and automated threat intelligence feeds. A dedicated Compliance Officer serves as the contact point between the organization and regulatory bodies. They maintain the documentation for the audits of GDPR, DORA, and NIS2^[28].

To enhance accountability, the RACI (Responsible, Accountable, Consulted, Informed) matrix formalises everyone's role in the governance framework. This makes the decisions visible and the responsibilities traceable. The structure promotes teamwork between technical/engineering and policy groups so that all aspects of AI security governance are understood. This organizational model increases control, ownership and communication, with compliance part of continuous operations as opposed to an annual audit exercise.

4. Security Policies and Procedures.

Resilient Intellect Ltd manages its governance framework using security policies and procedures that deliver consistent legal, technical and behavioral standards. According to NIST (2023) the strategic governance principles converted into auditable and improvable control mechanisms are called policies.

The first cornerstone is the AI Usage and Data Governance Policy. This defines the acceptable usage of datasets, the protocols of model training, and the verification of AI outputs. It requires that data must be minimized, anonymous and traceable in accordance with GDPR and ISO/IEC 42001. Through bias testing and model-explainability documentation, ethical oversight mechanisms are built into the AI systems to meet the transparency obligations of the EU AI Act (European Commission, 2024).

Another important policy adopted by the ECB is the Secure Cloud and SaaS Configuration Policy, which is in line with Zero Trust principles and DORA's resilience requirements. According to EIOPA (2024), the guidelines state that configuration monitoring should be continuous in nature. Further, they require that data in transit and at rest should be encrypted. Lastly, identity-access management should be strict and on multi-factor authentication. To limit exposures to cloud security intrusions, periodic compliance scans and automated alerts are used to catch misconfigurations in real-time.

The policy for endpoint security and remote access is the Hybrid Workforce and BYOD. It needs device registration, endpoint detection, and response (EDR) installation; and ultimately, a division of our personal and corporate information through VPN. This makes sure that remote workers always do their work in the right settings.

The company imposes Change-Control, Incident Response, and Third Party Risk Management processes in compliance with the NIS2 and DORA regulations. Every time we update our system or vendor, we always check the security of the event (European Commission, 2023). After an incident, we will contain the incident, a sure notification will occur, and finally, we will review it.

All policies go through an annual review and an internal audit, which ensures it remains relevant as the regulations and technology evolve. Training and awareness programs strengthen the processes through the deeply embedded culture of security among employees. To sum up the above paragraphs, the policies establish the mechanics of governance that come with accountability, standardization and continuous assurance at Resilient Intellect Ltd (Floridi, 2024).

5. DORA, NIS2 and GDPR are all Compliance.

Regulatory compliance is a key element of Resilient Intellect Ltd's governance structure, ensuring compliance with European and international data protection, resilience and functional standards. The compliance approach brings together the Digital Operational Resilience Act (DORA), the Network and Information Systems Directive 2 (NIS2), and the General Data Protection Regulation (GDPR) into a single system for managing legal obligations and audit readiness (European Commission, 2024).

As per DORA, the firm must showcase continuous monitoring and operational resilience across all ICT systems. Resilient Intellect Ltd's compliance roadmap includes regular pen testing, threat-led resilience assessments and documentation of critical service dependency. In the event of a significant incident identified by the industry, the AI sub-component or tool must possess adequate resilience.

NIS2 raises these expectations by making incident reporting and supply-chain oversight stricter. To comply, the company keeps a registry of supplier risk and conducts continuous security assessments across the extended tech stack. Cloud service providers must have the same security and compliance standards by contract to protect downstream (ISO, 2024). The internal SOC team is trained to respond within report a critical cyber incident within 24 hours.

GDPR covers all personal data processed in the AI pipelines

The data used to train the model should be pseudonymized and collected based on consent. The data protection impact assessments (DPIAs) will be carried out regularly to ensure fairness and transparency, especially in automated decision-making systems. The organization's AI Use Policy backs up the guidelines by requiring bias checks and traceability logs for any model they release.

DORA, NIS2 and GDPR create a multilayered compliance structure that achieves a proper balance between resilience, privacy and ethical accountability. By using technical measures alongside regulatory controls, Resilient Intellect Ltd not only lowers its chances of incurring a fine but also further demonstrates governance to regulators and clients, and hence positions compliance as a competitive advantage.

6. Governance And Strategy In NIST CSF 2.0.

The security governance of ResilientIntellect Ltd is strategically framed using the NIST Cybersecurity Framework (CSF) 2.0. The five functions of Identify, Protect, Detect, Respond and Recover guide the enterprise's overall security posture in both the technical and the organisational (NIST 2023). This section translates those functions into strategies that align with the operations of an AI-driven company and the regulatory environment.

The organisation keeps a central record for cataloguing its AI models, datasets, endpoints, and APIs of third parties. Each asset gets a ranking based on how critical it is, and controls from DORA, NIS2, and GDPR are mapped to them. This process establishes risk visibility and ownership.

Shutting down the threat prevents any further disruption to our operations or other security actors. Metrics, such as mean time to patch and access-review compliance rates, measure how well protections are working.

The SOC employs constant monitoring using AI tools and regular SIEM analytics. Through this function, alerts are generated when risk tolerance thresholds are reached. Toward the end of training, indicators of data drift and model deviation are also included to maintain the integrity of AI performance.^[29]

The response framework takes playbooks based on incident class and severity. Automated containment tools isolate impacted services, and the GRC team oversees communication, root-cause analysis, and regulatory notifications within the required timeframes.

Business continuity is strengthened through regular simulations and redundant AI environments. The “recover” function helps with DORA’s requirements on restoring essential services and the principles of ISO/IEC 22301 for operations continuity.

CSF 2.0 improves governance integration, or the linking of technical control outcomes to management oversight and performance evaluation. ResilientIntellect Ltd’s governance strategy achieves this by embedding performance indicators for cybersecurity into management dashboards that boards see and on which they base their decisions.

By aligning itself with the other two workstreams, governance becomes less of a discrete compliance checklist and more of a continuous improvement process that protects innovation while mitigating systemic risk^[30].

7. Risk Assessment and Management.

ResilientIntellect Ltd combines quantitative and qualitative models to assess its risks. Each risk is assessed based on its operational, technological, or ethical nature (NIST, 2023). The places that colours represent across four areas - Data Governance, Model Reliability, Supply Chain Security, and Operational Resilience.

For those threats related to AI models, such as model poisoning or data drift, risk evaluation involves sensitivity analysis and scenario modelling to foresee failure points (Goodfellow et al., 2022). When serious risks are detected, automatic procedures kick in. For example, we can isolate models to prevent problems. Alternatively, they can retrain the model using verified data after the fact.

We evaluate third-party and cloud risks through supplier audits and ongoing security scoring. If vendors don’t follow the compliance standards, they are moved up for replacement or remediation. Internal risks are caused by human error or by insiders and are helped by access governance and training for staff (ISO, 2024).

A critical outcome of risk management is the linkage of every risk event to controls and accountable owners through the governance dashboards of the organization. This tracking makes it possible to be accountable and support reporting for DORA and NIS2. In this model, risk assessment becomes not simply reactive, but predictive—enabling the data-based intelligence and adaptive governance of resilience (Floridi, 2024).

8. Contingency and Disaster Recovery Planning.

A comprehensive Disaster Recovery and Business Continuity Plan (DR/BCP) is essential for services to function well during a cyber disruption. Resilient Intellect Ltd’s plan meets the DORA requirement for operational resilience and NIST SP 800-34’s contingency management guideline.

AI threat-detection platforms, customer databases, analytics pipelines, and other critical systems are replicated in geographically redundant cloud regions to failover. For each service tier, Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) are defined based on business impact analysis^[31].

The DR strategy has playbooks for ransomware containment and escalation protocols for incidents, which explain immediate actions to isolate, restore and communicate with stakeholders. Encrypted backups cannot be changed, and only some trusted people can access them. They are checked daily (EIOPA, 2024).

We regularly conduct tests to see if we can recover, including fake simulations and drills. The experience lessons taken from such exercises are inputted to post-incident reviews for upscaling resilience. Resilient Intellect Ltd ensures that corporate resilience constitutes an integral part of the operational fabric of the governance framework of the firm (Floridi, 2024). The goal is to maintain business sustainability and client confidence.

9. Protection Mechanisms and Control Implementation.

Resilient Intellect Ltd.'s Zero-Trust strategy is effective when protected by layered defences that address the human and technical weaknesses faced by organizations. The foremost relevant capabilities currently are EDR, XDR, DLP, and PAM (NIST, 2023), which are classified as key controls for security planning.

With EDR and XDR solutions gaining traction, security teams can more easily spot early anomalies and contain threats. DLP policies scan internal systems and external interfaces to automatically stop unauthorised transfers or downloads from occurring. PAM uses just-in-time access to limit insider-risk exposure and improve auditability.

Additionally, phishing-resistant multifactor authentication (MFA) and adaptive access control secure the hybrid workforce. Moreover, verification of biometrics for privileged accounts is integrated. AI is helpful to the existing threat intelligence solutions and is able to correlate internal telemetry and external feeds.

By embedding these mechanisms, adherence to operational controls in compliance with DORA and NIS2 is guaranteed. Technical mappings towards the NIST CSF 2.0 facilitate the measurement of performance continuously. In sum, these measures will create a resilient, always verifiable environment capable of evolving with the threats—it will be compliant and trustworthy (Floridi, 2024).

10. Program Roadmap and Assurance.

A phased roadmap is used to execute the Zero-Trust-aligned project. Establishment of Policy, Training and Initial Architectural Deployment, 1st Phase. Phase 2 is about continuous monitoring, automated compliance checks and dashboards for risk metrics. Phase 3 can achieve full maturity through audits at the organisation and third-party certification (ISO/IEC 42001) and ongoing optimisations based on threat intelligence and post-incident reviews.

To ensure assurance, quarterly governance reviews, internal compliance checks, and performance information by the executive are performed. Accredited auditors provide independent assurance and allow for the incorporation of lessons learned into governance. The use of a circular model converts compliance into continuous assurance, resulting in ResilientIntellect Ltd becoming the benchmark for data-centric accountability for cybersecurity governance (Floridi, 2024).

11. Summary Task 2.

The zero-trust aligned security and compliance programme proposed shall result in a coherent future-ready governance model for ResilientIntellect Ltd. The framework changes security from a safeguard in the background to an enabler of innovation, through an integrated approach of regulatory compliance, risk-based management and technical resilience. Integrating DORA, NIS2, GDPR, and NIST CSF 2.0 ensures accountability and operational transparency.

In the end, the success of the organisation in the EU market will depend on continuous verification, ethical AI use and cross-functional cooperation. ResilientIntellect Ltd can maintain trust, comply with regulations, and gain a competitive advantage in a more volatile digital world through adaptive governance and embedded resilience^[32].

Conclusion

The detailed study and recommendation for application show that the AI and Cybersecurity governance applied science depends on an integrated and risk-based approach in which zero-trust strategies and regulatory alignment play an important role. The report creates a model that achieves the right balance between innovation and accountability by drawing from the guidance provided under the NIST CSF 2.0, ISO/IEC 42001, EU DORA, NIS2, and GDPR. At ResilientIntellect Ltd, governance isn't about compliance but about constant verification, ethical oversight and resilience in action. Policies, technical controls and measurable assurance mechanisms will ensure the right governance becomes a dynamic system able to adapt to emerging threats and the law. In short, this alignment enhances organisational integrity, safeguards stakeholder trust and ensures sustainable growth of business in the data economy, where transparency, accountability and resilience drive a competitive advantage.

References

- ^[1] European Commission (2024) *Artificial Intelligence Act: Regulation of AI systems in the European Union*. Available at: <https://artificial-intelligence.europa.eu> (Accessed: 19 October 2025).
- ^[2] NIST (2023) *AI Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD: National Institute of Standards and Technology. Available at: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (Accessed: 19 October 2025).
- ^[3] European Commission (2024) *Artificial Intelligence Act: Regulation of AI systems in the European Union*. Available at: <https://artificial-intelligence.europa.eu> (Accessed: 19 October 2025).
- ^[4] Floridi, L. (2024) ‘AI governance: Ethics, accountability and regulatory transformation in the digital era’, *Philosophy & Technology*, 37(1), pp. 45–63.
- ^[5] Goodfellow, I., Bengio, Y. and Courville, A. (2022) *Deep Learning*. 2nd edn. Cambridge, MA: MIT Press.
- ^[6] Jagielski, M., Carlini, N., Berthelot, D., Kurakin, A. and Papernot, N. (2020) ‘High-confidence data poisoning attacks on neural networks’, *Proceedings of the IEEE Security and Privacy Symposium*, 38(2), pp. 43–58.
- ^[7] European Commission (2023) *Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2 Directive)*. Available at: <https://digital-strategy.ec.europa.eu/en/policies/nis2-directive> (Accessed: 20 October 2025).
- ^[8] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A. and Amodei, D. (2023) *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*. Oxford: Oxford University Press.
- ^[9] Shokri, R., Stronati, M., Song, C. and Shmatikov, V. (2021) ‘Membership inference attacks against machine learning models’, *IEEE Symposium on Security and Privacy*, 40(3), pp. 20–34.
- ^[10] ISO (2024) *ISO/IEC 27001:2022 Information Security, Cybersecurity and Privacy Protection*. Available at: <https://www.iso.org/standard/82875.html> (Accessed: 21 October 2025).
- ^[11] NIST (2023) *Special Publication 800-34 Rev. 1: Contingency Planning Guide for Federal Information Systems*. Available at: <https://csrc.nist.gov/publications> (Accessed: 19 October 2025).
- ^[12] NIST (2024) *Cybersecurity Framework 2.0 Concept Paper*. Available at: <https://www.nist.gov/cyberframework> (Accessed: 18 October 2025).

-
- ^[13] European Commission (2024) *Artificial Intelligence Act: Regulation of AI systems in the European Union*. Available at: <https://artificial-intelligence.europa.eu> (Accessed: 19 October 2025).
- ^[14] Floridi, L. (2024) 'AI governance: Ethics, accountability and regulatory transformation in the digital era', *Philosophy & Technology*, 37(1), pp. 45–63.
- ^[15] Mouzakitis, A., Nikolaidou, M. and Anagnostopoulos, D. (2023) 'AI governance frameworks and the ISO/IEC 42001 paradigm: Towards accountable artificial intelligence systems', *Journal of Information Systems Engineering*, 18(3), pp. 102–118.
- ^[16] European Commission (2025) *Digital Operational Resilience (DORA) Guidance for Financial Entities*. Available at: <https://finance.ec.europa.eu/dora> (Accessed: 22 October 2025).
- ^[17] Mouzakitis, A., Nikolaidou, M. and Anagnostopoulos, D. (2023) 'AI governance frameworks and the ISO/IEC 42001 paradigm: Towards accountable artificial intelligence systems', *Journal of Information Systems Engineering*, 18(3), pp. 102–118.
- ^[18] NIST (2024) *Cybersecurity Framework 2.0 Concept Paper*. Available at: <https://www.nist.gov/cyberframework> (Accessed: 18 October 2025).
- ^[19] Floridi, L. (2024) 'AI governance: Ethics, accountability and regulatory transformation in the digital era', *Philosophy & Technology*, 37(1), pp. 45–63.
- ^[20] Jagielski, M., Carlini, N., Berthelot, D., Kurakin, A. and Papernot, N. (2020) 'High-confidence data poisoning attacks on neural networks', *Proceedings of the IEEE Security and Privacy Symposium*, 38(2), pp. 43–58.
- ^[21] Shokri, R., Stronati, M., Song, C. and Shmatikov, V. (2021) 'Membership inference attacks against machine learning models', *IEEE Symposium on Security and Privacy*, 40(3), pp. 20–34.
- ^[22] NIST (2023) *AI Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD: National Institute of Standards and Technology. Available at: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (Accessed: 19 October 2025).
- ^[23] EIOPA (2024) *Digital Operational Resilience Act (DORA): Supervisory expectations and implementation roadmap*. Available at: <https://www.eiopa.europa.eu/document-library> (Accessed: 18 October 2025).
- ^[24] Institute of Standards and Technology. Available at: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (Accessed: 19 October 2025).
- ^[25] European Commission (2024) *Artificial Intelligence Act: Regulation of AI systems in the European Union*. Available at: <https://artificial-intelligence.europa.eu> (Accessed: 19 October 2025).
- ^[26] EIOPA (2024) *Digital Operational Resilience Act (DORA): Supervisory expectations and implementation roadmap*. Available at: <https://www.eiopa.europa.eu/document-library> (Accessed: 18 October 2025).
- ^[27] ISO (2024) *ISO/IEC 42001:2024 Artificial Intelligence — Management System*. Available at: <https://www.iso.org/standard/83110.html> (Accessed: 21 October 2025).
- ^[28] ISO (2024) *ISO/IEC 42001:2024 Artificial Intelligence — Management System*. Available at: <https://www.iso.org/standard/83110.html> (Accessed: 21 October 2025).
- ^[29] Shokri, R., Stronati, M., Song, C. and Shmatikov, V. (2021) 'Membership inference attacks against machine learning models', *IEEE Symposium on Security and Privacy*, 40(3), pp. 20–34.
- ^[30] Floridi, L. (2024) 'AI governance: Ethics, accountability and regulatory transformation in the digital era', *Philosophy & Technology*, 37(1), pp. 45–63.

^[31] ISO (2024) *ISO/IEC 42001:2024 Artificial Intelligence — Management System*. Available at: <https://www.iso.org/standard/83110.html> (Accessed: 21 October 2025).

^[32] European Commission (2025) *Digital Operational Resilience (DORA) Guidance for Financial Entities*. Available at: <https://finance.ec.europa.eu/dora> (Accessed: 22 October 2025).

AI and Machine Learning

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

75/100

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

NAOD HABTOM

75

ENGAGEMENT WITH LITERATURE

All work must be appropriately researched and referenced with an extensive range of relevant and credible literature. High-level referencing skills consistently applied. Most references should be current and Laws must cite current legislation in the appropriate government site(s). (20 marks)



very good engagement with recognised frameworks (NIST AI RMF, ISO/IEC 42001, EU AI Act). Sources are relevant and mostly appropriate. However, citation quality is inconsistent.

16

KNOWLEDGE AND UNDERSTANDING

You will need to show application of skills and research in the assignment by demonstrating critical understanding of the well-established concepts and principles of the subject and the main methods of enquiry. May go beyond established theories.(10 marks)



Demonstrates very good understanding of AI governance, Zero Trust, and EU regulatory drivers. Concepts are accurate and well articulated. Knowledge remains within expected boundaries

8

COGNITIVE AND INTELLECTUAL

Marks will be awarded: For critical analysis and evaluation, based on the quality of the documentation/report/paper. How relevant conclusions are drawn and justified. (25 marks)



There is clear analytical particularly in Task 1, with structured discussion of risks, governance frameworks, and regulatory overlap. Task 2 justification is very good but conservative, with limited exploration of alternative strategies or trade-offs.

18

PRACTICAL SKILLS

Marks will be awarded based on the quality of and level of detail provided for the design approach; diagrams, tables, charts and any other design methods utilise will enhance marks. These should be properly explained and contextualised. (25 marks)



Diagrams, and organization chart are provided in additional files. However, these should be part of the report, numbered, labelled and discussed.

15

TRANSFERABLE SKILLS FOR LIFE

This will be assessed based on the written and presentational quality of your report: • Flow • Professionalism • Comprehension (grammar and spelling) • Completeness and presentation of report (20 marks)



Very professionally done. IEEE two-column used.

18