INTRODUCTION TO AI RISK MANAGEMENT FRAMEWORKS

AND SAMPLE SECURITY CONTROLS

SkillWeed

TABLE OF CONTENTS

INTRODUCTION TO AI AND RISKS AROUND LARGE LANGUAGE MODELS (LLMS)

Artificial Intelligence (AI) has rapidly evolved over the past decade, revolutionizing various industries through advancements in machine learning and deep learning. Among the most significant developments in AI are Large Language Models (LLMs). These models, such as OpenAI's GPT-3 and GPT-4, Google's BERT, and others, are designed to understand and generate human-like text based on the vast amounts of data they have been trained on. LLMs have demonstrated remarkable capabilities in tasks ranging from natural language processing to generating coherent and contextually relevant content.

WHAT ARE LARGE LANGUAGE MODELS (LLMS)?

Large Language Models are a type of deep learning model that leverage massive datasets and substantial computational resources to understand and generate natural language. They are built on neural networks, particularly transformers, which enable them to capture complex patterns in text data. LLMs can perform a wide range of tasks, including text completion, translation, summarization, question answering, and even creative writing.

The training process for LLMs involves exposing them to diverse text sources, such as books, articles, websites, and other digital content. Through this extensive training, LLMs learn the statistical relationships between words and phrases, allowing them to generate highly coherent and contextually accurate text.

BENEFITS OF LLMS

1. **Enhanced Natural Language Processing (NLP)**: LLMs have significantly improved the accuracy and efficiency of NLP tasks, making applications like chatbots, virtual assistants, and language translation more effective.
2. **Automation and Efficiency**: By automating tasks such as content creation, customer service, and data analysis, LLMs can boost productivity and reduce operational costs.
3. **Innovation and Creativity**: LLMs have opened new avenues for innovation, enabling creative applications in fields like content generation, gaming, and personalized marketing.

RISKS AND CHALLENGES ASSOCIATED WITH LLMS

While LLMs offer numerous advantages, they also pose several risks and challenges that need to be carefully managed:

1. **Bias and Fairness**: LLMs can inadvertently perpetuate or amplify biases present in their training data, leading to biased or unfair outcomes. This is particularly concerning in applications such as hiring, lending, and law enforcement.

2. **Security and Privacy**: LLMs can generate content that mimics human behavior, making them susceptible to misuse for generating misleading information, phishing attacks, or other malicious activities. Additionally, they might inadvertently reveal sensitive information present in the training data.
3. **Ethical Concerns**: The use of LLMs raises ethical questions about accountability, transparency, and the potential for misuse. Ensuring that AI systems are used responsibly and ethically is a critical challenge.
4. **Hallucinations**: LLMs sometimes generate plausible-sounding but incorrect or nonsensical answers, known as "hallucinations." This can be problematic in applications requiring high accuracy and reliability, such as medical advice or legal information.
5. **Resource Intensity**: Training and deploying LLMs require substantial computational resources, which can be costly and environmentally taxing. The carbon footprint associated with large-scale AI models is an ongoing concern.
6. **Regulatory Compliance**: As the use of LLMs grows, so does the need for compliance with various regulations and standards related to data protection, privacy, and AI ethics. Navigating this complex regulatory landscape is a significant challenge for organizations.

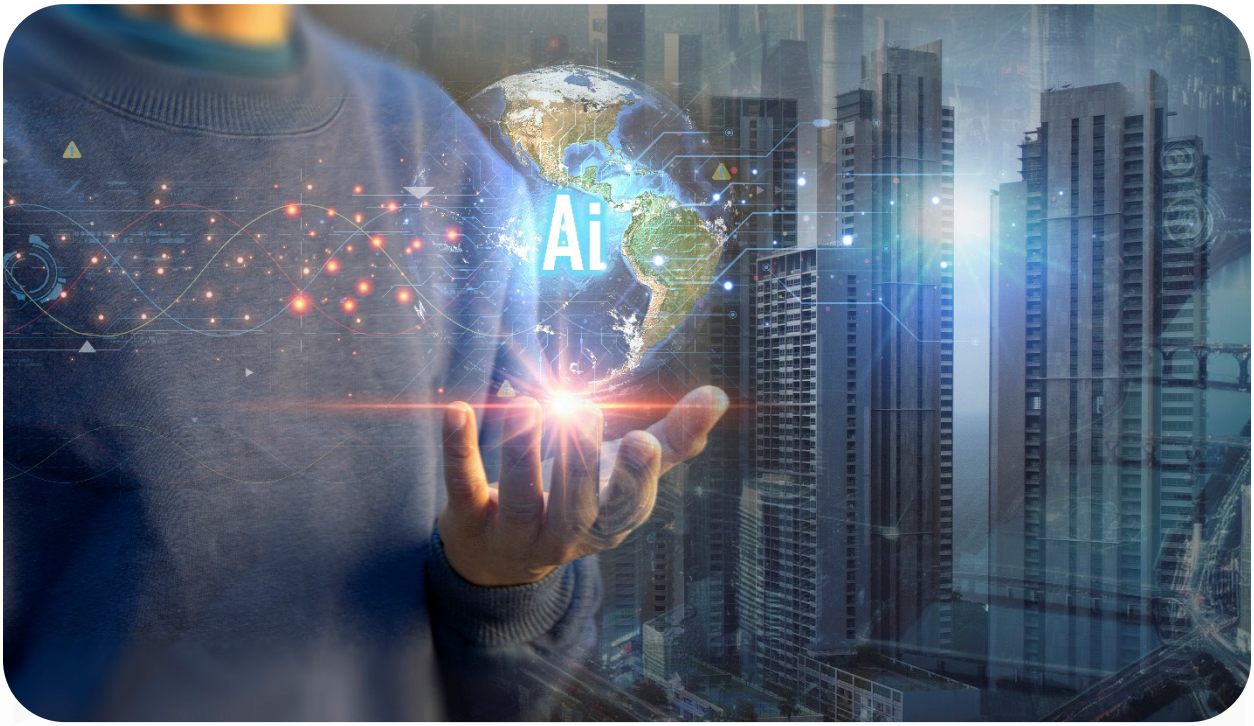
MITIGATING RISKS

To harness the benefits of LLMs while mitigating the associated risks, organizations should adopt comprehensive risk management strategies, including:

1. **Bias Mitigation**: Implementing techniques to identify and reduce biases in AI models, such as diverse and representative training data, bias audits, and fairness-aware algorithms.
2. **Security Measures**: Ensuring robust security protocols to protect against malicious use of AI-generated content and safeguarding sensitive information in training data.
3. **Transparency and Explainability**: Developing methods to make AI models more transparent and their decision-making processes explainable, enhancing trust and accountability.

4. **Ethical Guidelines**: Establishing ethical guidelines and governance frameworks to ensure responsible and ethical use of AI technologies.
5. **Regulatory Adherence**: Staying informed about and compliant with relevant regulations and standards, fostering collaboration with regulatory bodies and industry groups.
6. **Continuous Monitoring**: Regularly monitoring and evaluating AI systems to detect and address any emerging risks or issues.

INTRODUCTION TO AI RISK MANAGEMENT FRAMEWORKS



To manage the risks associated with LLMs and other AI systems effectively, several frameworks and guidelines have been developed by various organizations. These frameworks provide structured approaches to implementing, assessing, and maintaining AI systems securely and ethically. Here are twelve key frameworks:

1. **NIST AI Risk Management Framework**
    - A comprehensive framework developed by the National Institute of Standards and Technology to manage risks associated with AI, focusing on risk assessment, management, and mitigation.
2. **MITRE ATT&CK**
    - A globally accessible knowledge base of adversary tactics and techniques based on real-world observations, providing a framework for threat modeling, detection, and mitigation.
3. **ISO/IEC 27001 with AI Extensions**
    - An international standard for information security management systems, with extensions tailored for AI applications, emphasizing information security controls and risk management.

4. **AI in Cybersecurity Guidelines by ENISA**
    - Guidelines by the European Union Agency for Cybersecurity on the use of AI in cybersecurity, covering best practices, risk management, and ethical considerations.
5. **IEEE P7003**
    - A standard for Algorithmic Bias Considerations to address ethical concerns in AI applications, including cybersecurity, focusing on fairness, accountability, and transparency.
6. **OECD Principles on AI**
    - Guidelines by the Organisation for Economic Co-operation and Development to promote trustworthy AI, emphasizing ethical AI, robustness, security, transparency, and accountability.
7. **EU AI Act**
    - Proposed regulation by the European Union to ensure the safe and ethical use of AI across various sectors, including cybersecurity, adopting a risk-based approach and compliance requirements.
8. **CSA AI in Cybersecurity Working Group**
    - Guidelines and best practices developed by the Cloud Security Alliance for using AI in cybersecurity, covering cloud security, AI-based threat detection, and risk management.
9. **DHS AI Strategy for Cybersecurity**
    - The U.S. Department of Homeland Security's strategy for incorporating AI into cybersecurity efforts, focusing on national security, AI-driven threat detection, and incident response.
10. **AI Ethics Guidelines by the Partnership on AI**
    - Guidelines developed by a consortium of organizations to promote ethical AI usage, including in cybersecurity, emphasizing fairness, transparency, accountability, and best practices.

11. **AI Security Framework by Gartner**
    - A framework developed by Gartner to help organizations integrate AI into their cybersecurity strategies, focusing on AI governance, risk management, AI-driven security solutions, and best practices.
12. **ISO/IEC JTC 1/SC 42**
    - A subcommittee of ISO/IEC focusing on standardization in the area of AI, including its applications in cybersecurity, addressing AI terminology, risk management, ethical considerations, and standards.

These frameworks and guidelines provide comprehensive approaches to implementing, assessing, and maintaining AI systems securely and ethically, ensuring that organizations can leverage the power of AI while managing its inherent risks.

SUMMARY OF FRAMEWORK AND GUIDELINES

#	Framework/Guideline	Subsection	Description	Key Features
1	NIST AI Risk Management Framework	Risk Management, Trustworthiness	A framework developed by the National Institute of Standards and Technology to manage risks associated with AI.	Risk assessment and management, trustworthiness, reliability, security, privacy.
2	MITRE ATT&CK	Tactics, Techniques, Procedures (TTPs)	A globally accessible knowledge base of adversary tactics and techniques based on real-world observations.	Detailed adversary behavior, threat modeling, detection, and mitigation strategies.
3	ISO/IEC 27001 with AI Extensions	Information Security Management System (ISMS)	An international standard for information security management systems, with extensions for AI applications.	Information security controls, risk management, AI-specific guidelines.
4	AI in Cybersecurity Guidelines by ENISA	Best Practices, Risk Management	Guidelines by the European Union Agency for Cybersecurity on the use of AI in cybersecurity.	Best practices, risk management, ethical considerations, and AI deployment.
5	IEEE P7003	Algorithmic Bias Considerations	A standard for Algorithmic Bias Considerations to address ethical concerns in AI applications, including cybersecurity.	Fairness, accountability, transparency, and mitigation of bias in AI.

#	Framework/Guideline	Subsection	Description	Key Features
6	OECD Principles on AI	Ethical AI, Security	A set of guidelines by the Organisation for Economic Co-operation and Development to promote trustworthy AI.	Ethical AI, robustness, security, transparency, and accountability.
7	EU AI Act	Risk-based Approach, Compliance	Proposed regulation by the European Union to ensure the safe and ethical use of AI across various sectors, including cybersecurity.	Risk-based approach, compliance requirements, transparency, and accountability.
8	CSA AI in Cybersecurity Working Group	Cloud Security, AI-based Threat Detection	Guidelines and best practices developed by the Cloud Security Alliance for using AI in cybersecurity.	Cloud security, AI-based threat detection, incident response, and risk management.
9	DHS AI Strategy for Cybersecurity	National Security, Incident Response	The U.S. Department of Homeland Security's strategy for incorporating AI into cybersecurity efforts.	National security, AI-driven threat detection, and incident response.
10	AI Ethics Guidelines by the Partnership on AI	Ethical Considerations, Transparency	Guidelines developed by a consortium of organizations to promote ethical AI usage, including in cybersecurity.	Ethical considerations, transparency, accountability, and best practices.
11	AI Security Framework by Gartner	AI Governance, Risk Management	A framework developed by Gartner to help organizations integrate AI into their cybersecurity strategies.	AI governance, risk management, AI-driven security solutions, and best practices.

#	Framework/Guideline	Subsection	Description	Key Features
12	ISO/IEC JTC 1/SC 42	AI Standardization, Risk Management	A subcommittee of ISO/IEC focusing on standardization in the area of AI, including its applications in cybersecurity.	AI terminology, risk management, ethical considerations, and standards.

NIST AI RISK MANAGEMENT FRAMEWORK

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
1	NIST AI Risk Management Framework	**Access Control**: Implement MFA, role-based access control.	Risk Management, Trustworthiness	ISO/IEC 27001 A.9, CIS Control 5, GDPR Article 32	1. Verify that MFA is implemented and enforced for all critical systems. 2. Ensure role-based access controls are properly configured.	1. Access control policy document. 2. MFA logs. 3. User access review reports.	1. Implement MFA for all critical systems. 2. Review and update role-based access control configurations.
		Data Protection: Use encryption for data at rest and in transit.	Risk Management, Trustworthiness	ISO/IEC 27001 A.10, GDPR Article 32, SOC 2 Security	1. Check if data encryption methods meet industry standards. 2. Review encryption key management practices.	1. Data encryption policies. 2. Encryption certificates and logs. 3. Key management procedures.	1. Update encryption methods if necessary. 2. Improve key management practices to ensure security.
		Incident Response: Develop and test incident	Risk Management, Trustworthiness	ISO/IEC 27001 A.16, NIST SP 800-	1. Review the incident response plan for	1. Incident response plan document. 2.	1. Conduct regular incident response drills. 2. Update incident response

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		response plans.		61, CIS Control 19	completeness. 2. Check incident response test logs and after-action reports.	Test results and after-action reports.	plan based on test outcomes.
		Continuous Monitoring: Implement continuous monitoring tools.	Risk Management, Trustworthiness	NIST SP 800-137, ISO/IEC 27001 A.12, CIS Control 6	1. Verify that continuous monitoring tools are deployed and configured correctly. 2. Review monitoring alerts and logs.	1. Monitoring tool configuration settings. 2. Monitoring logs and alerts.	1. Ensure monitoring tools are up-to-date and properly configured. 2. Regularly review and act on monitoring alerts.

MITRE ATT&CK

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
2	MITRE ATT&CK	**Endpoint Detection and Response (EDR)**: Implement EDR solutions to detect and respond to threats.	Tactics, Techniques, Procedures (TTPs)	ISO/IEC 27001 A.12, NIST SP 800-137, CIS Control 6	1. Verify EDR solution deployment on all endpoints. 2. Test EDR capabilities to detect and respond to simulated threats.	1. EDR deployment reports. 2. Logs showing detected threats and response actions.	1. Deploy or update EDR solutions on all endpoints. 2. Train staff on EDR use and response protocols.
		Network Segmentation: Implement network segmentation to limit lateral movement.	Tactics, Techniques, Procedures (TTPs)	ISO/IEC 27001 A.13, NIST SP 800-41, CIS Control 1	1. Review network architecture for segmentation. 2. Test segmentation by attempting lateral movement in a controlled environment.	1. Network diagrams showing segmentation. 2. Results of lateral movement tests.	1. Redesign network architecture to enhance segmentation. 2. Implement access controls to enforce segmentation.
		User and Entity Behavior Analytics (UEBA):	Tactics, Techniques, Procedures (TTPs)	ISO/IEC 27001 A.12, NIST SP 800-137, CIS Control 5	1. Verify deployment of UEBA tools. 2. Review logs for detected	1. UEBA deployment reports. 2. Logs showing detected	1. Deploy or update UEBA tools. 2. Configure UEBA to monitor critical

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		Implement UEBA tools to detect anomalies.			anomalies and corresponding actions taken.	anomalies and responses.	assets and detect anomalies.
		Threat Intelligence Integration: Integrate threat intelligence feeds into security operations.	Tactics, Techniques, Procedures (TTPs)	ISO/IEC 27001 A.12, NIST SP 800-150, CIS Control 19	1. Verify integration of threat intelligence feeds. 2. Review usage of threat intelligence in incident response activities.	1. Configuration settings for threat intelligence integration. 2. Incident response reports showing use of threat intelligence.	1. Integrate updated threat intelligence feeds into security operations. 2. Train staff on leveraging threat intelligence in incident response.
		Incident Response: Develop and test incident response plans specific to TTPs in	Tactics, Techniques, Procedures (TTPs)	ISO/IEC 27001 A.16, NIST SP 800-61, CIS Control 19	1. Review incident response plans for TTPs coverage. 2. Conduct tabletop exercises	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover identified TTPs. 2. Regularly conduct and update tabletop exercises.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		MITRE ATT&CK.			simulating TTPs.		

ISO/IEC 27001 WITH AI EXTENSIONS

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
3	ISO/IEC 27001 with AI Extensions	**Access Control**: Implement role-based access control (RBAC) and AI-driven access analytics.	A.9 Access Control	NIST SP 800-53 AC-2, CIS Control 5	1. Verify implementation of RBAC policies. 2. Check AI-driven access analytics for anomalies.	1. Access control policy document. 2. Access logs and AI analytics reports.	1. Update RBAC policies as needed. 2. Enhance AI analytics for better anomaly detection.
		Data Encryption: Encrypt sensitive data at rest and in transit using AI-enhanced encryption methods.	A.10 Cryptographic Controls	NIST SP 800-53 SC-13, CIS Control 13	1. Verify data encryption methods meet industry standards. 2. Check encryption key management practices.	1. Encryption policies and procedures. 2. Encryption certificates and logs.	1. Implement or update encryption methods. 2. Improve key management practices.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		AI Model Risk Management: Implement controls for managing risks associated with AI models.	A.15 Supplier Relationships	NIST SP 800-53 SA-11, CIS Control 12	1. Review risk management policies for AI models. 2. Assess third-party AI model risk management practices.	1. AI model risk management policy document. 2. Third-party risk assessment reports.	1. Develop or update AI model risk management policies. 2. Improve third-party risk management practices.
		Incident Response: Develop AI-specific incident response plans.	A.16 Information Security Incident Management	NIST SP 800-61, CIS Control 19	1. Review AI-specific incident response plans. 2. Conduct tabletop exercises for AI incidents.	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover AI-specific incidents. 2. Regularly conduct tabletop exercises.
		Continuous Monitoring: Implement AI-driven continuous monitoring tools.	A.12 Operations Security	NIST SP 800-137, CIS Control 6	1. Verify deployment of AI-driven monitoring tools. 2. Review monitoring logs for anomalies.	1. Monitoring tool deployment reports. 2. Logs showing detected anomalies.	1. Ensure monitoring tools are up-to-date. 2. Train staff on interpreting AI-driven monitoring results.

AI IN CYBERSECURITY GUIDELINES BY ENISA

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
4	AI in Cybersecurity Guidelines by ENISA	**AI-Driven Threat Detection**: Implement AI-based threat detection systems to identify and mitigate threats.	Best Practices, Risk Management	NIST SP 800-94, ISO/IEC 27001 A.12, CIS Control 6	1. Verify deployment of AI-based threat detection systems. 2. Test the system with simulated threats to assess effectiveness.	1. Deployment reports of AI-based systems. 2. Logs showing detection and mitigation of threats.	1. Ensure AI-based threat detection systems are fully deployed. 2. Regularly update and test the detection algorithms.
		AI Model Explainability: Ensure AI models used in cybersecurity are interpretable and explainable.	Best Practices, Ethical Considerations	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for explainability. 2. Conduct assessments to ensure models can be interpreted.	1. Documentation of AI models. 2. Reports from explainability assessments.	1. Update AI models to enhance explainability. 2. Provide training on interpreting AI model outputs.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		Data Privacy and Security: Implement measures to ensure data used in AI models is secure and privacy is maintained.	Best Practices, Risk Management	GDPR Articles 25 and 32, ISO/IEC 27701	1. Verify data encryption and anonymization methods. 2. Review data access controls.	1. Data encryption and anonymization logs. 2. Access control policies.	1. Enhance data encryption and anonymization techniques. 2. Improve access control mechanisms.
		Bias and Fairness in AI Models: Implement controls to detect and mitigate bias in AI models.	Ethical Considerations, Best Practices	IEEE P7003, ISO/IEC TR 24027	1. Review AI models for potential biases. 2. Test models with diverse datasets to assess fairness.	1. Bias assessment reports. 2. Results from fairness testing.	1. Update AI models to reduce biases. 2. Regularly assess and mitigate biases in models.
		Incident Response: Develop and test AI-specific incident response plans.	Best Practices, Risk Management	NIST SP 800-61, ISO/IEC 27001 A.16	1. Review AI-specific incident response plans. 2. Conduct tabletop exercises for AI incidents.	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover AI-specific incidents. 2. Regularly conduct tabletop exercises.

IEEE P7003

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
5	IEEE P7003	**Bias Detection and Mitigation**: Implement controls to detect and mitigate bias in AI models.	Algorithmic Bias Considerations	ISO/IEC TR 24027, NIST SP 800-53 SA-11, GDPR Article 22	1. Conduct bias assessment on AI models. 2. Use diverse datasets to test AI models for fairness.	1. Bias assessment reports. 2. Test results showing fairness metrics.	1. Update AI models to reduce identified biases. 2. Implement continuous monitoring for bias.
		Transparency and Explainability: Ensure AI models are interpretable and decisions can be explained.	Algorithmic Bias Considerations	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for transparency. 2. Test the explainability of AI decisions.	1. Documentation of AI models. 2. Reports from transparency assessments.	1. Update AI models to enhance transparency. 2. Provide training on interpreting AI model outputs.
		Ethical AI Usage: Implement guidelines for ethical use of AI in cybersecurity.	Algorithmic Bias Considerations	OECD Principles on AI, ISO/IEC 27001 A.18	1. Review policies for ethical AI usage. 2. Conduct regular ethics audits of AI systems.	1. Policies on ethical AI usage. 2. Ethics audit reports.	1. Develop or update ethical AI usage policies. 2. Conduct training on ethical AI practices.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		Data Governance: Ensure proper data governance practices for datasets used in AI.	Algorithmic Bias Considerations	ISO/IEC 27701, GDPR Article 5	1. Verify data governance policies. 2. Review data usage and management practices.	1. Data governance policy documents. 2. Data management logs.	1. Improve data governance policies. 2. Regularly review and update data management practices.
		AI Risk Management: Implement risk management processes for AI applications.	Algorithmic Bias Considerations	ISO/IEC 31000, NIST SP 800-37	1. Review AI risk management policies. 2. Conduct risk assessments for AI applications.	1. AI risk management policy documents. 2. Risk assessment reports.	1. Update AI risk management policies. 2. Implement mitigation measures for identified risks.

OECD PRINCIPLES ON AI

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
6	OECD Principles on AI	**Robustness and Security**: Ensure AI systems are robust and secure.	Security, Robustness	ISO/IEC 27001 A.12, NIST SP 800-53, CIS Control 1	1. Conduct security assessments of AI systems. 2. Test AI robustness against various attacks.	1. Security assessment reports. 2. Logs from robustness tests.	1. Enhance security measures for AI systems. 2. Regularly test and update robustness measures.
		Transparency and Accountability: Ensure AI systems are transparent and accountable.	Transparency, Accountability	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI documentation for transparency. 2. Verify accountability measures for AI decisions.	1. AI system documentation. 2. Accountability logs and reports.	1. Update AI systems to improve transparency. 2. Implement accountability mechanisms for AI decisions.
		Fairness and Non-discrimination: Implement measures to ensure AI systems are fair and non-discriminatory.	Ethical AI, Fairness	IEEE P7003, ISO/IEC TR 24027, GDPR Article 22	1. Test AI systems for bias and discrimination. 2. Review fairness policies and practices.	1. Bias testing reports. 2. Fairness policy documents.	1. Update AI systems to mitigate biases. 2. Regularly review and update fairness practices.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		Human-centric AI: Ensure AI systems prioritize human well-being and values.	Ethical AI, Human-centric AI	ISO/IEC 27552, GDPR Article 25	1. Review policies for human-centric AI development. 2. Conduct impact assessments on human well-being.	1. Human-centric AI policy documents. 2. Impact assessment reports.	1. Develop or update human-centric AI policies. 2. Implement measures to prioritize human well-being in AI.
		AI Governance and Management: Implement governance frameworks for AI systems.	Governance, Management	ISO/IEC 38500, NIST			

EU AI ACT

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
7	EU AI Act	**Risk Management**: Implement risk management processes for AI applications.	Risk-based Approach, Compliance	ISO/IEC 31000, NIST SP 800-37	1. Review AI risk management policies. 2. Conduct risk assessments for AI applications.	1. AI risk management policy documents. 2. Risk assessment reports.	1. Update AI risk management policies. 2. Implement mitigation measures for identified risks.
		Transparency and Documentation: Ensure comprehensive documentation and transparency of AI systems.	Transparency, Accountability	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review documentation for AI systems. 2. Verify transparency of AI decision-making processes.	1. AI system documentation. 2. Transparency reports.	1. Update documentation practices. 2. Implement measures to improve transparency in AI systems.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		Data Governance: Implement stringent data governance practices to ensure data integrity and privacy.	Data Quality, Data Governance	GDPR Articles 5 and 32, ISO/IEC 27701	1. Verify data governance policies and practices. 2. Conduct data quality and privacy assessments.	1. Data governance policy documents. 2. Data quality assessment reports.	1. Improve data governance policies. 2. Enhance data privacy

CSA AI IN CYBERSECURITY WORKING GROUP

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
8	CSA AI in Cybersecurity Working Group	**AI-Driven Threat Detection and Response**: Implement AI-based systems for detecting and responding to threats.	Cloud Security, AI-based Threat Detection	NIST SP 800-94, ISO/IEC 27001 A.12, CIS Control 6	1. Verify deployment of AI-based threat detection systems. 2. Test system effectiveness with simulated threats.	1. Deployment reports. 2. Logs showing threat detection and response actions.	1. Ensure full deployment of AI-based threat detection systems. 2. Regularly update and test detection algorithms.
		Data Privacy and Protection: Implement measures to ensure data used in AI models is secure and privacy is maintaine.	Cloud Security, Risk Management	GDPR Articles 25 and 32, ISO/IEC 27701	1. Verify data encryption and anonymization methods. 2. Review data access controls.	1. Data encryption and anonymization logs. 2. Access control policies.	1. Enhance data encryption and anonymization techniques. 2. Improve access control mechanisms.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		AI Governance: Implement governance frameworks to oversee the use and management of AI systems.	Cloud Security, Governance	ISO/IEC 38500, NIST SP 800-37	1. Review AI governance policies. 2. Assess governance practices for AI systems.	1. AI governance policy documents. 2. Governance assessment reports.	1. Develop or update AI governance policies. 2. Implement robust governance frameworks.
		Transparency and Explainability: Ensure AI models used in cybersecurity are interpretable and explainable.	Cloud Security, Best Practices	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for explainability. 2. Conduct assessments to ensure models can be interpreted.	1. Documentation of AI models. 2. Reports from explainability assessments.	1. Update AI models to enhance explainability. 2. Provide training on interpreting AI model outputs.
		Incident Response: Develop and test incident response plans specific to AI-related incidents.	Cloud Security, Incident Response	NIST SP 800-61, ISO/IEC 27001 A.16	1. Review AI-specific incident response plans. 2. Conduct tabletop exercises for AI incidents.	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover AI-specific incidents. 2. Regularly conduct tabletop exercises.

DHS AI STRATEGY FOR CYBERSECURITY

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
9	DHS AI Strategy for Cybersecurity	**AI-Driven Threat Detection and Response**: Implement AI-based systems to detect and respond to cyber threats.	National Security, Incident Response	NIST SP 800-94, ISO/IEC 27001 A.12, CIS Control 6	1. Verify deployment of AI-based threat detection systems. 2. Test system effectiveness with simulated threats.	1. Deployment reports. 2. Logs showing threat detection and response actions.	1. Ensure full deployment of AI-based threat detection systems. 2. Regularly update and test detection algorithms.
		Data Privacy and Protection: Ensure data used in AI models is secure and privacy is maintained.	Data Security, Privacy	GDPR Articles 25 and 32, ISO/IEC 27701	1. Verify data encryption and anonymization methods. 2. Review data access controls.	1. Data encryption and anonymization logs. 2. Access control policies.	1. Enhance data encryption and anonymization techniques. 2. Improve access control mechanisms.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		AI Governance: Implement governance frameworks for AI systems in cybersecurity.	Governance, Risk Management	ISO/IEC 38500, NIST SP 800-37	1. Review AI governance policies. 2. Assess governance practices for AI systems.	1. AI governance policy documents. 2. Governance assessment reports.	1. Develop or update AI governance policies. 2. Implement robust governance frameworks.
		Transparency and Explainability: Ensure AI models in cybersecurity are interpretable and explainable.	Transparency, Accountability	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for explainability. 2. Conduct assessments to ensure models can be interpreted.	1. Documentation of AI models. 2. Reports from explainability assessments.	1. Update AI models to enhance explainability. 2. Provide training on interpreting AI model outputs.
		Incident Response: Develop and test incident response plans specific to AI-related incidents.	Incident Response, National Security	NIST SP 800-61, ISO/IEC 27001 A.16	1. Review AI-specific incident response plans. 2. Conduct tabletop exercises for AI incidents.	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover AI-specific incidents. 2. Regularly conduct tabletop exercises.

AI ETHICS GUIDELINES BY THE PARTNERSHIP ON AI

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
10	AI Ethics Guidelines by the Partnership on AI	**Fairness and Bias Mitigation**: Implement controls to detect and mitigate bias in AI models.	Ethical Considerations, Fairness	IEEE P7003, ISO/IEC TR 24027, GDPR Article 22	1. Conduct bias assessments on AI models. 2. Test models with diverse datasets to ensure fairness.	1. Bias assessment reports. 2. Test results showing fairness metrics.	1. Update AI models to reduce identified biases. 2. Implement continuous monitoring for bias.
		Transparency and Explainability: Ensure AI models are interpretable and decisions can be explained.	Transparency, Accountability	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for transparency. 2. Test the explainability of AI decisions.	1. Documentation of AI models. 2. Reports from transparency assessments.	1. Update AI models to enhance transparency. 2. Provide training on interpreting AI model outputs.
		Human-Centric Design: Ensure AI systems prioritize human well-being and ethical considerations.	Ethical Considerations, Human-Centric AI	ISO/IEC 27552, GDPR Article 25	1. Review policies for human-centric AI development. 2.		

AI SECURITY FRAMEWORK BY GARTNER

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
11	AI Security Framework by Gartner	**AI-Driven Threat Detection**: Implement AI-based systems for detecting and responding to cyber threats.	AI Governance, Risk Management	NIST SP 800-94, ISO/IEC 27001 A.12, CIS Control 6	1. Verify deployment of AI-based threat detection systems. 2. Test system effectiveness with simulated threats.	1. Deployment reports. 2. Logs showing threat detection and response actions.	1. Ensure full deployment of AI-based threat detection systems. 2. Regularly update and test detection algorithms.
		Data Privacy and Protection: Ensure data used in AI models is secure and privacy is maintained.	AI Governance, Data Security	GDPR Articles 25 and 32, ISO/IEC 27701	1. Verify data encryption and anonymization methods. 2. Review data access controls.	1. Data encryption and anonymization logs. 2. Access control policies.	1. Enhance data encryption and anonymization techniques. 2. Improve access control mechanisms.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		AI Governance: Implement governance frameworks for AI systems in cybersecurity.	AI Governance, Risk Management	ISO/IEC 38500, NIST SP 800-37	1. Review AI governance policies. 2. Assess governance practices for AI systems.	1. AI governance policy documents. 2. Governance assessment reports.	1. Develop or update AI governance policies. 2. Implement robust governance frameworks.
		Transparency and Explainability: Ensure AI models in cybersecurity are interpretable and explainable.	AI Governance, Best Practices	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for explainability. 2. Conduct assessments to ensure models can be interpreted.	1. Documentation of AI models. 2. Reports from explainability assessments.	1. Update AI models to enhance explainability. 2. Provide training on interpreting AI model outputs.
		Incident Response: Develop and test incident response plans specific to AI-related incidents.	AI Governance, Incident Response	NIST SP 800-61, ISO/IEC 27001 A.16	1. Review AI-specific incident response plans. 2. Conduct tabletop exercises for AI incidents.	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover AI-specific incidents. 2. Regularly conduct tabletop exercises.

ISO/IEC JTC 1/SC 42

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
12	ISO/IEC JTC 1/SC 42	**AI Risk Management**: Implement risk management processes for AI applications.	AI Standardization, Risk Management	ISO/IEC 31000, NIST SP 800-37	1. Review AI risk management policies. 2. Conduct risk assessments for AI applications.	1. AI risk management policy documents. 2. Risk assessment reports.	1. Update AI risk management policies. 2. Implement mitigation measures for identified risks.
		Transparency and Explainability: Ensure AI models are interpretable and decisions can be explained.	AI Terminology, AI Risk Management	ISO/IEC 27001 A.18, NIST SP 800-53 SA-15	1. Review AI model documentation for transparency. 2. Test the explainability of AI decisions.	1. Documentation of AI models. 2. Reports from transparency assessments.	1. Update AI models to enhance transparency. 2. Provide training on interpreting AI model outputs.
		Data Privacy and Security: Implement	AI Risk Management, AI Standardization	GDPR Articles 25 and 32,	1. Verify data encryption and anonymization methods. 2.	1. Data encryption and anonymization	1. Enhance data encryption and anonymization techniques. 2.

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		measures to ensure data used in AI models is secure and privacy is maintained.		ISO/IEC 27701	Review data access controls.	logs. 2. Access control policies.	Improve access control mechanisms.
		Bias and Fairness in AI Models: Implement controls to detect and mitigate bias in AI models.	AI Standardization, Ethical Considerations	IEEE P7003, ISO/IEC TR 24027, GDPR Article 22	1. Review AI models for potential biases. 2. Test models with diverse datasets to assess fairness.	1. Bias assessment reports. 2. Results from fairness testing.	1. Update AI models to reduce biases. 2. Regularly assess and mitigate biases in models.
		Incident Response: Develop and test incident response plans specific to	AI Risk Management, AI Standardization	NIST SP 800-61, ISO/IEC 27001 A.16	1. Review AI-specific incident response plans. 2. Conduct tabletop	1. Incident response plans. 2. Results of tabletop exercises.	1. Update incident response plans to cover AI-specific incidents. 2. Regularly

#	Framework/Guideline	Security Controls	Sub Area	Mapping	Testing Steps	Evidence Required	Remediation
		AI-related incidents.			exercises for AI incidents.		conduct tabletop exercises.

CONCLUSION

The advancements in Artificial Intelligence, particularly through the development of Large Language Models (LLMs), have revolutionized many aspects of our society and industries, enhancing capabilities in natural language processing, automation, and innovation. However, these advancements come with significant risks and challenges, including biases, security threats, ethical dilemmas, and regulatory compliance issues.

Effectively managing these risks is crucial to harnessing the full potential of AI while ensuring that its deployment remains safe, ethical, and beneficial. This involves implementing robust security controls, maintaining transparency and explainability, safeguarding data privacy, and addressing biases and fairness in AI models.

To support organizations in navigating these complexities, numerous frameworks and guidelines have been developed by leading institutions and regulatory bodies. These frameworks provide structured approaches to AI risk management, offering best practices, standards, and detailed recommendations to ensure the responsible use of AI technologies. From the NIST AI Risk Management Framework to the EU AI Act and the AI Ethics Guidelines by the Partnership on AI, each framework brings valuable insights and strategies tailored to different aspects of AI governance, security, and ethics.

In conclusion, the integration of AI into various domains presents both immense opportunities and substantial risks. By adhering to established frameworks and guidelines, organizations can mitigate these risks, promote trust and accountability, and leverage AI to drive innovation and efficiency. Continuous monitoring, ethical governance, and proactive risk management are essential to ensuring that AI technologies contribute positively to society while minimizing potential adverse impacts. As AI continues to evolve, so too must our strategies for managing its risks, ensuring a balanced approach that maximizes benefits while safeguarding against threats.


SkillWeed