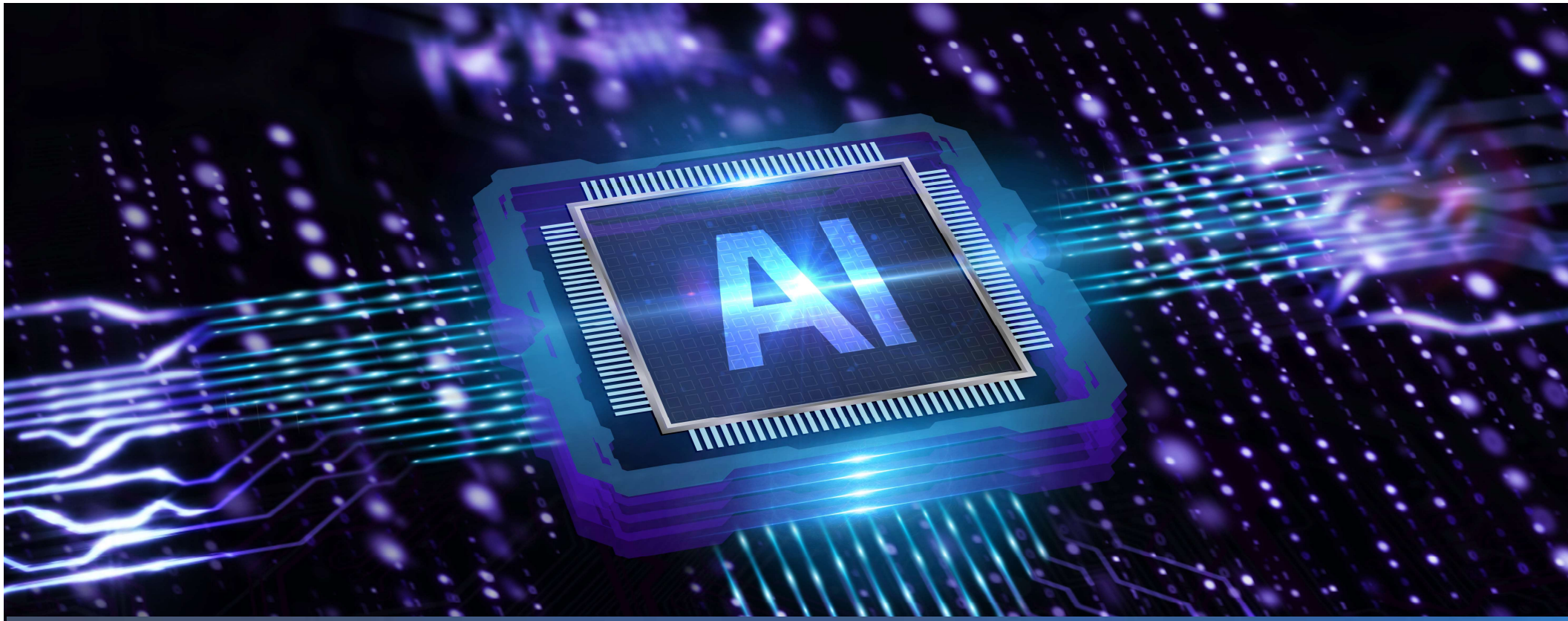


Job Aid 1 - AI Security Assessment Framework by Tonex

TONEX

TONEX AI SECURITY ASSESSMENT FRAMEWORK FOR MITIGATING RISKS IN LLM AND GENAI

EDITOR: CHARLES ALEXI

ASSESSMENT QUESTIONS FOR SECURITY TEAM MEMBERS TO ASK THEIR CUSTOMERS

1. Assessing the security posture of customers before they use AI-powered tools like Generative AI (GenAI) and Large Language Models (LLMs) for applications such as code generation
    1. To ensure that they understand and mitigate any associated risks

SCENARIOS BASED ON THE VULNERABILITIES OUTLINED IN OWASP LLM TOP 10 AND HOW ATTACKERS MIGHT EXPLOIT THEM

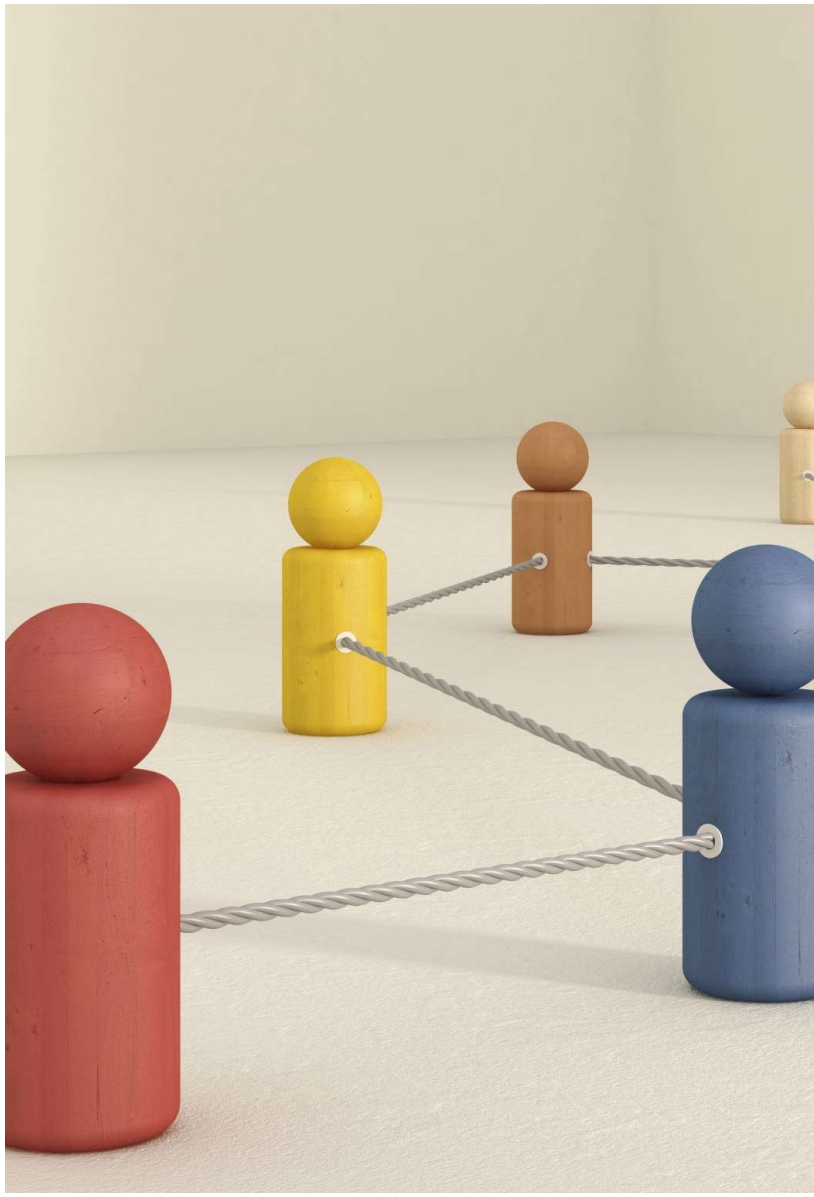
1. **Insecure Output Handling**: Attackers could exploit weaknesses in how an LLM handles its outputs. For instance, they might trick the LLM into disclosing confidential data or performing actions that compromise security, such as bypassing access controls.
2. **Training Data Poisoning**: By tampering with the training data of an LLM, attackers could introduce biases or vulnerabilities. This could lead to the LLM generating inaccurate or harmful outputs, affecting decision-making processes within a Business Unit.
3. **Model Denial of Service**: Attackers might target the resource-intensive nature of LLMs to cause service degradation or disruption. This could be achieved through sending a large number of requests or inputs that overload the LLM's processing capabilities.
4. **Supply Chain Vulnerabilities**: Exploiting vulnerabilities in third-party components or services used by an LLM can lead to security compromises. For example, an attacker might compromise a plugin or dataset used by the LLM to gain unauthorized access or manipulate its behavior.
5. **Sensitive Information Disclosure**: Attackers could exploit flaws in an LLM's logic or configuration to reveal sensitive information in its responses. This could include disclosing personally identifiable information (PII) or proprietary data.
6. **Insecure Plugin Design**: Vulnerabilities in LLM plugins could be exploited by attackers to gain unauthorized access or execute malicious code within the LLM environment. For instance, a plugin with insufficient access controls could be leveraged for privilege escalation attacks.
7. **Excessive Agency**: If an LLM has excessive permissions or autonomy, attackers might abuse these privileges to perform unauthorized actions or manipulate its outputs for malicious purposes.
8. **Overreliance**: Attackers could exploit situations where Business Units overly depend on LLMs without proper oversight or validation. This could lead to the dissemination of misinformation, or the execution of incorrect actions based on flawed LLM outputs.
9. **Model Theft**: Unauthorized access to proprietary LLM models could result in economic losses, compromised competitive advantage, or the exposure of sensitive information.

TRADITIONAL OWASP TOP 10 RISKS MIGHT TRANSLATE TO CONCERNS WITH GENAI/LLM

1. **Injection Flaws**: An LLM might be vulnerable to injection attacks if an attacker can feed it malicious input in the form of crafted prompts that could cause the model to output sensitive data, execute unintended actions, or reveal parts of its training data (data leakage).
2. **Broken Authentication**: If an LLM-based service doesn't properly implement authentication, attackers could gain unauthorized access to the model and use it for malicious purposes, including generating harmful content or accessing restricted functionalities.
3. **Sensitive Data Exposure**: An LLM might inadvertently expose sensitive information if it's been trained on datasets containing such information and hasn't been properly sanitized.
4. **XML External Entities (XXE)**: This risk is less relevant to LLMs as it pertains to older web services that parse XML input. However, if an LLM is part of a service that processes XML, it might be an indirect vector for XXE attacks.
5. **Broken Access Control**: LLMs must have strict access controls to prevent unauthorized users from accessing admin-level functions or generating content that could lead to harm or misuse.
6. **Security Misconfiguration**: An LLM that is not configured with security in mind may leak information through verbose error messages or debugging information, or it might allow access to administrative interfaces.
7. **Cross-Site Scripting (XSS)**: If an LLM-generated content is displayed on web applications without proper output encoding, it could be used as a vector for XSS attacks.
8. **Insecure Deserialization**: In the context of LLMs, insecure deserialization could occur if the model's parameters or state are manipulated before being loaded into an application, leading to arbitrary code execution or other attacks.
9. **Using Components with Known Vulnerabilities**: If the LLM or any associated software components have known vulnerabilities that aren't patched, it could be exploited by attackers.
10. **Insufficient Logging & Monitoring**: Without proper logging and monitoring, abusive or anomalous interactions with the LLM might not be detected or responded to in a timely manner.

KNOWN ATTACK VECTORS AND RISKS

1. **Data Poisoning Attacks:**
    1. Attackers influenced the training process of an LLM or GenAI system by introducing poisoned data into the training dataset. This led the model to learn and replicate unwanted biases or to respond to specific inputs in a certain way.
2. **Model Inversion Attacks:**
    1. By carefully crafting inputs and analyzing the outputs of LLMs, attackers attempted to reverse-engineer aspects of the training data, potentially exposing sensitive information.
3. **Adversarial Attacks:**
    1. Similar to adversarial attacks on image recognition systems, attackers devised inputs that are specially designed to trick the LLM into generating incorrect or harmful outputs.
4. **Misuse of Generated Content:**
    1. This isn't an attack on the model itself, but rather a misuse of the technology. LLMs and GenAI were used to create convincing phishing emails, fake news, or propaganda.
5. **Output Manipulation via Prompt Injection:**
    1. Attacker has access to the prompt or input for an LLM, they were able to manipulate the system into generating outputs that included malicious content, phishing links or malware.
6. **Evasion of Content Filters:**
    1. Attackers used LLMs to generate content that was designed to evade detection by content filters, enabling the spread of spam, malware harmful content.



PREVENTATIVE MEASURES

When these technologies are employed in security-sensitive environments, it's crucial for developers and businesses to be aware of such potential misuse and take preventative measures, such as:

Implementing rigorous data sanitization and validation practices.	Monitoring the use of the models to detect and respond to malicious activity.	Applying adversarial training techniques to make the models more robust against attacks.	Limiting the exposure of sensitive data during training and inference.	Employing strong access controls and user authentication to prevent unauthorized use.

The exact methods and mechanisms of any such attacks depend on the attackers' specific objectives and the vulnerabilities they seek to exploit. As these technologies become more prevalent, the security community will likely uncover and document more instances of LLM and GenAI exploitation.

BASIC QUESTIONS THAT A SECURITY TEAM MIGHT ASK THEIR CUSTOMERS

1. What is your current understanding of Generative AI and LLMs?
2. Have you used AI-powered tools for code generation or other purposes before?
3. What are your primary objectives for using these AI tools?
4. How do you plan to integrate AI-generated code into your development processes?

CASE STUDY 1 - CODE GENERATION

1. What measures do you have in place to review and validate the code generated by AI?
2. Are you aware of the potential risks of using AI-generated code, such as security vulnerabilities or biases?
3. What is your process for updating and maintaining code that includes AI-generated components?
4. How do you ensure that the data used to train or prompt the AI is free from sensitive information?
5. Do you have a designated individual responsible for AI cybersecurity concerns?

CASE STUDY 1: INCIDENT RESPONSE PLAN

1. Can you describe your incident response plan in the event of a security issue related to AI-generated content?
2. How do you plan to stay informed about the evolving landscape of AI and its implications for security?
3. What is your policy regarding the use of open-source code generated by AI in your projects?
4. How will you ensure compliance with regulations and standards applicable to AI-generated code?
5. Can you detail your data governance practices concerning the data used by AI tools?
6. What steps will you take to educate your team about secure practices when using AI for code generation?

STAKEHOLDERS

1. Security and Risk Management Team Members
2. Blue and Red Team Members
3. AI Team Members (Business and Technology)
4. AI Governance and Ethics
5. Systems Engineers
6. Software Developers
7. Product Managers
8. Project Managers
9. Portfolio Managers

TASK 1: ANALYZE AND ASSESS STAKEHOLDERS

1. Security and Risk Management Team Members
2. Blue and Red Team Members
3. AI Team Members (Business and Technology)
4. AI Governance and Ethics
5. Systems Engineers
6. Software Developers
7. Product Managers
8. Project Managers
9. Portfolio Managers
10. Vendor manager
11. Legal

TASK 2: OWASP LLM TOP 10 ASSESSMENT QUESTIONS

1. How familiar are you with the OWASP LLM Top 10 vulnerabilities and their impact on AI systems?
2. What steps do you take to prevent prompt injection attacks in AI-based applications?
3. How do you handle insecure output handling to prevent XSS, CSRF, and other security risks?
4. Can you explain the measures taken to detect and mitigate training data poisoning in AI models?
5. How do you protect AI models from denial-of-service (DoS) attacks and resource exhaustion?
6. What strategies are in place to mitigate supply chain vulnerabilities in AI solutions?
7. How do you prevent sensitive information disclosure in AI-generated outputs?
8. Can you describe the design considerations to ensure secure plugins and prevent remote code execution?
9. What measures are in place to prevent excessive agency and unintended consequences in AI systems?
10. How do you address overreliance on AI systems and ensure appropriate human oversight and intervention?

TASK 3: MITRE ATLAS INSPIRED ASSESSMENT QUESTIONS (GENERIC)

1. How do you assess and defend against reconnaissance tactics targeting AI systems?
2. What strategies are implemented to prevent resource development attacks in AI environments?
3. Can you describe the measures taken to secure initial access to AI systems and data?
4. How do you protect AI model access from unauthorized users or malicious actors?
5. What measures are in place to detect and respond to execution-based attacks targeting AI systems?
6. How do you ensure persistence and availability of AI services while preventing attacks?
7. Can you explain the strategies for detecting and mitigating privilege escalation in AI environments?
8. What measures are in place to evade AI system defenses and detect evasion tactics?
9. How do you protect against credential access attacks targeting AI systems and data?

TASK 3: MITRE ATLAS INSPIRED ASSESSMENT QUESTIONS – PART 1

1. Can you describe your reconnaissance tactics as they pertain to securing large language models (LLMs)?
2. How do you develop and manage resources related to LLM security?
3. What measures do you have in place to prevent unauthorized initial access to your LLM infrastructure?
4. How do you control access to and secure your ML model data?
5. What strategies do you employ for secure execution of LLM-related tasks and processes?
6. How do you ensure persistence and continuity in LLM security measures?
7. What mechanisms do you have for detecting and preventing privilege escalation attacks against your LLMs?
8. Can you describe your defense evasion techniques specific to protecting LLMs?
9. How do you secure credentials used in LLM-related operations and access?
10. What methods do you use for discovery and identification of potential LLM-related vulnerabilities?
11. How do you collect and analyze data related to LLM security incidents?
12. Can you describe your strategies for staging ML attacks against LLMs for testing and validation?
13. What measures do you have in place to prevent exfiltration of sensitive LLM-related data?
14. How do you assess the impact of security incidents or attacks on your LLM infrastructure?
15. Are there specific threat modeling techniques you use for LLM security?

TASK 3: MITRE ATLAS TACTICS - PART 2

1. Do you regularly update and patch your LLMs and related systems to address known vulnerabilities?
2. How do you conduct security testing and validation of your LLMs before deployment?
3. What incident response procedures do you have in place for LLM-related security events?
4. Can you describe your access control mechanisms for LLM-related resources and data?
5. How do you monitor and analyze LLM-related logs and events for security purposes?
6. What encryption methods do you use to protect data processed by LLMs?
7. How do you manage, and secure API endpoints used by LLMs for data exchange?
8. Can you provide examples of security awareness training specific to LLM-related risks for your personnel?
9. How do you handle security incidents involving LLMs that impact customer data?
10. What strategies do you use for secure storage and management of LLM training data?
11. Do you have redundancy and failover mechanisms in place for critical LLM components?
12. How do you ensure compliance with regulatory requirements related to LLM security?
13. What role-based access controls (RBAC) do you implement for LLM administration and usage?
14. Can you describe your disaster recovery plans specific to LLM-related disruptions?
15. How do you manage and secure user authentication and authorization for LLM access?
16. What network segmentation measures do you employ to isolate LLM-related traffic?
17. How do you validate the integrity and authenticity of LLM models and outputs?
18. Can you describe your secure coding practices for LLM-related software and applications?
19. How do you handle and mitigate insider threats related to LLMs?
20. What data loss prevention (DLP) measures do you have in place for LLM-related data?
21. How do you conduct vulnerability assessments and penetration testing for LLMs?
22. Can you provide examples of LLM-specific security incidents you have encountered and resolved?
23. How do you stay updated on emerging threats and vulnerabilities relevant to LLMs?
24. What encryption key management practices do you follow for LLM-related operations?
25. Can you describe your incident response playbook for LLM security breaches?

TASK 4: RESOURCES ASSESSMENT QUESTIONS

1. How do you collaborate with security teams and stakeholders to address AI security challenges?
2. Can you describe the role of AI security policies, procedures, and standards in your Business Unit?
3. What tools and technologies do you use for AI security testing, monitoring, and analysis?
4. How do you handle incident response and recovery in the event of AI security incidents?
5. Any knowledge of threat modeling and risk assessment in AI security?
6. How do you communicate AI security risks and strategies to non-technical stakeholders?

TASK 5: AI BUSINESS ASSESSMENT QUESTIONS

1. How does your AI strategy align with the Business Unit's overall business objectives and goals?
2. How do you ensure that AI solutions comply with legal and regulatory requirements, including data privacy laws?
3. What measures are in place to address ethical considerations and biases in AI decision-making processes?
4. Can you describe the process for evaluating and selecting AI technologies and vendors based on security and compliance criteria?
5. How do you assess the potential risks and impact of AI failures or vulnerabilities on business operations?
6. What strategies are implemented to monitor and audit AI systems for security incidents and anomalies?
7. How does the Business Unit handle data governance, access controls, and data protection in AI projects?
8. What are the procedures for reporting and responding to AI-related security incidents and breaches?
9. How do you measure the ROI (Return on Investment) and business value derived from AI implementations?

TASK 6: AI DEVELOPER ASSESSMENT QUESTIONS

1. What security best practices do you follow during the development and deployment of AI models and applications?
2. Can you explain how you mitigate potential risks associated with OWASP LLM Top 10 vulnerabilities in AI systems?
3. What steps are taken to ensure data integrity, confidentiality, and availability in AI projects?
4. How do you validate and test AI models for accuracy, robustness, and reliability?
5. What measures are in place to address bias, fairness, and transparency in AI algorithms and decision-making?
6. Can you describe the process for securing AI training data and preventing data poisoning attacks?
7. How do you implement secure coding practices and secure software development lifecycle (SDLC) in AI projects?
8. What security controls are integrated into AI applications to protect against insider threats and unauthorized access?
9. How do you manage and update AI models to address evolving security threats and vulnerabilities?
10. Can you provide examples of AI-related security challenges you've encountered and how they were resolved?

TASK 7: QUESTIONS RELATED TO HOW A BUSINESS UTILIZES AI TOOLS SUCH AS GENAI FOR CODE GENERATION, REQUIREMENTS MATCHING, AND OTHER FUNCTIONS PART 1

1. How does your business utilize AI tools like GenAI for code generation and automation of development tasks?
2. Can you describe the specific use cases where AI tools are employed for code generation within your Business Unit?
3. What benefits have you observed from using AI tools for code generation, such as increased productivity or improved code quality?
4. How do you ensure that the AI-generated code meets your Business Unit's coding standards and best practices?
5. What measures are in place to validate the accuracy and reliability of AI-generated code before deployment?
6. How does your business leverage AI tools for requirements gathering and matching processes?
7. Can you explain the integration of AI tools into your requirements management and analysis workflows?
8. What challenges or limitations have you encountered when using AI tools for requirements matching?
9. How do you ensure that AI-generated requirements align with stakeholders' expectations and business goals?
10. What strategies do you employ to handle complex or ambiguous requirements using AI tools?
11. How does your business address security considerations when utilizing AI tools for code generation and requirements matching?
12. Can you describe the access controls and permissions associated with AI tools used for development tasks?
13. How do you manage and protect sensitive data involved in AI-driven code generation and requirements analysis?
14. What measures are in place to prevent unauthorized access or manipulation of AI-generated code or requirements?
15. How does your Business Unit handle compliance with regulatory requirements related to AI tools used in development processes?
16. Can you provide examples of successful projects where AI tools significantly contributed to code generation and requirements matching?
17. What strategies do you use to monitor and evaluate the performance of AI tools in code generation and requirements analysis?
18. How do you handle the integration of AI-generated code with existing software systems and applications?

QUESTIONS RELATED TO HOW A BUSINESS UTILIZES AI TOOLS SUCH AS GENAI FOR CODE GENERATION, REQUIREMENTS MATCHING, AND OTHER FUNCTIONS PART 2

1. What considerations are taken into account when selecting or customizing AI models for code generation and requirements matching?
2. How does your Business Unit ensure the scalability and reliability of AI tools for handling large-scale development projects?
3. Can you describe the training and support provided to developers and teams using AI tools for code generation?
4. How do you measure the return on investment (ROI) of AI tools used in development processes?
5. What future enhancements or improvements do you anticipate in AI tools for code generation and requirements matching?
6. How does your business manage the ethical implications of using AI tools in software development, particularly in sensitive areas like privacy or security?
7. Can you share any lessons learned or best practices related to integrating AI tools into development workflows?
8. How do you handle version control and documentation for AI-generated code and requirements?
9. What role does AI play in assisting developers with tasks such as debugging or code refactoring?
10. How do you balance the use of AI tools with human expertise and decision-making in development processes?
11. Can you describe any challenges or obstacles faced when transitioning to AI-driven development practices?
12. How does your Business Unit ensure transparency and explainability in AI-driven code generation and requirements analysis?
13. What measures are in place to prevent bias or discrimination in AI-generated code or requirements?
14. How do you address concerns about job displacement or changes in roles due to the adoption of AI tools in development?

QUESTIONS RELATED TO HOW A BUSINESS UTILIZES AI TOOLS SUCH AS GENAI FOR CODE GENERATION, REQUIREMENTS MATCHING, AND OTHER FUNCTIONS PART 3

1. Can you provide insights into the cost implications of using AI tools for code generation and requirements matching?
2. How does your business handle data privacy and security when utilizing AI tools that may process sensitive information?
3. What steps are taken to ensure the reliability and accuracy of AI-generated requirements for software projects?
4. How do you evaluate the performance and effectiveness of AI tools in meeting development objectives?
5. Can you describe any regulatory or industry standards that impact the use of AI tools in software development?
6. How does your Business Unit approach risk management in relation to AI-driven development processes?
7. What strategies do you use to ensure continuous improvement and optimization of AI tools for code generation and requirements matching?
8. How does AI contribute to streamlining development workflows and reducing time-to-market for software projects?
9. Can you share examples of collaboration between AI tools and human developers in achieving project goals?
10. How does your business handle data quality and integrity when using AI tools for code generation and analysis?
11. What considerations are given to data governance and access control in AI-driven development environments?
12. How do you address challenges related to the interpretability and explainability of AI-generated code or requirements?
13. Can you describe any challenges or successes in implementing AI tools for cross-platform code generation?
14. How does AI support decision-making processes in software development, such as feature prioritization or architecture design?
15. What measures are in place to prevent malicious actors from exploiting vulnerabilities in AI-generated code or requirements?
16. How do you ensure collaboration and communication between teams working with AI tools in development projects?
17. Can you provide examples of innovation or unique solutions achieved through the use of AI tools in software development?
18. How does your Business Unit stay informed about advancements and trends in AI technologies relevant to code generation and requirements matching?

TASK 8: QUESTIONS BASED ON OWASP TOP 10 FOR LLM AND MITRE ATLAS TACTICS

1. Do you have measures in place to prevent prompt injection attacks against your large language models (LLMs)?
2. How do you handle outputs from LLMs to ensure there are no vulnerabilities related to insecure output handling?
3. What strategies do you employ to prevent training data poisoning in your LLMs?
4. Have you implemented safeguards to mitigate model denial of service attacks targeting your LLMs?
5. How do you ensure the security of your LLMs' supply chain, including third-party datasets and pre-trained models?
6. What steps do you take to prevent sensitive information disclosure by your LLMs in their responses?
7. Are your LLM plugins designed securely with proper input validation and access controls?
8. How do you prevent LLMs from taking excessive actions or exhibiting excessive agency?
9. Do you have mechanisms in place to detect and address overreliance on LLMs within your systems?
10. How do you protect your LLM models from theft and unauthorized access?

TASK 9: QUESTIONS RELATED TO HOW SECURITY WILL BE MANAGED AND WHAT MITIGATION STRATEGIES ARE NEEDED USING OWASP TOP 10 FOR LLM (LARGE LANGUAGE MODELS) – PART 1

1. How does your Business Unit plan to address prompt injection vulnerabilities in large language models (LLMs)?
2. What measures are in place to ensure secure handling of LLM outputs and prevent insecure output handling vulnerabilities?
3. Can you describe the strategies for mitigating training data poisoning risks in LLMs?
4. How will your Business Unit prevent and mitigate model denial of service attacks targeting LLMs?
5. What steps are taken to secure the LLM supply chain and prevent supply chain vulnerabilities?
6. How do you implement data sanitization and strict user policies to prevent sensitive information disclosure by LLMs?
7. What security controls are in place to ensure secure design and implementation of LLM plugins?
8. How will your Business Unit prevent LLM-based systems from taking excessive actions or exhibiting excessive agency?
9. What strategies are in place to mitigate overreliance on LLMs and prevent associated security vulnerabilities?
10. How does your Business Unit plan to protect LLM models from theft and unauthorized access?
11. Can you describe the process for identifying and remediating prompt injection vulnerabilities in LLMs?
12. What role-based access controls (RBAC) are implemented for managing LLM outputs and preventing insecure output handling?
13. How does your Business Unit validate and verify the integrity of LLM training data to prevent training data poisoning?
14. What monitoring and alerting mechanisms are in place to detect and respond to model denial of service attacks against LLMs?
15. How are third-party components and services in the LLM supply chain assessed and secured to prevent supply chain vulnerabilities?
16. Can you provide examples of sensitive information that could be disclosed by LLMs and the measures in place to protect against such disclosures?
17. What security assessments are conducted for LLM plugins to ensure they are designed securely and have proper access controls?
18. How does your Business Unit enforce permissions and restrictions to prevent LLM-based systems from undertaking unintended actions?
19. Can you describe the process for evaluating the reliability and accuracy of LLM outputs to mitigate overreliance risks?
20. How are access controls and encryption used to protect proprietary LLM models from theft and unauthorized access?
21. What tools or techniques are used to assess and mitigate prompt injection vulnerabilities in LLMs during development and testing phases?
22. How do you ensure that LLM outputs undergo thorough validation and sanitization to prevent insecure output handling?
23. Can you describe the process for auditing and validating LLM training data to detect and mitigate data poisoning attempts?

QUESTIONS RELATED TO HOW SECURITY WILL BE MANAGED AND WHAT MITIGATION STRATEGIES ARE NEEDED USING OWASP TOP 10 FOR LLM (LARGE LANGUAGE MODELS) – PART 2

1. What contingency plans are in place to handle and mitigate the impact of model denial of service attacks on LLMs?
2. How does your Business Unit conduct security assessments and audits of third-party components and services used in LLM applications?
3. What mechanisms are in place to detect and prevent sensitive information disclosure by LLMs in real-time?
4. How are secure coding practices integrated into the development of LLM plugins to prevent insecure plugin design vulnerabilities?
5. Can you describe the process for implementing fail-safe mechanisms to prevent excessive agency by LLM-based systems?
6. What controls and monitoring mechanisms are in place to detect and mitigate overreliance on LLMs within your Business Unit?
7. How does your Business Unit implement encryption and access controls to protect against model theft and unauthorized access to LLM models?
8. What measures are in place to ensure secure deployment and configuration of LLMs to prevent prompt injection vulnerabilities in production environments?
9. How do you conduct ongoing security assessments and penetration testing of LLM outputs to identify and remediate insecure output handling vulnerabilities?
10. Can you describe the process for monitoring and analyzing LLM training data for signs of poisoning or manipulation?
11. What incident response procedures are in place to address and mitigate the impact of model denial of service attacks on LLMs?
12. How does your Business Unit assess and manage security risks associated with third-party components and services used in LLM applications?
13. What measures are in place to monitor and audit LLM responses for potential sensitive information disclosures?
14. How are secure development practices enforced for LLM plugins to prevent insecure plugin design vulnerabilities?
15. Can you describe the process for implementing checks and balances to prevent excessive agency and unintended consequences in LLM-based systems?
16. What strategies are in place to educate and train personnel on the risks of overreliance on LLMs and how to mitigate those risks?
17. How does your Business Unit implement data encryption and access controls to protect proprietary LLM models from theft and unauthorized access?
18. What procedures are in place to validate and verify the security configurations of LLM deployments to prevent prompt injection vulnerabilities?
19. How do you conduct regular security assessments and audits of LLM outputs to identify and address insecure output handling vulnerabilities?
20. Can you describe the process for monitoring and detecting anomalies in LLM training data that may indicate poisoning attempts?
21. What response plans are in place to quickly mitigate the impact of model denial of service attacks on LLMs and restore normal operations?
22. How does your Business Unit assess and manage security risks associated with third-party datasets, pre-trained models, and plugins used in LLM applications?
23. What mechanisms are in place to monitor and prevent unauthorized access to sensitive information disclosed by LLMs?
24. How are secure coding standards enforced for LLM plugins to prevent insecure plugin design and implementation?
25. Can you describe the process for establishing governance and oversight to prevent excessive agency and ensure responsible use of LLM-based systems?
26. What training and awareness programs are in place to educate personnel on the risks of overreliance on LLMs and promote security best practices?
27. How does your Business Unit implement strong authentication, authorization, and encryption measures to protect LLM models and prevent model theft?

TASK 10 - QUESTIONS FOCUSING ON DATA POISONING, ATTACKS ON MODELS, AND THE INTEGRATION OF 3RD PARTY AI, GENAI, AND LLM TOOLS INTO THE BUSINESS

1. How does your Business Unit integrate 3rd party AI tools, such as GenAI and LLM, into your existing technological infrastructure?
2. What types of data are processed by these AI tools, and what measures are in place to protect against data poisoning attempts?
3. How do you assess and mitigate the risks of model attacks targeting the AI tools used in your business operations?
4. Can you describe the security implications and safeguards involved in the generation and maintenance of Java code using AI tools?
5. What potential vulnerabilities could arise during the integration of 3rd party AI tools, and how are these vulnerabilities addressed?
6. How does your Business Unit ensure the security and integrity of data inputs and outputs for AI tools like GenAI and LLM?
7. Can you explain the measures in place to prevent unauthorized access and modifications to AI models and data?
8. What strategies are employed to detect and mitigate data poisoning attempts targeting AI tools within your Business Unit?
9. How are AI models and algorithms protected against adversarial attacks and malicious inputs?
10. Can you describe the process for evaluating and selecting 3rd party AI tools based on their security features and capabilities?
11. What measures are in place to monitor and analyze AI model behavior for signs of compromise or abnormal activities?
12. How does your Business Unit handle the secure integration of GenAI and LLM tools into your software development and business processes?
13. Can you provide examples of security controls implemented to prevent unauthorized access to AI-generated code and outputs?
14. How do you address the risks associated with using AI tools from different vendors and ensuring compatibility and security?
15. What steps are taken to validate and verify the accuracy and reliability of AI-generated outputs before deployment?
16. How does your Business Unit handle security incidents or breaches related to AI tools, including data poisoning or model attacks?
17. Can you describe the process for securing AI models and data against insider threats and unauthorized access?
18. What mechanisms are in place to ensure the accountability and transparency of AI tools used in your business?
19. How do you stay updated on emerging threats and vulnerabilities related to AI tools, including GenAI and LLM?
20. Can you provide insights into the secure integration of AI tools into your business workflows, including data handling and processing?

TASK 11: QUESTIONS CRAFTED BY THE SECURITY TEAM THAT THE AI TEAM, WHICH MAY NOT BE WELL-VERSED IN SECURITY, WOULD ANSWER REGARDING DATA POISONING, ATTACKS ON MODELS, AND INTEGRATION OF 3RD PARTY AI TOOLS

1. How is the data processed by AI tools like GenAI and LLM, and what steps are taken to ensure the accuracy and reliability of this data?
2. Can you describe the measures in place to protect AI models and algorithms from adversarial attacks and malicious inputs?
3. How does the AI team assess and mitigate the risks of data poisoning targeting AI tools within the Business Unit?
4. What strategies are employed to monitor and analyze AI model behavior for signs of compromise or abnormal activities?
5. Can you explain the process for evaluating and selecting 3rd party AI tools based on their security features and capabilities?
6. How does the AI team handle the secure integration of GenAI and LLM tools into the software development and business processes?
7. Can you provide examples of security controls implemented by the AI team to prevent unauthorized access to AI-generated code and outputs?
8. What mechanisms are used by the AI team to ensure the accountability and transparency of AI tools used in the Business Unit?
9. How does the AI team stay updated on emerging threats and vulnerabilities related to AI tools, including GenAI and LLM?
10. Can you provide insights into the secure integration of AI tools into business workflows, including data handling and processing?

TASK 12: QUESTIONS CRAFTED BY THE SECURITY TEAM FOR THE AI TEAM, FOCUSING ON DATA POISONING, ATTACKS ON MODELS, AND THE INTEGRATION OF 3RD PARTY AI TOOLS – PART 1

1. How does the AI team ensure the integrity and reliability of data processed by GenAI and LLM tools?
2. Can you describe the measures taken to detect and prevent data poisoning in AI models?
3. What steps are in place to protect AI models from adversarial attacks and malicious inputs?
4. How does the AI team assess and mitigate the risks of model attacks targeting GenAI and LLM tools?
5. Can you explain the process for evaluating and selecting 3rd party AI tools based on their security capabilities?
6. What strategies are employed to monitor and analyze AI model behavior for signs of compromise or abnormal activities?
7. How does the AI team handle the secure integration of 3rd party AI tools like GenAI and LLM into the Business Unit's infrastructure?
8. Can you provide examples of security controls implemented to prevent unauthorized access to AI-generated code and outputs?
9. What mechanisms are used by the AI team to ensure the accountability and transparency of AI tools used in the Business Unit?
10. How does the AI team stay updated on emerging threats and vulnerabilities related to AI tools?
11. Can you describe the process for securing AI models and data against insider threats?
12. What measures are in place to prevent unauthorized modifications to AI models and algorithms?
13. How does the AI team ensure data privacy and confidentiality when processing sensitive information?
14. Can you explain the role of encryption in protecting AI models and data from unauthorized access?
15. What steps are taken to validate and verify the accuracy of AI-generated outputs before deployment?
16. How does the AI team handle security incidents or breaches related to AI tools, including data poisoning or model attacks?
17. Can you describe the process for conducting security assessments and audits of AI tools and models?
18. What strategies are employed to mitigate the risks of using 3rd party AI tools with potential security vulnerabilities?
19. How does the AI team address the challenges of integrating AI tools into existing security frameworks and policies?
20. Can you provide insights into the secure development practices implemented by the AI team for AI models and algorithms?
21. What measures are in place to ensure the traceability and auditability of AI model outputs?
22. How does the AI team handle the secure storage and management of AI models and training data?
23. Can you describe the process for validating and verifying the accuracy and reliability of AI models?

QUESTIONS CRAFTED BY THE SECURITY TEAM FOR THE AI TEAM, FOCUSING ON DATA POISONING, ATTACKS ON MODELS, AND THE INTEGRATION OF 3RD PARTY AI TOOLS – PART 2

1. What mechanisms are used to monitor and detect anomalies in AI model behavior that may indicate security threats?
2. How does the AI team collaborate with the security team to address security concerns related to AI tools and models?
3. Can you explain the process for implementing access controls and permissions for AI tools and data?
4. What role does the AI team play in ensuring compliance with regulatory requirements related to AI tools and data?
5. How does the AI team handle the secure deployment and configuration of AI tools in production environments?
6. Can you provide examples of security incidents or breaches that the AI team has successfully mitigated in the past?
7. What measures are in place to ensure the reliability and availability of AI tools and models?
8. How does the AI team handle the secure exchange of data between AI tools and external systems?
9. Can you describe the process for implementing authentication and authorization mechanisms for AI tools and users?
10. What steps are taken to ensure the security and integrity of AI model training data?
11. How does the AI team address the risks associated with using AI models in critical or sensitive applications?
12. Can you explain the process for conducting vulnerability assessments and penetration testing of AI tools and models?
13. What measures are in place to ensure the secure disposal of AI models and data when they are no longer needed?
14. How does the AI team handle the secure transmission of data between AI tools and cloud services?
15. Can you describe the process for implementing secure APIs for AI tools and data integration?
16. What role does encryption play in protecting AI models and data at rest and in transit?
17. How does the AI team ensure the accuracy and fairness of AI models, particularly in applications involving sensitive data or decision-making?
18. Can you explain the process for implementing secure logging and monitoring for AI tools and models?
19. What measures are in place to prevent AI models from leaking sensitive information or producing biased results?
20. How does the AI team handle the secure updating and patching of AI tools and models?
21. Can you describe the process for conducting risk assessments and threat modeling for AI tools and models?
22. What role does the AI team play in incident response and recovery efforts related to AI tools and models?
23. How does the AI team ensure the compatibility and interoperability of AI tools with other systems and applications?
24. Can you explain the process for implementing secure data preprocessing and feature engineering for AI models?
25. What measures are in place to ensure the confidentiality and privacy of AI model outputs and predictions?
26. How does the AI team address the challenges of explainability and interpretability in AI models and decisions?
27. Can you describe the process for documenting and communicating security best practices and policies related to AI tools and models within the Business Unit?

TASK 13: QUESTIONS THAT FOCUS ON UNDERSTANDING BUSINESS PROCESSES, AGREEMENTS, AND RELATED ASPECTS IN THE CONTEXT OF AI TOOLS, INCLUDING DATA POISONING, ATTACKS ON MODELS, AND INTEGRATION WITH 3RD PARTY AI TOOLS – PART1

1. How does the AI team collaborate with different business units to understand their specific needs and requirements for AI tools?
2. Can you describe the key stakeholders involved in decision-making processes related to AI tools within the Business Unit?
3. What legal and regulatory requirements govern the use of AI tools, especially concerning data privacy and security?
4. How does the AI team ensure compliance with data protection laws and regulations when processing sensitive information?
5. Can you provide insights into the business processes that AI tools are integrated into, and how they contribute to operational efficiency?
6. What agreements or contracts are in place with 3rd party AI tool providers, and how are they managed and reviewed for security aspects?
7. How does the AI team ensure that AI tools align with the Business Unit's strategic goals and objectives?
8. Can you describe the process for evaluating and selecting AI tools based on their compatibility with existing business systems and processes?
9. What measures are in place to ensure data accuracy and reliability when using AI tools for decision-making?
10. How does the AI team address the challenges of explainability and transparency in AI-driven business processes?
11. Can you provide examples of successful implementations of AI tools in improving business outcomes and customer experiences?
12. What steps are taken to educate and train employees on the use of AI tools and their role in business processes?
13. How does the AI team handle the integration of AI tools with existing IT infrastructure and applications?
14. Can you describe the process for assessing and managing risks associated with AI tool deployment in business operations?
15. What mechanisms are used to monitor and measure the performance and effectiveness of AI tools in achieving business objectives?
16. How does the AI team ensure that AI tool outputs and predictions are aligned with business requirements and expectations?
17. Can you provide examples of challenges or obstacles encountered in integrating AI tools into business processes, and how they were overcome?
18. How are AI tools used to automate repetitive tasks and streamline workflows in various business departments?
19. What data governance policies and procedures are in place to manage AI tool usage and data access across the Business Unit?
20. How does the AI team collaborate with legal and compliance teams to ensure that AI tools adhere to industry regulations and standards?
21. Can you describe the process for documenting and maintaining AI tool configurations and settings for audit and compliance purposes?

QUESTIONS THAT FOCUS ON UNDERSTANDING BUSINESS PROCESSES, AGREEMENTS, AND RELATED ASPECTS IN THE CONTEXT OF AI TOOLS, INCLUDING DATA POISONING, ATTACKS ON MODELS, AND INTEGRATION WITH 3RD PARTY AI TOOLS – PART 2

1. What measures are in place to protect sensitive business data when using AI tools for analysis and decision-making?
2. How does the AI team address the scalability and resource requirements of AI tools as business needs evolve?
3. Can you provide insights into the cost-benefit analysis conducted for AI tool adoption and implementation in business processes?
4. What contingency plans are in place to address disruptions or failures in AI tool functionality during critical business operations?
5. How does the AI team ensure that AI tool usage aligns with ethical guidelines and values of the Business Unit?
6. Can you describe the process for evaluating the ROI (Return on Investment) of AI tools and their impact on business outcomes?
7. What data retention and deletion policies are followed when using AI tools to process and store business data?
8. How does the AI team handle the integration of AI tool outputs with reporting and analytics systems for decision support?
9. Can you provide examples of data security incidents or breaches related to AI tool usage, and how they were mitigated?
10. What steps are taken to ensure that AI tool usage complies with contractual agreements and service-level agreements (SLAs)?
11. How does the AI team collaborate with vendors and partners to leverage AI tools effectively in joint business initiatives?

QUESTIONS THAT FOCUS ON UNDERSTANDING BUSINESS PROCESSES, AGREEMENTS, AND RELATED ASPECTS IN THE CONTEXT OF AI TOOLS, INCLUDING DATA POISONING, ATTACKS ON MODELS, AND INTEGRATION WITH 3RD PARTY AI TOOLS – PART 3

1. Can you describe the process for evaluating and addressing biases or inaccuracies in AI tool outputs that may impact business decisions?
2. What measures are in place to protect intellectual property and proprietary business information when using AI tools?
3. How does the AI team ensure that AI tool outputs are reliable and consistent across different business scenarios and use cases?
4. Can you provide examples of successful AI tool implementations that have led to cost savings or revenue growth for the Business Unit?
5. What role does the AI team play in developing and implementing AI governance frameworks within the Business Unit?
6. How are AI tools used to enhance customer experiences and satisfaction in various business processes?
7. Can you describe the process for conducting user training and onboarding for AI tools across different business units?
8. What measures are in place to monitor and audit AI tool usage to detect and prevent unauthorized access or misuse?
9. How does the AI team address the challenges of data integration and interoperability when using AI tools across diverse business systems?
10. Can you provide insights into the strategic planning and roadmap for AI tool adoption and expansion within the Business Unit?
11. What metrics and KPIs (Key Performance Indicators) are used to measure the success and impact of AI tools on business operations?
12. How does the AI team collaborate with internal and external stakeholders to gather feedback and improve AI tool functionality?
13. Can you describe the process for evaluating and selecting AI tool vendors based on their track record, expertise, and support capabilities?
14. What measures are in place to ensure data sovereignty and compliance with international data transfer regulations when using AI tools?
15. How does the AI team handle the integration of AI tools with business intelligence (BI) platforms and data visualization tools?
16. Can you provide examples of challenges faced in scaling AI tool usage across the Business Unit, and how they were addressed?
17. What steps are taken to ensure continuous monitoring and optimization of AI tool performance and efficiency?
18. How does the AI team collaborate with business leaders to identify new opportunities for AI tool deployment and innovation?

TASK 14: QUESTIONS RELATED TO FUNCTIONALITY, PERFORMANCE, AND SECURITY MEASURES

- Functionality:

1. Does the AI tool meet the operational requirements specified for its intended use?
2. How effectively does the tool handle complex queries and inputs?
3. Can the tool seamlessly integrate with existing systems and workflows within the Business Unit?
4. What mechanisms are in place to ensure data accuracy and consistency in the tool's outputs?
5. How does the tool handle scalability and performance when processing large volumes of data?
6. Can the tool adapt to evolving business requirements and user needs over time?
7. What level of customization and configuration options does the tool offer to meet specific use cases?
8. How user-friendly is the tool's interface, and does it support efficient workflows for users?
9. Can the tool automate repetitive tasks and streamline manual processes effectively?
10. What measures are in place to ensure continuous availability and uptime of the tool?

TASK 15: PERFORMANCE: WHAT IS THE TOOL'S CAPACITY FOR HANDLING CONCURRENT REQUESTS AND WORKLOAD SPIKES?

1. How well does the tool perform in terms of response time and latency under normal operating conditions?
2. Can the tool maintain performance levels during peak usage periods or heavy workloads?
3. How does the tool perform when subjected to stress testing or load testing scenarios?
4. What strategies are in place to optimize the tool's performance and resource utilization?
5. How does the tool handle resource constraints or limitations, such as memory and processing power?
6. Can the tool efficiently process and analyze data in real-time or near real-time?
7. What measures are in place to monitor and evaluate the tool's performance metrics regularly?
8. How does the tool ensure data integrity and consistency across different operations and transactions?
9. Can the tool provide meaningful insights and analytics based on performance data and metrics?

TASK 16: WHAT SECURITY MEASURES ARE IMPLEMENTED TO PROTECT SENSITIVE DATA PROCESSED BY THE TOOL?

1. How are authentication and access controls managed to prevent unauthorized access to the tool and its data?
2. Can the tool detect and mitigate potential security threats such as SQL injection, cross-site scripting (XSS), and other common attacks?
3. What encryption protocols are used to secure data in transit and at rest within the tool?
4. How does the tool handle user permissions and roles to enforce least privilege access principles?
5. Can the tool detect and respond to anomalies or suspicious activities that may indicate security breaches?
6. What measures are in place to secure the tool's communication channels and APIs?
7. How resilient is the tool against DDoS (Distributed Denial of Service) attacks and other forms of network-based threats?
8. Can the tool provide audit logs and logs of security-related events for monitoring and analysis?
9. How often are security assessments and penetration testing conducted to evaluate the tool's security posture?

TASK 17: DOES THE TOOL SUPPORT INTEGRATION WITH SECURITY INFORMATION AND EVENT MANAGEMENT (SIEM) SYSTEMS FOR CENTRALIZED MONITORING?

1. How does the tool facilitate incident response and remediation processes in case of security incidents?
2. Can the tool detect and block malicious activities, such as malware infections or unauthorized access attempts?
3. What measures are in place to ensure data privacy and confidentiality within the tool?
4. How does the tool handle compliance requirements related to data protection regulations and industry standards?
5. Can the tool enforce data retention policies and data lifecycle management practices?
6. What backup and recovery mechanisms are in place to protect against data loss and ensure business continuity?
7. How does the tool address the risks associated with insider threats and internal vulnerabilities?
8. Can the tool detect and prevent data exfiltration attempts and unauthorized data transfers?
9. What encryption standards and protocols are used to secure data stored within the tool's databases?

TASK 18: **PERFORMANCE ASSESSMENT**

1. How does the tool handle high-volume data processing and analytics tasks?
2. Can the tool scale horizontally and vertically to accommodate growing data volumes and user demands?
3. What measures are in place to optimize resource utilization and minimize performance bottlenecks?
4. How does the tool perform in terms of data processing speed and efficiency compared to industry benchmarks?
5. Can the tool provide real-time monitoring and alerts for performance anomalies and degradation?
6. What strategies are used to optimize query performance and minimize latency in data retrieval?
7. How does the tool handle data caching and pre-fetching to improve performance for frequently accessed data?
8. Can the tool support distributed computing and parallel processing to enhance performance?
9. What load balancing and failover mechanisms are in place to ensure high availability and fault tolerance?
10. How does the tool handle data synchronization and replication across distributed environments for performance optimization?

QUESTION AND ANSWER