Applying AIS Domain of CCM to Generative AI

1. Introduction

The Cloud Control Matrix (CCM) is a cybersecurity control framework for cloud computing that's developed by the Cloud Security Alliance (CSA). It's designed to provide organizations with the necessary structure, detail, and clarity relating to information security tailored to the cloud industry.

The number of controls and their descriptions have evolved over the years with new versions of the CCM. CCM v4.0 is the latest version and has a total of 197 control objectives spread across 17 domains. This white paper focuses on Application & Interface Security (AIS) domain: Ensures secure software, application development, and lifecycle management processes.

The Cloud Control Matrix (CCM) offers a uniquely suitable framework for assessing controls for generative AI, owing to its distinct attributes:

1. Comprehensive Coverage: The CCM encompasses a broad spectrum of security controls relevant to cloud environments, which aligns well with the multifaceted security needs of generative AI models often operated in the cloud.

2. Flexible Adaptation: Designed originally for cloud security, the CCM's modular structure enables easy tailoring and expansion to cater to the specific requirements of generative AI systems.

3. Industry Acknowledgment: The CCM enjoys widespread recognition and esteem within the industry, serving as a robust foundation in sync with established best practices.

4. Regulatory Compliance: Crafted with global regulations in mind, applying the CCM to generative AI ensures both security and adherence to international standards.

5. Methodical Evaluation: Organized into domains like "Application & Interface Security (AIS)," the CCM facilitates a structured assessment approach, leaving no security aspect unaddressed.

6. Community-Driven Updates: Continuously refined with input from a diverse community of security experts, the CCM remains relevant and responsive to emerging threats in the rapidly evolving realm of generative AI.

7. Audit Emphasis: Given the opacity of many AI models, the CCM's focus on audit assurance and compliance proves vital for consistent security and ethical evaluation.

In essence, the CCM's comprehensive, adaptable, and structured nature, coupled with its industry acclaim and global compliance alignment, positions it ideally for evaluating and implementing controls for Generative AI(GenAI) systems.

This paper initially centers on applying the "Application & Interface Security (AIS)" domain of the CCM to the realm of GenAI

2. AIS Controls: What they are, and their Application to Generative AI:

"Application & Interface Security (AIS)" domain of the CCM includes 7 controls; we will review these 7 controls and then list their applicability to GenAI.

2.1 Review of AIS Controls

1. AIS-01: Application and Interface Security Policy and Procedures: Establish, document, approve, communicate, apply, and update a policy and procedures for application and interface security.

2. AIS-02: Application Security Baseline Requirements: Establish, document, and maintain baseline requirements for application and interface security.

3. AIS-03: Application Security Metrics:Define and implement technical and operational metrics for application and interface security.

4. AIS-04: Secure Application Design and Development: Define and implement a SDLC process for application and interface security.

5. AIS-05: Automated Application Security Testing: Implement a testing strategy, including criteria for security testing tools and their effectiveness.

6. AIS-06: Automated Secure Application Deployment: Establish and implement strategies and capabilities for secure application and interface deployment.

7. AIS-07: Application Vulnerability Remediation: Define and implement a process to remediate application and interface security vulnerabilities.

2.2 AIS Control and Applicability for Generative AI

The following table gives the applicability of AIS controls to Generative AI

Table 1: AIS Controls and their Applicability for Generative AI

Control ID	Control Title	Control Specification	Applicability for Generative AI
AIS-01	Application and Interface Security Policy and Procedures	Establish, document, approve, communicate, apply, and update a policy and procedures for application and interface security.	Policies governing AI model access, interactions, and reviews ensure robust security as models evolve.
AIS-02	Application Security Baseline Requirements	Establish, document, and maintain baseline requirements for application and interface security.	Baseline security standards protect generative AI from unauthorized access and unintentional data leaks.
AIS-03	Application Security Metrics	Define and implement technical and operational metrics for application and interface security.	Metrics such as unauthorized access attempts or quality of generated content offer insights into AI system operation.

AIS-04	Secure Application Design and Development	Define and implement a SDLC process for application and interface security.	Security mechanisms, like those preventing model inversion attacks, should be integrated from the design phase.
AIS-05	Automated Application Security Testing	Implement a testing strategy, including criteria for security testing tools and their effectiveness.	Automated testing ensures AI behaves as expected, verifying content adherence to guidelines and checking vulnerabilities.
AIS-06	Automated Secure Application Deployment	Establish and implement strategies and capabilities for secure application and interface deployment.	Automated checks during AI model updates ensure no compromise in security, verifying generated content and vulnerabilities.
AIS-07	Application Vulnerability Remediation	Define and implement a process to remediate application and interface security vulnerabilities.	Swift remediation is crucial for vulnerabilities in generative AI, which may involve patching models or updating training data.

3. AIS Controls and Their Concrete Application to Generative AI in Banking

This section using GenAI in Banking as an example, to discuss Application & Interface Security (AIS) controls examples in the safe adoption of generative AI in banking.

AIS-01: Application and Interface Security Policy and Procedures

- Context: In the banking domain, AI models, especially chatbots, handle sensitive user queries ranging from account balances to loan inquiries.

- Example: Consider a generative AI chatbot, 'BankBot', which assists users in navigating their online banking portal. The policies for 'BankBot' must clearly define who can train and modify the model, the exact process it employs to handle and respond to customer queries, and the frequency at which these policies are reviewed and updated. This ensures that 'BankBot' provides accurate information without compromising user data.

AIS-02: Application Security Baseline Requirements

- Context: Banking applications often deal with highly sensitive user data, making it crucial for AI models in this sector to meet stringent security standards.

- Example: An AI model predicting loan eligibility based on user profiles must employ encryption standards to protect user data, have robust identity management protocols to prevent unauthorized access, and ensure that every piece of data used is handled with utmost confidentiality.

AIS-03: Application Security Metrics

- Context: Metrics help in quantitatively gauging the performance and security of AI models.

- Example: For an AI model used in banking to predict potential loan defaults, metrics could include its accuracy in predictions, the number of unauthorized access attempts, and its response time. Consistently monitoring these metrics ensures that the model performs optimally and securely.

AIS-04: Secure Application Design and Development

- Context: Banking applications demand high standards of security given the sensitive nature of financial transactions.

- Example: A generative AI model forecasting stock market trends for the bank's investment wing must be designed to securely handle financial data, ensuring that potential data leaks or biases are addressed right from the design phase.

AIS-05: Automated Application Security Testing

- Context: Automation ensures that security checks are consistent and continuous.

- Example: Automated tests for a chatbot in banking, like 'BankBot', would ensure that it doesn't inadvertently share sensitive information such as account details, previous transactions, or other confidential data in its generated responses.

AIS-06: Automated Secure Application Deployment

- Context: As AI models evolve, ensuring their secure deployment is crucial.

- Example: Before rolling out an updated version of a fraud detection model, automated checks must verify its efficacy. This ensures that the updated model can accurately detect fraudulent transactions without generating a high rate of false positives or negatives.

AIS-07: Application Vulnerability Remediation

- Context: The discovery of vulnerabilities in banking applications can have significant repercussions, making swift remediation vital.

- Example: If a vulnerability is found in 'BankBot', where it mistakenly leaks user transaction histories in certain scenarios, immediate action must be taken to patch the model. Moreover, affected customers must be informed, and steps should be implemented to prevent such occurrences in the future.

Table 2: AIS Controls and Their Concrete Application to Generative AI in Banking

Control ID	Control Title	Applicability for Generative AI in Banking
AIS-01	Application and Interface Security Policy and Procedures	For a banking chatbot, policies dictate who can train the model, how customer queries are processed, and how often policies undergo review.
AIS-02	Application Security Baseline Requirements	All AI models used in banking, from fraud detection to investment suggestions, must meet a minimum encryption standard to protect user data.
AIS-03	Application Security Metrics	Metrics for a loan prediction AI might include accuracy in loan default predictions, unauthorized access attempts, or response time.
AIS-04	Secure Application Design and Development	A financial forecasting AI in banking should be designed to handle sensitive financial data securely, avoiding potential leaks.
AIS-05	Automated Application Security Testing	Automated tests ensure that a chatbot handling banking queries doesn't inadvertently share account details or transaction histories.

AIS-06	Automated Secure Application Deployment	Before deploying an updated fraud detection model, automated checks verify that it doesn't produce false positives/negatives at a high rate.
AIS-07	Application Vulnerability Remediation	If a banking chatbot is found leaking user information, immediate steps must be taken to patch the model and inform affected customers.

Table 2 provides a succinct overview of the AIS controls and their application in generative AI scenarios.

4: AIS Domain Implementation Guidelines for GenAI

AIS-01: Application and Interface Security Policy and Procedures

Guideline 1: The policy should include defined roles and responsibilities.

- Generative AI Application: When deploying a generative AI model, roles and responsibilities should be clearly defined. For instance, certain team members may be responsible for model training, while others handle deployment or monitor outputs.

Guideline 2: Provide a description of the application environment.

- Generative AI Application: Documenting the environment where the generative AI model operates is essential. This can include the hardware it runs on, the data sources it interacts with, and any third-party integrations.

Guideline 3: Mandate regular review processes.

- Generative AI Application: Given the rapid evolution of AI, periodic reviews of the model's performance, outputs, and security are vital. This can ensure the model remains relevant, accurate, and secure over time.

AIS-02: Application Security Baseline Requirements

Guideline : At a minimum, baseline requirements should include security controls, encryption standards, and identity management protocols.

- Generative AI Application: A generative AI model that produces text content for a website should have baseline security measures in place. This includes ensuring outputs are encrypted, access to the model is authenticated, and security protocols are adhered to.

AIS-03: Application Security Metrics

Guideline: Actionable metrics should be defined with considerations for the type of application and its criticality.

- Generative AI Application: For a generative AI model creating art, metrics can include the uniqueness of generated pieces, user engagement rates, and any potential copyright infringements.

AIS-04: Secure Application Design and Development

Guideline: Defining security requirements should be the first step in the development lifecycle.

- Generative AI Application: Before developing a model that generates personalized content for users, security requirements like data privacy, content filtering, and user consent should be established.

AIS-06: Automated Secure Application Deployment

Guideline: The strategies should include defined security checks, approval processes, and monitoring.

- Generative AI Application: When deploying a generative AI model that suggests product recommendations, there should be automated checks ensuring the suggestions are appropriate, an approval process for model updates, and continuous monitoring of model outputs.

AIS-07: Application Vulnerability Remediation

Guideline: Application security remediation should adhere to established policies, ensuring timely response and mitigation.

- Generative AI Application: If a generative AI chatbot starts producing inappropriate responses, there should be a defined process to quickly rectify the model, address the vulnerability, and inform affected users, if necessary.

5. Potential New Controls Needed:

Generative AI's unique capabilities suggest the need for additional controls tailored to its challenges. Table 3 is the initial attempt at defining these controls.

Table 3: Proposed New Controls for AIS Domain Focusing on Application and API Interfaces

Control ID	Control Title	Control Specification
AIS-08	Generative Content Monitoring & Filtering	Implement mechanisms to monitor the content generated by AI models, including filters to prevent the production of inappropriate, harmful, or biased content.
AIS-09	Data Source Authenticity Verification	Ensure that generative AI models verify the authenticity of data sources, especially when integrating with third-party APIs, to prevent data tampering or poisoning.
AIS-10	Rate Limiting & Anomaly Detection	Implement rate limiting for AI-generated requests to APIs and other systems. Incorporate anomaly detection to identify unusual patterns indicative of malicious intent or system malfunctions.
AIS-11	Generative Model Feedback Loop	Establish a feedback mechanism for users or other systems to report issues or anomalies in the content generated by AI, facilitating continuous model improvement.
AIS-12	Secure Model Sharing & Deployment	Define protocols for securely sharing generative AI models, especially when integrating with external systems or platforms, ensuring that model integrity is preserved.
AIS-13	Transparency in Generative Decisions	Provide mechanisms for users or administrators to understand the decision-making process of the generative AI, especially when interfacing with applications or APIs.
AIS-14	API Input Validation for Generative Models	Enhance security by validating and sanitizing inputs from APIs interfacing with generative AI models to prevent injection attacks or other malicious manipulations.

6. Conclusion and Future Work:

This white paper is the initial attempt to apply one of CSA CCM security domains to GenAI systems and applications. Future work can include refine this paper and apply the same research to other security domains inside CCM.