# AI Risk Management Framework for CCP and CCA Practice Exam Generation: A Responsible and Trustworthy Approach

March 16, 2024

## 1. Introduction

This document outlines the risk management approach for an AI system developed to generate practice exams for individuals preparing for the Certified CMMC Professional (CCP) and Certified CMMC Assessor (CCA) certifications. The system leverages two large language models (LLMs), Claude Opus and ChatGPT 4.0, to generate questions and answers, score responses, provide question intent justification, and validate answer choices.

The primary objective of this AI system is to provide a realistic and comprehensive training environment for aspiring CCP and CCA professionals, enabling them to assess their knowledge and preparedness for the certification exams. By harnessing the capabilities of advanced natural language processing models, the system aims to create a diverse and challenging set of practice questions that accurately reflect the breadth and depth of the CMMC (Cybersecurity Maturity Model Certification) domain.

The scope of this risk management document encompasses the entire AI system lifecycle, from data acquisition and model training to system deployment and continuous monitoring. It aligns with the principles and guidelines outlined in the NIST AI Risk Management Framework (AI RMF) and the associated AI RMF Playbook, providing a structured approach to identifying, assessing, and mitigating potential risks associated with the development and operational use of the AI system.

By proactively addressing risk management considerations, this document aims to ensure the responsible and trustworthy deployment of the AI system, safeguarding the integrity of the CCP and CCA certification processes while promoting transparency, fairness, and accountability in the use of AI technologies.

## 2. AI System Description

The AI system for generating CCP and CCA practice exams comprises two state-of-the-art large language models (LLMs), Claude Opus and ChatGPT 4.0. These models have been carefully selected for their respective strengths and capabilities, ensuring a robust and comprehensive approach to creating high-quality practice questions and supporting materials.

Claude Opus, developed by Anthropic, is renowned for its strong reasoning abilities, deep subject matter comprehension, and adherence to ethical principles. Its strengths lie in generating coherent and context-aware responses, making it well-suited for crafting scenario-based questions and providing detailed justifications for answer choices.

On the other hand, ChatGPT 4.0, developed by OpenAI, excels in its breadth of knowledge, creative problem-solving skills, and natural language generation capabilities. Its ability to produce human-like responses and its versatility across various domains make it an ideal choice for generating diverse and challenging multiple-choice, true/false, and open-ended questions.

The AI system is designed to leverage the complementary strengths of these two LLMs by allowing them to alternate in generating practice exam questions and answers, question intent explanations, and answer justifications. This approach ensures a well-rounded and balanced assessment experience for the CCP and CCA candidates.

Furthermore, the system incorporates a feedback mechanism that allows students to rate both the questions and the supporting narratives provided by the LLMs. This valuable feedback will be used to continuously improve the system's performance and address any potential issues or inconsistencies. To cater to varying skill levels, the LLMs are programmed to create questions across three difficulty tiers: Foundational, Moderate, and Advanced. This feature ensures that candidates at different stages of their preparatory journey can find suitable challenges and effectively gauge their progress.

The AI system is capable of generating four types of questions: multiple-choice, true/false, open-ended, and scenario-based. This diversity in question formats ensures a comprehensive assessment of candidates' knowledge and problem-solving abilities, better preparing them for the real-life challenges they may encounter as CCP or CCA professionals.

It is important to note that the LLMs were trained exclusively on publicly available Federal government, CyberAB and CMMC Assessors and Instructors Certification Organization (CAICO) documentation, ensuring that no proprietary or sensitive information was used in the training process. All data sources are from the public domain, promoting transparency and adherence to ethical standards.

The CMMC Training Academy (a service of Cyber Security Training and Consulting LLC) will adopt a phased approach to the system's release, beginning with a Beta testing period. During this phase, the system's performance will be thoroughly analyzed, and any identified risks, such as invalid questions and answers or inaccurate intents and justifications provided by the LLMs, will be addressed through optimization and refinement processes. This iterative approach ensures that the AI system meets the highest standards of quality and reliability before its full-scale deployment.

## 3. Risk Management Framework

To ensure a comprehensive and structured approach to managing risks associated with the AI system for generating CCP and CCA practice exams, the NIST AI Risk Management Framework (AI RMF) has been adopted. The AI RMF provides a holistic and risk-based methodology for addressing the unique challenges posed by AI systems throughout their lifecycle.

The AI RMF components have been mapped to the specific context of the AI system, as follows:

- AI Actor: The CMMC Training Academy (a service of Cyber Security Training and Consulting LLC) is the primary AI Actor responsible for the development, deployment, and maintenance of the AI system.
- AI Environment: The AI system operates within the context of the CCP and CCA certification programs, providing practice exams and training resources for candidates aspiring to obtain these certifications.
- AI Life Cycle: The AI system's life cycle encompasses data acquisition, model training, system integration, deployment, monitoring, and continuous improvement phases.
- AI Use Case: The primary use case for the AI system is to generate realistic and challenging practice exam questions, answer choices, and supporting narratives (question intent and answer justifications) to assist CCP and CCA candidates in their preparation efforts.

By aligning with the AI RMF, the risk management approach for the AI system follows a structured and standardized process, ensuring that potential risks are systematically identified, analyzed, and addressed throughout the system's development and operational phases.

The AI RMF tasks have been mapped to the various sections of this risk management document, providing a comprehensive framework for addressing risks related to areas such as data quality, model performance, system security, bias and fairness, transparency, and accountability.

Adherence to the AI RMF not only promotes the responsible and trustworthy deployment of the AI system but also facilitates compliance with relevant laws, regulations, and ethical guidelines governing the use of AI technologies in the context of professional certification programs.

By adopting the AI RMF, the CMMC Training Academy demonstrates its commitment to prioritizing risk management and fostering trust in the AI system's outputs, thereby ensuring the integrity and credibility of the CCP and CCA certification processes.

## 4. Risk Assessment

The risk assessment process is a critical component of the AI Risk Management Framework, as it involves identifying, analyzing, and evaluating potential risks associated with the AI system for generating CCP and CCA practice exams. This section outlines the key risks identified and their corresponding assessment.

### 4.1 Risk Identification

The following potential risks have been identified:

- Data Quality and Representativeness: Risks related to the quality, completeness, and representativeness of the data used to train the LLMs, which could lead to biased or incomplete question sets.
- Model Performance and Accuracy: Risks associated with the accuracy and reliability of the LLMs' outputs, including the generation of incorrect or inconsistent questions, answers, and supporting narratives.
- System Security and Privacy: Risks related to potential security vulnerabilities or data privacy breaches that could compromise the confidentiality, integrity, and availability of the AI system and its associated data.
- Bias and Fairness: Risks of introducing biases or unfair treatment in the generated questions, which could disadvantage certain groups of candidates or misrepresent the certification domains.
- Transparency and Interpretability: Risks related to the lack of transparency and interpretability in the LLMs' decision-making processes, making it challenging to explain and justify the generated outputs.
- Ethical and Legal Compliance: Risks of non-compliance with relevant ethical guidelines, laws, and regulations governing the use of AI technologies in professional certification programs.

### 4.2 Risk Analysis and Evaluation

Each identified risk has been analyzed and evaluated based on its likelihood of occurrence and potential impact on the AI system's performance, integrity, and credibility. The risk analysis process involves assessing the inherent risk levels and determining the residual risk levels after considering existing controls and mitigation strategies.

The risk evaluation process considers various factors, including the severity of potential consequences, the likelihood of risk occurrence, and the potential impact on stakeholders (candidates, certification bodies, and the broader professional community).

Based on the risk analysis and evaluation, appropriate risk treatment options (avoidance, mitigation, transfer, or acceptance) will be selected and implemented, as outlined in Section 5: Risk Treatment.

By conducting a comprehensive risk assessment, the CMMC Training Academy aims to proactively identify and address potential vulnerabilities, ensuring the responsible and trustworthy deployment of the AI system for generating CCP and CCA practice exams.

## 5. Risk Treatment

Based on the risks identified and evaluated in Section 4, this section outlines the risk treatment strategies and associated controls implemented to mitigate or manage the potential risks associated with the AI system for generating CCP and CCA practice exams.

### 5.1 Data Quality and Representativeness

Risk Mitigation Strategies:

- Rigorous data curation and verification processes to ensure the quality, completeness, and representativeness of the training data.
- Diverse data sourcing from authoritative and reputable sources, including official government and industry publications, subject matter expert inputs, and real-world scenarios.
- Continuous monitoring and updating of the data sources to reflect the latest developments and changes in the CMMC and cybersecurity domains.

### 5.2 Model Performance and Accuracy

Risk Mitigation Strategies:

- Robust model training and validation procedures, including cross-validation and human evaluation, to ensure the accuracy and reliability of the LLMs' outputs.
- Continuous monitoring and refinement of the LLMs' performance, leveraging feedback from subject matter experts and end-users (CCP and CCA candidates).
- Implementation of quality assurance checks to identify and address potential errors or inconsistencies in the generated questions, answers, and supporting narratives.

### 5.3 System Security and Privacy

Risk Mitigation Strategies:

- Adoption of industry-standard security practices, including secure data storage, encryption, access controls, and regular security audits.
- Implementation of robust data privacy protocols, ensuring compliance with relevant regulations (e.g., GDPR, CCPA) and protecting the confidentiality of user data.
- Regular security updates and vulnerability management processes to address potential security risks and threats.

### 5.4 Bias and Fairness

Risk Mitigation Strategies:

- Rigorous bias testing and evaluation procedures to identify and mitigate potential biases in the generated content.
- Diverse and representative training data to promote fairness and inclusivity in the AI system's outputs.
- Collaboration with subject matter experts and stakeholders to ensure the fair and equitable representation of certification domains and candidate groups.

### 5.5 Transparency and Interpretability

Risk Mitigation Strategies:

- Implementation of explainable AI techniques to provide insights into the LLMs' decision-making processes and rationales for generated outputs.

- Clear documentation and communication of the AI system's capabilities, limitations, and potential biases to end-users and stakeholders.
- Establishment of feedback mechanisms to gather user input and continuously improve the transparency and interpretability of the AI system.

### 5.6 Ethical and Legal Compliance

Risk Mitigation Strategies:

- Adherence to relevant ethical guidelines, such as the AI Ethics Principles and the CMMC Code of Professional Conduct (Ethics), throughout the AI system's development and deployment.
- Continuous monitoring of legal and regulatory developments related to the use of AI in professional certification programs and prompt implementation of necessary compliance measures.
- Collaboration with certification bodies, industry associations, and regulatory authorities to ensure alignment with best practices and compliance requirements, where possible.

The implementation of these risk treatment strategies and associated controls will be undertaken in a phased approach, with continuous monitoring, evaluation, and refinement processes in place to ensure the effective management of risks throughout the AI system's lifecycle.

## 6. Monitoring and Evaluation

Continuous monitoring and evaluation are essential components of the risk management approach for the AI system used to generate CCP and CCA practice exams. This section outlines the procedures and metrics employed to assess the effectiveness of the implemented risk controls and identify opportunities for improvement.

### 6.1 Monitoring Procedures

The following monitoring procedures will be implemented:

- Ongoing Data Monitoring: Regular audits and reviews of the training data sources to ensure their quality, currency, and representativeness of the CMMC and cybersecurity domains.
- Model Performance Tracking: Continuous monitoring of the LLMs' outputs, including generated questions, answers, and supporting narratives, to identify any potential errors, inconsistencies, or performance degradation.
- User Feedback Collection: Systematic collection and analysis of feedback from CCP and CCA candidates, subject matter experts, and other stakeholders through surveys, ratings, and open-ended comments.
- Security and Privacy Monitoring: Regular security assessments, vulnerability scans, and compliance audits to identify and address potential security and privacy risks.
- Regulatory and Ethical Compliance Monitoring: Continuous monitoring of legal and regulatory developments, as well as updates to ethical guidelines and industry best practices, to ensure ongoing compliance.

### 6.2 Evaluation Criteria and Metrics

The effectiveness of the risk controls will be evaluated using the following criteria and metrics:

- Data Quality Metrics: Measures of data completeness, accuracy, relevance, and diversity, ensuring the representativeness of the training data.
- Model Performance Metrics: Accuracy, precision, recall, and other relevant metrics for evaluating the LLMs' outputs, as well as measures of consistency and coherence in the generated content.
- User Satisfaction Metrics: User ratings, feedback scores, and qualitative feedback analysis to assess the perceived quality, fairness, and usefulness of the practice exams.
- Security and Privacy Metrics: Measures of system security posture, including vulnerability counts, incident response times, and compliance with data privacy regulations.
- Ethical and Legal Compliance Metrics: Assessments of adherence to ethical principles, industry guidelines, and relevant laws and regulations governing the use of AI in professional certification programs.

### 6.3 Reporting and Communication

The monitoring and evaluation results will be documented and communicated through the following channels:

- Internal Reporting: Regular reports and dashboards for the CMMC Training Academy's leadership and development team, highlighting key performance indicators, identified risks, and recommended actions.
- External Reporting: Periodic reports and updates for certification bodies, industry associations, and other relevant stakeholders, demonstrating the commitment to responsible and trustworthy use of AI technologies.
- Transparency and Disclosure: Public-facing documentation and resources, such as white papers and case studies, to promote transparency and foster trust in the AI system's capabilities and risk management approach.

By implementing robust monitoring and evaluation procedures, the CMMC Training Academy aims to continuously improve the AI system's performance, address emerging risks, and maintain the highest standards of quality, fairness, and accountability in the generation of CCP and CCA practice exams.

## 7. Governance and Accountability

As a small business, Cyber Security Training and Consulting LLC recognizes the importance of establishing effective governance and accountability measures to ensure the responsible development, deployment, and management of the AI system used for generating CCP and CCA practice exams. While operating with limited resources, the company remains committed to upholding ethical principles, maintaining transparency, and promoting accountability throughout the AI system's lifecycle.

### 7.1 Roles and Responsibilities

The following key roles have been identified:

- AI System Lead: A designated individual responsible for overseeing the development, deployment, and ongoing management of the AI system, ensuring alignment with ethical guidelines and regulatory requirements.
- Subject Matter Experts: A team of experienced professionals in the CMMC and cybersecurity domains, providing guidance, validation, and quality assurance for the AI system's outputs.
- Technical Team: A cross-functional team comprising data scientists, machine learning engineers, and software developers responsible for the design, implementation, and maintenance of the AI system.
- Legal and Compliance Advisor: An external consultant or legal professional providing guidance on compliance with relevant laws, regulations, and industry standards related to the use of AI in professional certification programs.

### 7.2 Accountability and Oversight

The following mechanisms have been established to promote accountability and oversight:

- Ethical Guidelines and Code of Conduct: A comprehensive set of ethical guidelines and a Code of Conduct outlining principles and best practices for responsible AI development and deployment, to which all individuals involved in the AI system's lifecycle must adhere.
- Risk Assessment and Mitigation Plan: A documented risk assessment and mitigation plan, regularly reviewed and updated, to identify potential risks and implement appropriate controls and safeguards.
- Quality Assurance and Testing: Rigorous quality assurance processes, including testing, validation, and review by subject matter experts, to ensure the accuracy, fairness, and reliability of the AI system's outputs.
- Stakeholder Engagement: Regular engagement and collaboration with certification bodies, industry associations, and other relevant stakeholders to gather feedback, align with best practices, and address any concerns or issues.
- Incident Response and Remediation: Defined processes for identifying, reporting, and addressing incidents or issues related to the AI system's performance, security, or ethical considerations, including root cause analysis and corrective action plans.
- Transparency and Reporting: Periodic reporting and communication to stakeholders and the broader public, promoting transparency about the AI system's capabilities, limitations, and adherence to ethical and legal requirements.

While operating with limited resources, Cyber Security Training and Consulting LLC remains committed to implementing robust governance and accountability measures to the extent possible. The company will continuously strive to uphold ethical principles, maintain public trust, and ensure the responsible and trustworthy deployment of the AI system for generating CCP and CCA practice exams.

## 8. Conclusion

The development and deployment of the AI system for generating CCP and CCA practice exams represent a significant step towards leveraging advanced technologies to support professional development and training in the cybersecurity domain. While this AI system is not officially endorsed by the Department of Defense (DoD), CyberAB or CAICO, the organization responsible for the CMMC certifications, Cyber Security Training and Consulting LLC has taken a proactive and responsible approach to managing risks associated with the use of AI in this context.

By adopting the NIST AI Risk Management Framework (AI RMF) and aligning with its principles and guidelines, we have established a structured and comprehensive risk management strategy. This strategy encompasses all phases of the AI system's lifecycle, from data acquisition and model training to system deployment and continuous monitoring.

Through rigorous risk assessment processes, we have identified and evaluated potential risks related to data quality, model performance, bias and fairness, transparency, security, privacy, and ethical and legal compliance. Corresponding risk treatment strategies and mitigation controls have been implemented to address these risks, ensuring the responsible and trustworthy development and operation of the AI system.

Robust governance and accountability measures, including the establishment of an Ethical Guidelines and Code of Conduct, a Risk Assessment and Mitigation Plan, and well-defined roles and responsibilities, further reinforce our commitment to upholding ethical principles and promoting transparency in the use of AI technologies.

Continuous monitoring, evaluation, and quality assurance processes have been integrated into our approach, enabling us to continuously improve the AI system's performance, address emerging risks, and align with evolving best practices and regulatory requirements.

Furthermore, we recognize the importance of stakeholder engagement and have established a commitment to actively involving certification candidates, subject matter experts, certification bodies, industry associations, and regulatory authorities. By incorporating their feedback, insights, and expertise, we aim to foster trust, promote transparency, and ensure the integrity and credibility of the AI-generated practice exams.

While this AI system is not an official offering from the DoD, CyberAB or CAICO, it represents a valuable resource for CCP and CCA candidates seeking to assess their knowledge and preparedness for the certification exams. By prioritizing responsible and trustworthy AI development, we strive to contribute to the advancement of cybersecurity professionals while upholding the highest standards of ethical conduct and accountability.

Looking ahead, we remain committed to continuous learning, adaptation, and improvement as the field of AI ethics and governance evolves. We will actively monitor and embrace new guidelines, best practices, and regulatory developments to ensure that our AI system remains aligned with the latest standards for responsible and trustworthy AI development and deployment.

Respectfully,

Jeffrey D. Crump
Principal
Cyber Security Training and Consulting LLC