List of potential questions for the course on AI security

# Agentic AI

**Background (for context):**

[Agentic AI](https://en.wikipedia.org/wiki/Agentic_AI?utm_source=chatgpt.com) refers to autonomous AI systems—often composed of multiple agents—that can plan, decide, and act with minimal human oversight. These systems are being applied across domains like cybersecurity, enterprise workflows, and more.

**Essay question ideas:**

1. **Explain and Evaluate**:

*“*[*What defines Agentic AI*](https://www.techradar.com/pro/the-age-of-agency-why-agentic-ai-will-redefine-the-future-of-work?utm_source=chatgpt.com) *and how does it differ from traditional generative AI? Assess its potential applications and ethical challenges in cybersecurity.”*

(Refer to: [autonomy, proactive behaviour, reinforcement learning, decision‑making](https://arxiv.org/abs/2506.04133?utm_source=chatgpt.com), as opposed to reactive generative models)

1. **Cybersecurity Impacts**:

*“Discuss how Agentic AI could transform cybersecurity defense and offense (e.g., automated threat detection, cyber‑arms‑race escalations). What are the risks?”*

(Based on insights from “[Agentic AI and the Cyber Arms Race](https://arxiv.org/abs/2503.04760?utm_source=chatgpt.com)”)

1. **Trust & Governance**:

*“How can trust, risk, and security management be structured in Agentic AI systems? Evaluate governance frameworks for ensuring accountability.”*

(Use the [TRiSM framework](https://arxiv.org/abs/2506.04133?utm_source=chatgpt.com): governance, explainability, security, etc.)

1. **Legal & Ethical Ramifications:**

*“Explore how Agentic AI challenges traditional notions of authorship, liability, and accountability. How should legal frameworks adapt?”*

(Refer to diffusion of responsibility, the “[moral crumple‑zone](https://arxiv.org/abs/2502.00289?utm_source=chatgpt.com)”)

# Post-Quantum Cryptography (PQC)

**Background (for context):**

[PQC](https://en.wikipedia.org/wiki/Post-quantum_cryptography?utm_source=chatgpt.com) involves designing cryptographic algorithms believed to be resilient to quantum‑computer attacks. Major institutions like NIST have already standardized some algorithms, and frameworks like Mosca’s theorem guide migration urgency.

**Essay question ideas:**

1. \*\*Fundamentals and Importance\*\*:

*“What is post‑quantum cryptography and why is it important in today’s digital infrastructure?”*

(Cover threats from quantum algorithms like Shor’s, “store‑now‑decrypt‑later,” and symmetric vs. public‑key differences)

1. **Time‑Sensitivity in Adoption:**

*“Using Mosca’s theorem, analyse how organizations should plan PQC adoption. What factors influence the urgency?”*

(Explain X (data shelf‑life) + Y (migration time) > Z (quantum arrival) per [Mosca’s formula](https://en.wikipedia.org/wiki/Michele_Mosca?utm_source=chatgpt.com))

1. **Algorithm Tradeoffs**:

*“Compare different PQC algorithm families—lattice‑based, hash‑based, code‑based—in terms of security, key sizes, and deployment considerations.”*

(Discuss key‑size trade‑offs, example schemes like NTRU, Kyber, McEliece, SPHINCS+  .)

1. **Practical Implementation Challenges:**

*“What obstacles arise when integrating PQC into existing systems? Consider performance, compatibility, and hybrid approaches.”*

(Mention challenges in integrating new algorithms, hybrid encryption approaches, and industry experiments)

# AI Security

While not a specific category, AI Security overlaps significantly with both Agentic AI and PQC. Here are essay prompts that bridge both themes:

1. **Securing AI Agent Architectures**:

*“What are the unique security risks of multi‑agent AI systems, and how can protocols like the Model Context Protocol (MCP) mitigate or exacerbate these risks?”*

(Discuss issues like [prompt injection, tool poisoning, and the trickiness of secure orchestration](https://en.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com))

1. **Defensive AI in Cybersecurity:**

*“How are organizations using autonomous AI (Agentic AI) defensively to counter cyber threats? What limitations or restrictions are necessary?”*

(Based on real‑world deployment: AI triaging alerts, limitations set by Southwest Airlines, balancing [dependency vs. reliability](https://www.wsj.com/articles/battered-by-constant-hacks-security-chiefs-turn-to-ai-be17c37f?utm_source=chatgpt.com))

1. **Proactive Threat Detection**:

*“Evaluate the role of AI in pre-emptive cyber defense. Can Agentic AI anticipate and neutralize attacks before they occur? What are the attendant risks?”*

(Tie in customization, false positives, adversarial behaviors, and oversight issues.)

**Summary Table of Essay Questions**

|  |  |
| --- | --- |
| **Theme** | **Sample Essay Question** |
| Agentic AI Fundamentals | Define Agentic AI, distinguish from generative AI; assess cybersecurity applications. |
| Cyber Arms Race | Impact of Agentic AI on cyber defense and offense strategies. |
| Trust & Governance | Frameworks for secure, explainable, accountable Agentic AI (e.g., TRiSM). |
| Legal & Ethical Issues | Attribution, authorship, and liability in Agentic AI systems. |
| PQC 101 | Explain the need for PQC and its relevance. |
| Migration Urgency | Use Mosca’s theorem to evaluate PQC adoption timelines. |
| Algorithm Comparison | Trade-offs among PQC algorithm families. |
| Integration Challenges | Discuss real-world deployment hurdles and mitigation strategies. |
| Agentic AI Security Risks | Identify risks in multi-agent AI and evaluate mitigations like MCP. |
| Defensive Agentic AI | Analyse benefits and limitations of Agentic AI in real-time security operations. |