kaspersky
expert training

Large language models security

Course
program

№	Track	What you will learn	What you will practice	Lesson	Practice	Evaluation
00	About the course	• About the learning process • About your trainer • Course intro	—	About the learning process Course intro Introduction to Virtual Lab		—
01	Attacks on LLMs	• LLM security intro • Jailbreaks: manual attack and automated jailbreaks • Alignment removal • Prompt injection and prompt extraction • Token smuggling and sponge attacks	Understanding of how the attacks on LLMs work	Course intro LLM security intro Jailbreaks: manual attacks Automated jailbreaks Alignment removal Prompt injections Prompt extraction and token smuggling Attacks on LLMs: recap	Lab 01 • Task 1. Querying open API • Task 2. Dialogue management • Task 3. Direct prompt injection • Task 4. Token smuggling • Task 5. Obfuscating the output • Task 6. Prompt extraction • Task 7. Phishing attack via RAG poisoning • Task 8. Running a tool • Task 9. Opensource jailbreaks • Task 10. Prompt Injection + Past Tense Jailbreak • Task 11. BoN. Implementing the jailbreak • Task 12. BoN. LLM-as-judge. Prompt	Checkpoint quiz
02	LLM defenses	• Why and how to protect LLMs • Model-level defenses: alignment and unlearning • Prompt-level, system-level and service-level defenses • LLM Security Toolbox	Protecting your LLM applications from various attacks, including using LLM to protect itself	Why and how to protect LLMs Model-level defenses: training Model-level defenses: unlearning Prompt-level defenses	Lab 02 • Task 1. Prompt modifier • Task 2. Sandwich prompting • Task 3. Document enclosure • Task 4. Spotlighting • Task 5. Guarded conversation	Checkpoint quiz

№	Track	What you will learn	What you will practice	Lesson	Practice	Evaluation
				System-level defenses	• Task 6. LLM-based prompt injection detection • Task 7. LLM moderation • Task 8. Detecting attacks using perplexity • Task 9. Get started with Garak • Task 10. Get started with LLM-guard	
				Service-level defenses		
				LLM defenses: recap		
				LLM Security Toolbox		
03	LLM Security Frameworks	• Approaches to LLM Security Analysis • Real cases • Further study and recap of the topic	–	Approaches to LLM Security Analysis	–	Checkpoint quiz
				LLM Security: cases		
				LLM Security: further study and recap		

Own the knowledge, outsmart the threat

kaspersky.com

Discord server: kas.pr/g2j8

Help page: kas.pr/ii9f

kaspersky