



สถาบันไอจีเนียส
IT GENIUS INSTITUTE

COURSE ID
CYC-05

COURSE OUTLINE · รายละเอียดหลักสูตรอบรม

CYBERSECURITY / AI SECURITY

AI Security and LLM Red Teaming

ทดสอบเจาะระบบและป้องกันแอป LLM และ AI Agent ตามแนวทาง OWASP

● 12 ชั่วโมง

● 2 วัน

● พื้นฐาน Python และ Web Security

● ระดับ Intermediate



ระยะเวลา: 2 วัน

12 ชั่วโมง (2 วัน วันละ 6 ชม.) · 09:00-16:00 น.



รูปแบบ: Classroom / Hybrid

Onsite หรือ Online ผ่าน Microsoft Teams · Public / In-house



รายละเอียดหลักสูตร

www.itgenius.co.th

แอปที่ใช้ LLM ทั้ง Chatbot ระบบ RAG และ AI Agent ที่เรียกเครื่องมือได้เอง กำลังเข้าไปอยู่ในระบบงานจริงขององค์กร แต่ช่องโหว่ของแอปเหล่านี้ต่างจากเว็บแอปทั่วไป ข้อความธรรมดาในอีเมลหรือเอกสารก็สั่งให้โมเดลทำสิ่งที่ไม่ควรทำได้ ข้อมูลลับใน Context อาจรั่วออกมาในคำตอบ และ Agent ที่มีสิทธิ์มากเกินไปอาจถูกหลอกให้ลบข้อมูลหรือส่งข้อมูลออกนอกองค์กร เครื่องมือสแกนแบบเดิมจึงมองไม่เห็นความเสี่ยงส่วนใหญ่

หลักสูตรนี้พาผู้เรียนทดสอบความปลอดภัยของแอป LLM แบบลงมือจริงในมุมมองของผู้ป้องกัน โดยใช้ OWASP Top 10 for LLM Applications ฉบับ 2026 และ OWASP Top 10 for Agentic Applications เป็นกรอบ ทุก Lab โจมตีเฉพาะแอปตัวอย่างที่ออกแบบให้มีช่องโหว่โดยเจตนาซึ่งรันบนเครื่องของผู้เรียนเอง เริ่มจาก Prompt Injection ทั้งแบบตรงและแบบอ้อม ข้อมูลรั่ว ช่องโหว่ของ RAG ไปจนถึงการโจมตี AI Agent ผ่าน Tool และ MCP จากนั้นใช้ promptfoo ทำ Automated Red Teaming วาง Guardrails แบบหลายชั้น ฝึการทดสอบไว้ใน CI/CD และปิดท้ายด้วย Capstone ที่ต้องเขียนรายงานผลและข้อเสนอแก้ไขให้ทีมพัฒนา (อบรม 2 วัน วันละ 6 ชั่วโมง รวม 12 ชั่วโมง ระดับ Intermediate)

เครื่องมือ



OWASP LLM Top 10



promptfoo



Guardrails



Python



AI Agents

ราคาค่าอบรม · PUBLIC / IN-HOUSE

8,900 บาท / ท่าน

ราคายังไม่รวมภาษีมูลค่าเพิ่ม (VAT 7%) · รวมเอกสารประกอบการอบรม Lab Guide ไฟล์ docker-compose ของแอปตัวอย่าง ชุด Config ของ promptfoo แม่แบบรายงาน Red Team และ Certificate · รับจัด In-house ขอใบเสนอราคาได้

สมัครผ่านเว็บไซต์ ลด 10%



เพิ่มเพื่อน
LINE

☎ 02-570-8449

☎ 088-807-9770

💬 @itgenius

🌐 www.itgenius.co.th

01 วัตถุประสงค์ Objectives

- อธิบายความเสี่ยงตาม OWASP Top 10 for LLM Applications 2026 และเชื่อมกับสถาปัตยกรรมของแอปจริงได้
- วางขอบเขตและกติกากการทดสอบ AI Red Teaming อย่างมีจริยธรรมกับระบบที่ได้รับอนุญาต
- ทดสอบ Prompt Injection แบบตรงและแบบอ้อม รวมถึงการรั่วของข้อมูลและ Hidden Context
- ประเมินช่องโหว่ของระบบ RAG ตั้งแต่ Vector Store ไปจนถึงข้อมูลที่ถูกลวงยา
- วิเคราะห์ภัยคุกคามของ AI Agent ตาม OWASP Top 10 for Agentic Applications
- ใช้ promptfoo สร้างชุดทดสอบ Red Team อัตโนมัติและอ่านผลได้ถูกต้อง
- ออกแบบ Guardrails และการควบคุมสิทธิ์แบบหลายชั้นที่ลดความเสียหายเมื่อโมเดลถูกหลอก
- เขียนรายงานผลการทดสอบพร้อมระดับความเสี่ยงและข้อเสนอแก้ไขที่ทีมพัฒนานำไปใช้ได้

02 กลุ่มเป้าหมาย Target Audience

- Security Engineer และ Penetration Tester ที่ต้องทดสอบระบบที่มี LLM หรือ AI Agent
- AI Engineer และนักพัฒนาที่สร้าง Chatbot, RAG หรือ Agent และต้องการทำให้ปลอดภัยตั้งแต่ต้น
- ทีม AppSec และ DevSecOps ที่ต้องวางมาตรฐานการทดสอบแอป AI ขององค์กร
- Security Architect ที่ต้องรีวิวกการออกแบบระบบ AI ก่อนขึ้น Production

03 จุดเด่นของหลักสูตร Highlights

- อัปเดต **OWASP 2026** LLM Top 10 ฉบับ 2026 และ Agentic Top 10
- **ลงมือกับแอปจริง** แอปที่มีช่องโหว่โดยเจตนา รันบนเครื่องผู้เรียน
- **ครอบคลุม AI Agent** Tool Misuse, MCP และ Goal Hijack
- **Red Team อัตโนมัติ** promptfoo และทดสอบซ้ำใน CI/CD
- **บุมผู้ป้องกัน** ทุกการโจมตีจบที่วิธีแก้และ Guardrails

04 ความรู้พื้นฐาน Prerequisites

- เขียน Python ขั้นพื้นฐานได้ และอ่านโค้ดแอปที่เรียกใช้ LLM API ได้
- เข้าใจพื้นฐาน Web Security เช่น XSS, SQL Injection และ HTTP Request
- เคยใช้ ChatGPT หรือ LLM อื่น และเข้าใจแนวคิด RAG และ AI Agent เบื้องต้น
- โน้ตบุ๊ก RAM อย่างน้อย 16 GB ที่ติดตั้ง Docker Desktop, Node.js และ Python ได้

05 สิ่งที่ต้องเตรียมและเครื่องมือ Setup & Tools

รายการ	รายละเอียด
อุปกรณ์	Notebook RAM 16 GB ขึ้นไป ติดตั้ง Docker Desktop และ Docker Compose ได้
เป้าหมายทดสอบ	แอป Chatbot, RAG และ AI Agent ที่มีช่องโหว่โดยเจตนา รันบน Docker ในเครื่อง
LLM	โมเดลขนาดเล็กผ่าน Ollama บนเครื่อง หรือ LLM API Key ของผู้เรียน
เครื่องมือทดสอบ	promptfoo, Python และ Burp Suite Community Edition
กรอบอ้างอิง	OWASP Top 10 for LLM Applications 2026, OWASP Agentic Top 10 และ MITRE ATLAS

หมายเหตุ: เครื่องมือหลักเป็นโอเพนซอร์สหรือมีแบบใช้ฟรี ทุก Lab รันบนเครื่องของผู้เรียน หากต้องการทดสอบกับ LLM เชิงพาณิชย์ ผู้เรียนใช้ API Key ของตนเองหรือของบริษัท ค่าใช้ API เป็นไปตามการใช้งานจริง การทดสอบทั้งหมดทำกับแอปใน Lab ที่ได้รับอนุญาตเท่านั้น หลักสูตรนี้ต่อยอดจาก Ethical Hacking and Penetration Testing Fundamentals โดยเจาะเฉพาะแอปที่ใช้ LLM และ AI Agent ชื่อเครื่องมือและฟีเจอร์ด้าน AI เปลี่ยนเร็ว เนื้อหาจะอัปเดตตามเวอร์ชันล่าสุด ณ วันอบรม

06 ภาพรวมการเรียนรู้ 2 วัน Big Picture

วัน	หัวข้อหลัก
Day 1 Section 1-6	วางกรอบการทดสอบอย่างมีจริยธรรม, ไล่ OWASP Top 10 for LLM Applications 2026 แล้วลงมือทดสอบ Prompt Injection, ข้อมูลรั่ว, Hidden Context Exposure, Output Handling และช่องโหว่ของ RAG บนแอปตัวอย่างที่มีช่องโหว่โดยเจตนา
Day 2 Section 7-12	ภัยคุกคามของ AI Agent ตาม OWASP Top 10 for Agentic Applications, ทดสอบ Tool และ MCP, Automated Red Teaming ด้วย promptfoo, ออกแบบ Guardrails, ทดสอบต่อเนื่องใน CI/CD และ Capstone พร้อมรายงานผล

DAY
1

ช่องโหว่ของแอป LLM ตาม OWASP 2026

เนื้อหาการอบรม · Section 1-6

รวม
6 ชั่วโมง

1 Lab: AI Red Teaming และการเตรียมสนามทดสอบ

- AI Red Teaming ต่างจาก Penetration Test แบบเดิมอย่างไร และทำไมโมเดลเพียงอย่างเดียวป้องกันไม่ได้
- ขอบเขต สิทธิ และ Rules of Engagement ทดสอบเฉพาะระบบที่ได้รับอนุญาตเป็นลายลักษณ์อักษร
- ติดตั้งแอปตัวอย่างที่มีช่องโหว่โดยเจตนาด้วย Docker และเชื่อมกับโมเดลผ่าน Ollama
- Lab: สำรวจแอปเป้าหมาย วาด Data Flow และระบุจุดที่ข้อมูลภายนอกไหลเข้าโมเดล

2 OWASP Top 10 for LLM Applications 2026

- ภาพรวมรายการ LLM01 ถึง LLM10 ฉบับ 2026 และสิ่งที่เปลี่ยนจากฉบับ 2025
- แนวคิดหลักของฉบับใหม่: ยอมรับว่าโมเดลถูกหลอกได้ แล้วออกแบบระบบรอบตัวให้เสียหายน้อยที่สุด
- Threat Modeling แอป LLM: ระบุทรัพย์สิน ผู้โจมตี และเส้นทางการโจมตี
- Workshop: จัดลำดับความเสี่ยงของแอปตัวอย่างตามรายการ OWASP

3 Lab: Prompt Injection แบบตรงและแบบอ้อม

- Direct Prompt Injection และ Jailbreak รูปแบบที่พบบ่อย
- Indirect Prompt Injection ผ่านเว็บ อีเมล ไฟล์ PDF และเอกสารที่โมเดลอ่าน
- เทคนิคหลบตัวกรอง เช่น การเข้ารหัสข้อความ การสลับภาษา และการแบ่งคำสั่งหลายรอบ
- Lab: ฝังคำสั่งในเอกสารให้ Chatbot ตัวอย่างทำงานผิดจากที่ออกแบบ แล้วบันทึกหลักฐาน

4 Lab: ข้อมูลรั่วและ Hidden Context Exposure

- Sensitive Information Disclosure: ข้อมูลส่วนบุคคล ความลับทางธุรกิจ และ Credential ในคำตอบ
- Hidden Context Exposure: การรั่วของ System Prompt, Tool Schema และกฎการทำงานภายใน
- แนวทางป้องกัน: แยกสิทธิ์ข้อมูล ปิดบังข้อมูล และตรวจสอบคำตอบก่อนส่งออก
- Lab: ดึง Hidden Context และข้อมูลลูกค้ายาลงออกจากแอปตัวอย่าง แล้วเสนอวิธีแก้

5 Lab: Output Handling และ Unbounded Consumption

- Improper Output Handling: เมื่อคำตอบของโมเดลกลายเป็น XSS, SQL หรือคำสั่ง Shell
- การ Render Markdown และรูปภาพที่เปิดช่องให้ข้อมูลรั่วออกนอกระบบ
- Unbounded Consumption: คำขอเดียวที่ทำให้เรียกโมเดลและ Tool วนซ้ำจนค่าใช้จ่ายพุ่ง
- Lab: ทดสอบให้คำตอบของโมเดลรันสคริปต์ในหน้าเว็บ แล้วแก้ด้วย Output Encoding

6 Lab: ช่องโหว่ของ RAG และ Supply Chain

- Vector and Embedding Weaknesses: สิทธิการค้นหาข้ามผู้ใช้และข้อมูลที่ไม่ควรถูกดึง
- Data and Model Poisoning: วางเอกสารปลอมในคลังความรู้เพื่อบิดคำตอบ
- Misinformation และ Supply Chain: เชื้อคำตอบโดยไม่ตรวจ และโมเดลหรือ Package ที่ไม่น่าเชื่อถือ
- Lab: ทำ RAG Poisoning กับคลังเอกสารตัวอย่าง แล้วออกแบบการตรวจแหล่งที่มาและสิทธิ์



DAY
2

Agentic Threats, Red Team อัตโนมัติ และการป้องกัน

เนื้อหาการอบรม · Section 7-12

รวม
6 ชั่วโมง

7 Excessive Agency และ Agentic Top 10

- Excessive Agency: สิทธิ์ ฟังก์ชัน และความเป็นอิสระที่มากเกินไปจนความจำเป็นของ Agent
- OWASP Top 10 for Agentic Applications: Goal Hijack, Tool Misuse และ Identity Abuse
- Memory Poisoning, Cascading Failures, Rogue Agents และการเชื่อ Agent มากเกินไป
- Workshop: Threat Model ของ Agent ตัวอย่างที่อ่านอีเมลและเรียก API ได้

8 Lab: ทดสอบ AI Agent, Tool และ MCP

- Agent Goal Hijack ผ่านข้อมูลที่ Agent อ่านระหว่างทำงาน
- Tool Misuse: ใช้เครื่องมือที่ได้รับอนุญาตต่อกันจนเกิดผลเสียหาย
- ความเสี่ยงของ MCP Server เช่น คำอธิบาย Tool ที่ถูกแก้ไขและสิทธิ์ที่กว้างเกินไป
- Lab: ชักนำ Agent ตัวอย่างให้ส่งข้อมูลออกนอกกระบวน แล้วปิดช่องด้วย Least Privilege

9 Lab: Automated Red Teaming ด้วย promptfoo

- ตั้งค่า promptfoo ให้ยังทดสอบไปยังแอปเป้าหมายผ่าน HTTP หรือ Python Provider
- เลือก Plugin และ Strategy ให้ตรงกับความเสี่ยงตาม OWASP ของแอป
- อ่านรายงานผล แยกผลบวกผลลบ และจัดลำดับช่องโหว่ที่ต้องแก้ไข
- Lab: สแกนแอปตัวอย่างทั้งก่อนและหลังแก้ไข แล้วเปรียบเทียบผล

10 Lab: Guardrails และ Defense in Depth

- ชั้นการป้องกัน: ตรวจสอบ Input, ตรวจสอบ Output, แยกสิทธิ์ Tool และ Human-in-the-loop
- แยกข้อมูลที่ไม่น่าเชื่อถือออกจากคำสั่ง และจำกัดผลกระทบเมื่อโมเดลถูกหลอก
- ขอการยืนยันจากมนุษย์ก่อนทำงานที่ย้อนกลับไม่ได้ และบันทึก Audit Log ของ Agent
- Lab: เพิ่ม Guardrails ให้แอปตัวอย่าง แล้ววัดผลด้วยชุดทดสอบเดิม

11 ทดสอบต่อเนื่องและเฝ้าระวังใน Production

- ฝังชุดทดสอบ Red Team ไว้ใน CI/CD เพื่อจับปัญหาเมื่อเปลี่ยน Prompt หรือโมเดล
- Logging และ Monitoring ของแอป LLM เพื่อตรวจจับการโจมตีที่เกิดขึ้นจริง
- ประเด็น PDPA เมื่อ Prompt และ Log มีข้อมูลส่วนบุคคล
- Lab: ตั้ง GitHub Actions ให้รัน promptfoo ทุกครั้งที่มีการแก้ไข Prompt

12 Workshop: Capstone Red Team Engagement

- รับโจทย์แอป AI Agent ตัวใหม่ พร้อมขอบเขตและกติกาการทดสอบ
- ทำ Threat Model ทดสอบด้วยมือ และสแกนด้วย promptfoo
- จัดระดับความรุนแรงของช่องโหว่และเชื่อมกับรายการ OWASP
- เขียนรายงานผลพร้อมข้อเสนอแก้ไข แล้วนำเสนอและรีวิวยร่วมกัน

07 เรียนจบแล้วได้อะไร Outcomes

- เข้าใจ OWASP 2026 LLM Top 10 และ Agentic Top 10
- ทดสอบแอป LLM ได้ Prompt Injection, ข้อมูลรั่ว และ RAG
- ประเมิน AI Agent ได้ Tool Misuse, MCP และ Goal Hijack
- ใช้ Red Team อัตโนมัติ promptfoo และทดสอบซ้ำใน CI/CD
- ออกแบบการป้องกัน Guardrails และ Least Privilege
- เขียนรายงานผลได้ ระดับความเสี่ยงและข้อเสนอแก้ไข

08 สิ่งที่คุณเรียนจะได้รับ Deliverables

- เอกสารประกอบการอบรมฉบับสมบูรณ์ครบทั้ง 2 วัน และ Lab Guide ที่ละขั้นตอน
- ไฟล์ docker-compose ของแอปตัวอย่างที่มีช่องโหว่โดยเจตนา สำหรับฝึกต่อบนเครื่องตนเอง
- ชุด Config ของ promptfoo, ตัวอย่าง Guardrails และ Workflow ของ GitHub Actions
- แม่แบบรายงาน Red Team และ Checklist ความปลอดภัยของแอป LLM
- ประกาศนียบัตรจบหลักสูตร (Certificate of Completion) และช่องทางสอบถามวิทยากรหลังจบหลักสูตร

09 ข้อมูลผู้สอน Instructor

อ.สามิตร โกยม ผู้สอนและผู้ปฏิบัติงานด้านความปลอดภัยไซเบอร์ที่ทดสอบและออกแบบการป้องกันให้ระบบที่ใช้ LLM และ AI Agent จุดเด่นคือการสอนทุกการโจมตีควบคู่กับวิธีป้องกัน ให้ผู้เรียนเห็นทั้งมุมมองผู้โจมตีและผู้แก้ไข เน้นจริยธรรมและการทดสอบเฉพาะระบบที่ได้รับอนุญาต และพาเขียนรายงานที่ทีมพัฒนานำไปแก้ไขจริง ไม่ใช่แค่รายการช่องโหว่

ราคาค่าอบรม 8,900 บาท/ท่าน · สนใจ IN-HOUSE ขอใบเสนอราคา

8,900 บาท / ท่าน (ยังไม่รวม VAT 7%)

สมัครผ่านเว็บไซต์ ลด 10%

☎ 02-570-8449

☎ 088-807-9770

LINE @itgenius

✉ contact@itgenius.co.th



เพิ่มเพื่อน
LINE