Agentic AI Testing & Evaluation for IT Professionals
(ทดสอบ ประเมิน และกำกับ AI Agent ก่อนขึ้น Production)

ระยะเวลาการอบรม 2 วัน

หลักการและเหตุผล

ปัจจุบันองค์กรไทยจำนวนมากเริ่มนำ AI Agent และ LLM-based Application เข้าสู่การใช้งานจริง ทั้งในรูปแบบ Chatbot, RAG-based Knowledge System และ Workflow Automation อย่างไรก็ตาม ความท้าทายสำคัญไม่ได้อยู่ที่ "การทำให้ Agent ทำงานได้" แต่อยู่ที่ "การทำให้มั่นใจว่า Agent ตอบถูก ปลอดภัย และควบคุมได้ก่อนขึ้น Production" ซึ่งครอบคลุมทั้งปัญหา Hallucination, Tool Misuse, Prompt Injection และการรั่วไหลของข้อมูลส่วนบุคคลตามข้อกำหนดของ พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 (PDPA) ในทางปฏิบัติยังขาดแนวทางการทดสอบและประเมินผลที่เป็นระบบสำหรับทีมพัฒนา ทำให้หลายองค์กรต้อง deploy โดยอาศัย "ความรู้สึก" มากกว่าตัวชี้วัดที่วัดได้จริง

หลักสูตรนี้ถูกออกแบบให้แตกต่างจากคอร์ส "สร้าง AI Agent" ทั่วไป โดยมุ่งเน้นเฉพาะ 3 ด้านที่เป็นช่องว่างสำคัญขององค์กร ได้แก่ Testing (การออกแบบ Test Suite และ Golden Dataset), Evaluation (การวัดผลด้วย Metrics เชิงคณิต เช่น Faithfulness, Answer Relevancy, Context Precision ผ่านเครื่องมือ RAGAS และ DeepEval) และ Governance (การตรวจจับ PII, Prompt Injection และการวาง Evaluation Gate ใน CI/CD Pipeline) จึงเหมาะเป็นพิเศษกับองค์กรในกลุ่ม Regulated Industry เช่น Banking, Insurance, Healthcare และหน่วยงานภาครัฐ ที่ต้องการหลักฐานเชิงประจักษ์ (Evidence-based Compliance) ก่อน deploy AI Agent สู่ผู้ใช้จริง เมื่อเรียนจบ ผู้เข้าอบรมจะได้ Lab Notebook, Golden Test Dataset และ CI/CD Configuration ที่สามารถนำกลับไปใช้กับ Repository ขององค์กรได้ทันที และด้วยขอบเขตเนื้อหาที่ครอบคลุมตั้งแต่ Agent Architecture, Failure Debugging, Evaluation Tooling จนถึง Governance และ CI/CD Integration หลักสูตรจึงออกแบบเป็น 2 วันเต็ม (Hands-on Lab 60% + บรรยาย 40%) เพื่อให้ผู้เรียนได้ลงมือทำ Lab ครบทั้ง 3 โมดูลในจังหวะที่เหมาะสม ไม่เร่งรัด และมีเวลาให้ผู้สอน debug ปัญหาจริงร่วมกับผู้เรียนในห้องได้

วัตถุประสงค์

- เข้าใจการทำงานของ AI Agent (ReAct Loop + Tool Architecture) และสามารถพัฒนา Agent ของตัวเองได้ในระดับเขียนโค้ด
- วิเคราะห์และแก้ปัญหา Agent ที่ทำงานผิดพลาด โดยใช้ log เพื่อ debug และระบุสาเหตุได้อย่างเป็นระบบ
- ออกแบบ Test Suite สำหรับ LLM / AI Agent โดยใช้ RAGAS, DeepEval และ pytest ให้เหมาะกับงานจริง
- วัดผลคุณภาพของ Agent ด้วย Metrics สำคัญ เช่น Faithfulness, Answer Relevancy และ Context Precision
- สร้าง Golden Dataset (20+ Test Cases) เพื่อทดสอบความถูกต้องและความเสถียรในสถานการณ์จริง
- ออกแบบการตรวจจับความเสี่ยง เช่น PII / Data Leakage สำหรับ prompt, log และ vector store
- วางระบบ Evaluation Gate ใน CI/CD เพื่อควบคุมคุณภาพก่อน deploy และหยุดอัตโนมัติเมื่อไม่ผ่านเกณฑ์
- สร้าง Lab Notebook และ GitHub Portfolio เพื่อแสดงผลงานและต่อยอดในสายอาชีพ

รายละเอียดหลักสูตร

Module 1 | Agent Architecture & Failure Debugging
อ่าน trace log เป็น รู้ว่า Agent พังตรงไหน ก่อนผู้ใช้จะเจอปัญหา
Key Topics

- เข้าใจโครงสร้างการทำงานของ Agent แบบลงมือทำ
  (ReAct Loop: Thought → Action → Observation พร้อม Python example)
- ออกแบบและใช้งาน MCP Tools Architecture
  - Tool Calling Protocol
  - Schema Design
  - Error Handling Pattern
- วิเคราะห์ Failure Mode หลักของ Agent (6 รูปแบบ)
  พร้อมแนวทางตรวจจับจาก log จริง
  - Hallucination (ตอบไม่ตรง source)
  - Tool Misuse (เรียก tool ผิด / parameter ไม่ถูก)
  - Infinite Loop (วนไม่จบ)

    - Prompt Injection (ถูก override instruction)
    - Data Leakage (ข้อมูลสำคัญหลุดออก)
    - Cost Runaway (token usage สูงผิดปกติ)
- Case Study: วิเคราะห์ trace log จากระบบจริง
  (เช่น Banking Chatbot / Internal Knowledge Base)

Hands-on Lab 1

- วิเคราะห์ trace log ของ Agent ที่มีปัญหา
- ใช้ Python parse log และระบุ failure mode
- สร้าง Test Coverage Map เพื่อป้องกัน

Module 2 | LLM Evaluation: RAGAS & DeepEval Hands-on
วัดผลให้ชัด รู้ว่า Agent ดีพอจะใช้งานจริงหรือยัง
Key Topics

- เข้าใจ Evaluation Metrics ที่ใช้จริงในงาน
  เช่น Faithfulness, Answer Relevancy, Context Precision, Context Recall
  (ทั้งความหมาย + วิธีคำนวณ)
- ออกแบบ Hallucination Detection Pipeline
  เพื่อจับคำตอบที่ "ดูถูก แต่ผิด"
- การใช้งานเครื่องมือ Evaluation
    - RAGAS Workflow: สร้าง dataset → run evaluate() → วิเคราะห์ผล
    - DeepEval Workflow: เขียน test case → ตั้ง threshold → integrate กับ pytest
- เปรียบเทียบการใช้งาน
    - RAGAS (เหมาะกับ research / analysis)
    - DeepEval (เหมาะกับ production / automation)
- การจัดการ Cost vs Latency ในการ evaluate

Hands-on Lab 2

- รัน RAGAS + DeepEval บน dataset (RAG-based Customer Support)
- เปรียบเทียบผลลัพธ์จาก 2 เครื่องมือ
- ระบุจุดที่ Agent ต้องปรับปรุง
- บันทึกผลลัพธ์ลง Notebook พร้อม insight

Module 3 | Governance Testing: PDPA, Bias & CI/CD Gate
มั่นใจก่อน deploy ว่าผ่านทั้งคุณภาพและ compliance
Key Topics

- PDPA Compliance Testing
    - ตรวจจับ PII ด้วย regex และ NER
    - ครอบคลุม prompt, log และ vector store
- Prompt Injection Testing
    - รวม attack pattern ที่ควร test ก่อนใช้งานจริง
- Bias Detection
    - ออกแบบ balanced test set
    - วัด fairness metric ในเชิงปฏิบัติ
- Golden Set Design
    - สร้าง test dataset (20+ cases)
    - ครอบคลุม edge case ที่เกิดขึ้นจริงในระบบ
- CI/CD Evaluation Gate
    - เชื่อม DeepEval กับ GitHub Actions / GitLab CI
    - ตั้ง threshold และ block การ deploy อัตโนมัติเมื่อคุณภาพไม่ผ่าน

Hands-on Lab 3

- สร้าง Golden Set สำหรับ use case ของตัวเอง
- รัน PDPA Scanner ตรวจจับข้อมูลสำคัญ
- เขียน CI/CD Pipeline (เช่น GitHub Actions) เพื่อควบคุมคุณภาพก่อน deploy

Output ที่ผู้เรียนจะได้

- Lab Notebook ครบทั้ง 3 Modules
- Test Dataset (Golden Set) สำหรับ use case ของตัวเอง
- CI/CD Config สำหรับ Eval Gate
- พร้อมต่อยอดเป็น GitHub Portfolio ใช้โชว์ทักษะหรือสมัครงานได้ทันที

สิ่งที่จำเป็นต้องมีก่อนการอบรม

(เพื่อให้เรียนได้เต็มประสิทธิภาพ และลงมือทำ Lab ได้ทันที)

- มีพื้นฐาน Python ระดับเขียน function และเรียก API ได้
- เคยใช้งาน LLM API เช่น OpenAI, Anthropic หรือเครื่องมือใกล้เคียง
- มี Laptop ส่วนตัว พร้อมใช้งาน และติดตั้ง Python 3.10+ เรียบร้อย

เครื่องมือที่แนะนำ (Optional)

- RAGAS, DeepEval, pytest
- Jupyter Notebook
- LangSmith หรือ LangFuse

ผู้ที่เหมาะจะเข้ารับการอบรม

- AI / ML Engineer
  ที่กำลังพัฒนา หรือเตรียม deploy AI Agent ในระบบจริง
- Backend / Full-stack Developer
  ที่มีการ integrate LLM API และต้องการยกระดับการทดสอบและควบคุมคุณภาพ
- DevOps / MLOps / Platform Engineer
  ที่ต้องออกแบบ pipeline สำหรับ evaluation และควบคุมก่อนขึ้น production
- QA / Test Engineer
  ที่ต้องทดสอบระบบที่มี LLM หรือ AI Agent เป็นส่วนประกอบ
- Data Engineer / Data Scientist
  ที่ทำงานกับ RAG, Vector Database และต้องการวัดคุณภาพผลลัพธ์อย่างเป็นระบบ

NETWORK TRAINING CENTER CO.,LTD. (NTC) | www.trainingcenter.co.th
Call us today 0-2634-7993-4
177/1 BUI Bldg., 14th Fl., Unit 1, 3 & 4, Surawongse Rd., Suriyawongse, Bangrak, Bangkok, THAILAND | Email: sales@trainingcenter.co.th