AI Security เบื้องต้น

รู้ทันความเสี่ยง ป้องกันภัยไซเบอร์ และใช้งาน AI อย่างปลอดภัย

Harmoniz
InterSoftware

เรียบเรียง — นรกิตติ์ ศรีสัมฤทธิ์

AI Security เบื้องต้น

รู้ทันความเสี่ยง ป้องกันภัยไซเบอร์ และใช้งาน AI อย่างปลอดภัย

ผู้เขียน/เรียบเรียง: นรกิตติ์ ศรีสัมฤทธิ์ (Norrakith Srisumrith)
บรรณาธิการ: ทีม HMZ
ออกแบบปก: ChatGPT
สำนักพิมพ์: บริษัท ฮาโมนิซ อินเตอร์ซอฟต์แวร์ จำกัด / www.harmoniz.biz
ปีที่พิมพ์: พ.ศ. 2569
ครั้งที่พิมพ์: พิมพ์ครั้งที่ 1 — พฤษภาคม 2569
ISBN (e-book) : 978-616-631-824-1
ราคา: 395 บาท

Harmoniz
InterSoftware

จัดทำโดย บริษัท ฮาโมนิซ อินเตอร์ซอฟต์แวร์ จำกัด

ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ

นรกิตติ์ ศรีสัมฤทธิ์.
AI Security เบื้องต้น รู้ทันความเสี่ยง ป้องกันภัยไซเบอร์ และใช้งาน AI อย่างปลอดภัย.—
นนทบุรี : [ม.ป.พ.], 2569.
198 หน้า.

1. ปัญญาประดิษฐ์. 2. คอมพิวเตอร์. I. ชื่อเรื่อง.

006.3
ISBN 978-616-631-824-1

คำนำ

โลกกำลังเปลี่ยนแปลงเร็วกว่าที่ระบบการศึกษาจะตามทัน

ในวันที่ผมเริ่มเขียนหนังสือเล่มนี้ (ต้นปี 2024) ChatGPT เพิ่งเปิดตัวได้เพียงปีเศษ แต่ผู้ใช้ทะลุหลักร้อยล้านคนไปแล้ว องค์กรทั่วโลกเริ่มทดลองใช้ Generative AI ในกระบวนการทำงาน นักศึกษาใช้ AI ช่วยทำรายงาน อาจารย์ใช้ AI ออกข้อสอบ วิศวกรใช้ AI เขียนโค้ด นักบัญชีใช้ AI ตรวจสอบตัวเลข แต่ในวันที่ผมเขียนคำนำนี้ (พฤษภาคม 2026) โลกเปลี่ยนไปอีกขั้น Agentic AI — ระบบ AI ที่สามารถวางแผน ตัดสินใจ และลงมือทำด้วยตัวเอง — ถูกปล่อยเข้าสู่ Production ในองค์กรชั้นนำ และในสัปดาห์ที่แล้วเอง มีรายงานว่า AI Agent ที่ได้รับสิทธิ์เกินจำเป็น ลบฐานข้อมูลของบริษัทแห่งหนึ่งภายใน 9 วินาที

นี่คือโลกที่เราอยู่

หนังสือ "AI Security เบื้องต้น" เล่มนี้ถือกำเนิดขึ้นจากคำถามง่าย ๆ ที่ผมถามตัวเอง: "คนทั่วไปที่ใช้ AI ทุกวัน — นักศึกษา, พนักงานออฟฟิศ, ผู้บริหาร, อาจารย์ — จะรู้ได้อย่างไรว่ากำลังใช้ AI อย่างปลอดภัย? จะรู้ได้อย่างไรว่าข้อมูลของตัวเองอยู่ที่ไหน? จะรู้ได้อย่างไรว่าอีเมลที่ได้รับเป็นของจริงหรือ AI ปลอม? และถ้าพวกเขาอยากเรียนเรื่องนี้อย่างจริงจัง จะเริ่มจากตรงไหน?"

คำตอบที่ผมพบคือ: **ไม่มีหนังสือภาษาไทยที่ตอบคำถามเหล่านี้อย่างครอบคลุมและอ่านง่ายเลย** ตำรา AI Security ส่วนใหญ่เป็นภาษาอังกฤษ เขียนสำหรับผู้เชี่ยวชาญทางเทคนิค หรือไม่ก็เป็น Whitepaper สั้น ๆ ที่ลงรายละเอียดไม่ถึง จึงเกิดเป็นหนังสือเล่มนี้ขึ้น

หนังสือเล่มนี้คืออะไร

หนังสือเล่มนี้ถูกออกแบบให้เป็น **"ตำราเล่มแรก"** สำหรับคนไทยที่ต้องการเข้าใจ AI Security อย่างเป็นระบบ โดยมีคุณสมบัติ 5 ประการ:

1. **ภาษาที่มนุษย์อ่านเข้าใจ:** หลีกเลี่ยงศัพท์เทคนิคที่ไม่จำเป็น ใช้ตัวอย่างจากชีวิตประจำวัน และอธิบายมโนทัศน์ซับซ้อนด้วยการเปรียบเทียบ
2. **ครอบคลุมทั้ง 360 องศา:** ตั้งแต่ภัยคุกคาม (Prompt Injection, ข้อมูลรั่วไหล, AI-Generated Phishing, Deepfake, Adversarial ML) ไปจนถึงการป้องกัน (การตั้งค่า ChatGPT ให้ปลอดภัย, DLP สำหรับ AI, AI Governance) และการต่อยอด (งานวิจัย, เส้นทางอาชีพ)
3. **อิงเหตุการณ์จริง:** หนังสือเล่มนี้ใช้กรณีศึกษาจากเหตุการณ์และรายงานข่าวในช่วงปี 2023–2026 เช่น Samsung, deepfake fraud, Vercel และ PocketOS โดยแต่ละกรณีจะระบุแหล่งข้อมูลและขอบเขตของหลักฐานให้ชัดเจน เพื่อหลีกเลี่ยงการสรุปเกินกว่าข้อมูลที่ยืนยันได้
4. **ใช้ได้จริง:** บทที่ 7 และบทที่ 9 คือ "คู่มือการใช้งาน" และ "Checklist" ที่คนทั่วไปและองค์กรสามารถนำไปปรับใช้ได้ทันที ไม่ต้องรอทีม IT
5. **เป็นตำราเรียนได้:** แต่ละบทมีวัตถุประสงค์การเรียนรู้ คำถามทบทวน และเอกสารอ้างอิง อาจารย์สามารถหยิบไปใช้เป็นตำราเสริมในวิชา Cybersecurity, AI Ethics, IT Governance หรือสร้างเป็นวิชาเลือกใหม่ "AI Security" ได้เลย

หนังสือเล่มนี้สำหรับใคร

- **นักศึกษาระดับปริญญาตรีและโท** สาขาวิทยาการคอมพิวเตอร์, IT, Cybersecurity, AI, หรือสาขาอื่นที่ต้องใช้ AI — หนังสือเล่มนี้จะช่วยให้คุณ

ใช้ AI อย่างรู้เท่าทัน และเป็นใบเบิกทางสู่อาชีพเกิดใหม่ในปี 2026 อย่าง AI Red Teamer, AI Governance Officer, หรือ ML Privacy Engineer

- **อาจารย์และผู้สอน** ที่กำลังมองหาตำราเสริมหรือตำราหลักสำหรับวิชา AI Security, AI Ethics, หรือ Generative AI Governance — หนังสือเล่มนี้ถูกออกแบบให้เป็น "Plug and Play": แต่ละบทมีคำถามทบทวน กรณีศึกษา และสามารถสอนแบบแยกบทได้
- **บุคลากรด้าน IT และ Cybersecurity** ที่ต้องการอัปสกิลสู่ AI Security — หนังสือเล่มนี้จะเป็นสะพานเชื่อมระหว่างความรู้ Cybersecurity แบบดั้งเดิมกับโลกใหม่ของ Adversarial ML, โมเดลภาษาใหญ่ (LLM), Security, และ AI Risk Management
- **ผู้บริหารและผู้ทำงานทั่วไป** ที่ใช้ AI ในชีวิตประจำวันและต้องการรู้ว่า "ข้อมูลของฉันปลอดภัยไหม", "ใช้ AI อย่างไรไม่ให้รั่ว", และ "จะรู้ได้อย่างไรว่าอีเมลที่ได้รับเป็น AI ปลอม"
- **ผู้สนใจทั่วไป** ที่อยากรู้เท่าทันเทคโนโลยีที่กำลังเปลี่ยนโลก โดยไม่จำเป็นต้องมีพื้นฐานเขียนโค้ดหรือคณิตศาสตร์ขั้นสูง

แต่ละบทจะเริ่มด้วย "ลองนึกภาพ" — สถานการณ์จำลองที่ทำให้คุณเห็นภาพความเสี่ยงก่อนที่จะอธิบายทางเทคนิค เสริมด้วยกรณีศึกษาจริง สถิติอ้างอิง และปิดท้ายด้วยคำถามทบทวนที่ใช้ในห้องเรียนได้

สิ่งที่หนังสือเล่มนี้ไม่ใช่

หนังสือเล่มนี้ **ไม่ใช่**:

- คู่มือการเขียนโค้ด AI Security ขั้นสูง (เราแนะนำแนวทาง แต่ไม่ลงลึกถึงขั้น Implementation)

- หนังสือที่อ้างว่า "รู้ทุกอย่างเกี่ยวกับ AI Security" — สาขานี้เคลื่อนไหวเร็วเกินกว่าหนังสือเล่มใดจะครอบคลุมทุกอย่างได้ (โดยเฉพาะเมื่อคุณอ่านบรรทัดนี้หลังจากปี 2026)
- หนังสือที่ฟันธงว่า "นี่คือวิธีป้องกันที่สมบูรณ์แบบ" — ปัญหาหลายอย่างใน AI Security ยังไม่มีทางแก้ที่สมบูรณ์ และหนังสือเล่มนี้จะบอกตามตรงเมื่อมันเป็นเช่นนั้น

หากคุณอ่านถึงตรงนี้แล้ว — ผมเชื่อว่าคุณคือคนที่ "ไม่แค่ใช้ AI" แต่ "พร้อมจะเข้าใจมัน"

ไม่ว่าคุณจะเป็นใคร — นักศึกษา นักพัฒนา ผู้บริหาร หรือประชาชนทั่วไป — การรู้เท่าทัน AI และการใช้มันอย่างปลอดภัยคือทักษะพื้นฐานของพลเมืองในทศวรรษนี้ ไม่ต่างจากการรู้จักล็อกประตูบ้าน หรือไม่บอกรหัสผ่านให้ใคร

ขอให้หนังสือเล่มนี้เป็นเพื่อนร่วมทางของคุณในการเดินทางสู่โลกของ AI Security

สิ้นสุดคำนำ

พฤษภาคม 2026

สารบัญ

บทที่ 1

AI คืออะไร และทำไมต้องมี AI Security

พื้นฐานสำคัญสู่การใช้งาน AI อย่างปลอดภัย

ลองนึกภาพ: คุณเป็นนักศึกษาที่กำลังนั่งเขียนรายงานตอนดึก เปิด ChatGPT ช่วยสรุปบทความวิชาการให้เสร็จทันเช้าวันรุ่งขึ้น สิ่งที่คุณคัดลอกลงในแชตมีทั้งเนื้อหาจากเอกสารภายในของอาจารย์ที่ยังไม่เผยแพร่ และข้อมูลส่วนตัวของกลุ่มตัวอย่างที่แอบส่งต่อกันมาเพียงไม่กี่วินาที ข้อความนั้นกลายเป็นส่วนหนึ่งของข้อมูลบนคลาวด์ที่คุณไม่รู้ว่าถูกเก็บไว้ที่ไหน ใครเข้าถึงได้บ้าง และจะถูกนำไปเทรนโมเดลต่อหรือไม่นี่ไม่ใช่พล็อตหนังไซไฟอีกต่อไป แต่มันคือความเสี่ยงจริงที่เกิดขึ้นได้ในยุคที่ **ปัญญาประดิษฐ์ หรือ AI** กลายเป็นเครื่องมือสามัญประจำโต๊ะทำงานของเรา

เพราะ AI ไม่ได้เป็นแค่ผู้ช่วยอัจฉริยะอีกแล้ว แต่มันคือโครงสร้างพื้นฐานดิจิทัลตัวใหม่ และเหมือนกับโครงสร้างพื้นฐานทุกชนิด มันต้องการ **ความมั่นคงปลอดภัย (Security)** ที่ออกแบบมาสำหรับมันโดยเฉพาะ

1.1 โลกที่ AI อยู่ในทุกอณูของชีวิต

คุณอาจไม่รู้ตัว แต่ AI ได้แทรกซึมเข้าสู่ชีวิตประจำวันอย่างแนบเนียน ไม่ว่าจะเป็นระบบแนะนำวิดีโอใน YouTube การปลดล็อกมือถือด้วยใบหน้า ฟิลเตอร์แต่งรูปในโซเชียลมีเดีย ไปจนถึงการช่วยแพทย์อ่านฟิล์มเอกซเรย์ ในภาคธุรกิจ AI ถูกใช้คาดการณ์ยอดขาย ตรวจจับธุรกรรมทุจริต และคัดกรองผู้สมัครงาน

ปลายปี 2022 พ.ศ. 2565 ChatGPT เปิดให้บริการสาธารณะด้วย Generative AI ที่สามารถ “สร้าง” เนื้อหาใหม่ ไม่ว่าจะเป็นข้อความ รูปภาพ โค้ด หรือแม้แต่วิดีโอ ภายในเวลาเพียง 2 เดือนมีผู้ใช้ทะลุ 100 ล้านคน สะท้อนว่า AI ไม่ใช่เทคโนโลยีสำหรับผู้เชี่ยวชาญอีกต่อไป แต่เป็น *ของใช้ทั่วไป* ที่ใครก็หยิบมาใช้ได้

แต่ของที่ใช้กันแพร่หลายย่อมตกเป็นเป้าหมายของการโจมตี **ยิ่งมีคนพึ่งพามากเท่าไร ภัยก็ยิ่งอันตรายมากขึ้นเท่านั้น**

1.2 AI จริง ๆ แล้วคืออะไรกันแน่

เพื่อให้เข้าใจ AI Security เราจำเป็นต้องเข้าใจก่อนว่า AI ในปัจจุบันคืออะไร (และไม่ใช่อะไร) เพราะสื่อมักทำให้ภาพของ AI ดูเกินจริง อย่างหุ่นยนต์มีสำนึกแบบในหนัง ซึ่งห่างไกลจากความเป็นจริง

ปัญญาประดิษฐ์ (Artificial Intelligence – AI) หมายถึง ระบบคอมพิวเตอร์ที่สามารถทำงานซึ่งปกติต้องใช้สติปัญญามนุษย์ เช่น การรู้จำภาพ การเข้าใจภาษา หรือการตัดสินใจ

ในทางเทคนิค ปัจจุบัน AI ส่วนใหญ่คือ **Machine Learning (ML)** หรือการเรียนรู้ของเครื่อง ที่เรียนรู้จากข้อมูลจำนวนมหาศาลเพื่อหาแบบแผนและทำนายผล โดยไม่ต้องเขียนกฎตายตัวลงไป และเมื่อเราพูดถึง **Deep Learning** ก็คือ ML ที่ใช้โครงข่ายประสาท

เทียมซับซ้อนหลายชั้น (Neural Networks) ซึ่งเป็นรากฐานของเทคโนโลยีล้ำสมัยเกือบทั้งหมดในปัจจุบัน

AI ยุคนี้โดยเฉพาะ **Generative AI (GenAI)** เช่น ChatGPT, Gemini, Claude หรือระบบสร้างภาพอย่าง Midjourney ทำงานโดยการ “ทำนายคำหรือพิกเซลต่อไป” จากข้อมูลที่ถูกเทรนมา โมเดลเหล่านี้ไม่ได้ “เข้าใจ” โลกอย่างมนุษย์ หากแต่เรียนรู้ความน่าจะเป็นของภาษาและภาพจนสามารถเลียนแบบการสร้างสรรค์ได้แนบเนียน

ประเภทของ AI	*คำอธิบายสั้น ๆ*	*ตัวอย่าง*
AI แบบแคบ (Narrow AI)	เก่งเฉพาะงานใดงานหนึ่ง ไม่มีสำนึก	ระบบแนะนำหนัง Netflix, ตรวจจับสแปม
AI ทั่วไป (General AI)	ฉลาดเทียบเท่ามนุษย์ทุกด้าน	(ยังไม่มีจริง)
Super AI	ฉลาดเหนือมนุษย์	(แค่ในนิยาย)

AI ที่เราใช้ทั้งหมดวันนี้คือ Narrow AI แม้จะดูเก่งกาจแค่ไหน มันก็คือระบบทางสถิติที่ถูกออกแบบมาเพื่อทำงานเฉพาะทาง และความ “เก่ง” นี้เองที่ทำให้เกิดความเสี่ยงรูปแบบใหม่ ๆ เพราะเรามักไว้ใจมากเกินไป

1.3 ทำไม AI ถึงต้องมีความปลอดภัยเฉพาะของมัน

คำถามสำคัญ: เรามี Cybersecurity (ความมั่นคงปลอดภัยไซเบอร์) อยู่แล้ว ทำไมต้องมี AI Security ต่างหาก?

คำตอบสั้น ๆ คือ เพราะ AI มี **พื้นผิวการโจมตี (Attack Surface)** ที่แปลกใหม่ และ **ความเสียหาย** ที่ขยายวงกว้างกว่า

- **โมเดล AI เป็น “กล่อง” ที่คาดเดายาก** ซอฟต์แวร์แบบดั้งเดิมทำงานตามกฎที่มนุษย์เขียน ทุกเงื่อนไขตรวจสอบได้ แต่โมเดล ML เรียนรู้จากข้อมูล ทำให้พฤติกรรมของมันอาจผิดพลาดหรือถูกชักจูงได้โดยที่เราไม่สามารถอ่านโค้ดเพื่อหาเหตุผลได้เหมือนเดิม

- **ข้อมูลคือหัวใจ** AI ถูกเทรนจากข้อมูลมหาศาล ซึ่งอาจมีข้อมูลส่วนบุคคล ความลับทางการค้า หรืออคติแฝงฝังอยู่ การโจมตีจึงไม่ได้หยุดที่ตัวระบบ แต่ลามมาถึงการขโมยหรือปนเปื้อนข้อมูล

- **Input แปลก ๆ เปลี่ยนผลลัพธ์ได้** ในระบบทั่วไป ถ้ามีคนพิมพ์คำสั่งผิดพลาด โปรแกรมอาจแจ้ง error แต่ AI อาจถูกหลอกให้ตอบคำถามอันตราย หรือเปิดเผยข้อมูลภายในด้วย **Prompt Injection** ที่ดูเหมือนประโยคสนทนาธรรมดา

- **AI ถูกใช้เป็นเครื่องมือโจมตีซะเอง** แฮกเกอร์ใช้ AI สร้างฟิชชิ่งอีเมลที่แนบเนียน ปลอมเสียงในวิดีโอคอล (Deepfake) หรือเขียนมัลแวร์อัตโนมัติ ทำให้การโจมตีเก่งขึ้น เร็วขึ้น และตรวจจับยากขึ้น

AI Security จึงเป็นสหวิทยาการที่ผสานระหว่าง Cybersecurity, Data Privacy, Machine Learning และจริยธรรม โดยมีเป้าหมายให้ระบบ AI มีคุณสมบัติ **CIA Triad** เช่นเดียวกับไอทีทั่วไป แต่ขยายความครอบคลุมไปถึง *ความปลอดภัยของโมเดล* และ *ความเป็นส่วนตัวของข้อมูล* ด้วย:

- **Confidentiality (ความลับ):** ป้องกันไม่ให้ข้อมูลส่วนตัวรั่วไหลผ่านการตอบของโมเดล หรือถูกขโมยด้วยเทคนิค Model Inversion
- **Integrity (ความสมบูรณ์):** รับรองว่าโมเดลไม่ถูกวางยาพิษข้อมูล (Data Poisoning) จนตัดสินใจผิดพลาด หรือถูกบิดเบือนด้วย Adversarial Examples
- **Availability (ความพร้อมใช้):** ทำให้ AI ยังให้บริการได้แม้ถูกโจมตีด้วยการป้อนคำถามมหาศาล (Denial of Service)
- **(เพิ่มเติม) Safety & Fairness:** AI ต้องไม่สร้างเนื้อหาอันตราย ให้คำแนะนำที่ผิดพลาดถึงชีวิต หรือตัดสินใจอย่างมีอคติ

1.4 ภัยไซเบอร์ยุค AI ไม่เหมือนเดิมอีกต่อไป — ฉากทัศน์ที่ทุกองค์กรต้องเตรียมรับมือ

เพื่อให้เห็นภาพชัดขึ้น ลองดูตัวอย่างเหตุการณ์ที่เกิดขึ้นจริงในช่วง 2–3 ปีที่ผ่านมา ซึ่งสะท้อนว่า AI ไม่ใช่แค่ซอฟต์แวร์ธรรมดา:

เหตุการณ์ที่ 1: Samsung แบน ChatGPT หลังข้อมูลรั่ว (2566) พนักงาน Samsung ป้อนซอร์สโค้ดลับและบันทึกการประชุมภายในลง ChatGPT เพื่อให้ช่วยสรุปและวิเคราะห์ ข้อมูลเหล่านั้นถูกส่งไปยังเซิร์ฟเวอร์ของ OpenAI และอาจถูกใช้เทรนโมเดลต่อได้ ทำให้บริษัทต้องออกกฎห้ามใช้ Generative AI ในอุปกรณ์ของบริษัทและกำลังสร้าง

AI ภายในของตนเอง → *สิ่งที่ควรรู้:* ข้อมูลที่คุณพิมพ์ลงในแชตบอตสาธารณะอาจไม่เป็นความลับอีกต่อไป แม้คุณจะลบประวัติแล้วข้อมูลก็อาจถูกเก็บไว้ในโครงสร้างของโมเดล (ในบทที่ 4 เราจะลงลึกเรื่องนี้)

เหตุการณ์ที่ 2: Deepfake CEO หลอกโอนเงิน 25 ล้านดอลลาร์ (2567) พนักงานบริษัทข้ามชาติในฮ่องกงถูกหลอกให้โอนเงินกว่า 25 ล้านดอลลาร์ หลังจากเข้าร่วมวิดีโอคอลที่มีทั้ง CFO และเพื่อนร่วมงานที่ดูเหมือนจริงทุกประการ แท้จริงแล้วทุกคนในสายคือ Deepfake ที่สร้างด้วย AI ทั้งสิ้น → *สิ่งที่ควรรู้:* AI สามารถเลียนแบบบุคคลได้แนบเนียนขนาดแยกไม่ออกด้วยตาเปล่า แม้แต่การยืนยันตัวตนแบบเห็นหน้าก็ไม่ปลอดภัยอีกต่อไป (อ่านเพิ่มบทที่ 5)

เหตุการณ์ที่ 3: แชตบอตรับบทบาทใหม่โดยไม่ตั้งใจ (2566) นักวิจัยพบว่าการเติมประโยคสั้น ๆ อย่าง “Ignore previous directions…” สามารถเปลี่ยนพฤติกรรมของ LLM ให้ลืมคำสั่งความปลอดภัยที่นักพัฒนาตั้งไว้ และทำตามคำขอใด ๆ ก็ตาม ซึ่งต่อมาชุมชนเรียกว่า Prompt Injection → *สิ่งที่ควรรู้:* โมเดลภาษาไม่มีเส้นแบ่งระหว่าง “คำสั่งระบบ” กับ “ข้อมูลผู้ใช้” อย่างชัดเจน นี่คือรากฐานของช่องโหว่ที่ใหญ่ที่สุดของ LLM (บทที่ 3 ทั้งบทจะว่าด้วยเรื่องนี้)

จากตัวอย่างทั้งสาม จะเห็นว่า **ภัย AI ไม่ใช่แค่ไวรัสคอมพิวเตอร์** แต่มันคือภัยที่โจมตี *วิธีคิดของมนุษย์* และ *จุดบอดของปัญญาประดิษฐ์* ไปพร้อมกัน

1.5 ทำไมต้องเรียน AI Security – ถึงเวลาของทุกคน ไม่ใช่แค่ผู้เชี่ยวชาญ

หากคุณเป็นนักศึกษาไอทีหรือไซเบอร์ซิเคียวริตี้ Cybersecurity คุณอาจกำลังคิดว่านี่คือทักษะเฉพาะทางสำหรับคนไม่กี่คน แต่ความเป็นจริงคือ **ในอีก 2–3 ปีข้างหน้า AI Security จะเป็นทักษะพื้นฐาน** เช่นเดียวกับที่ Firewall และ Antivirus เคยเป็น เพราะ:

- **ตลาดงานโตก้าวกระโดด** Gartner ประมาณการว่าภายในปี 2026 พ.ศ. 2569 องค์กรกว่า 80% จะนำ AI มาใช้ในกระบวนการหลัก และความต้องการผู้รู้ด้าน AI Trust, Risk and Security Management (AI TRiSM) จะเพิ่มขึ้นอย่างมหาศาล

- **กฎหมายเริ่มบังคับใช้จริง** สหภาพยุโรปออก EU AI Act กำหนดระดับความเสี่ยงและบทลงโทษสำหรับระบบ AI รวมถึงความปลอดภัยที่ต้องมี ประเทศอื่น ๆ รวมถึงไทยกำลังยกร่างกฎหมาย AI ทำให้องค์กรต้องปรับตัวอย่างเลี่ยงไม่ได้ (อ่านบทที่ 8)

- **เป็นภูมิคุ้มกันส่วนบุคคล** แม้ว่าคุณจะไม่ทำงานสาย IT เลย แต่คุณก็เป็นผู้ใช้ AI ทุกวัน การรู้วิธีปกป้องตัวเองจากการหลอกลวงด้วย Deepfake, Phishing ยุคใหม่ และการตั้งค่า Privacy ให้ถูกต้อง คือทักษะชีวิตในโลกดิจิทัล 2.0

ส่วนใครที่เป็นอาจารย์หรือผู้สอน หนังสือเล่มนี้ถูกออกแบบให้ใช้เป็นตำราเสริมวิชาเลือกหรือคอร์สอบรมระยะสั้นได้ทันที แต่ละบทจบในตัวเอง มีกรณีศึกษา แบบฝึกหัดท้ายบท และ checklist ให้ลงมือปฏิบัติจริง (ดูบทที่ 9)

1.6 หนังสือเล่มนี้จะพาคุณไปรู้จักอะไรบ้าง

หลังจากปูพื้นฐานในบทที่ 1 นี้แล้ว บทต่อจากนี้จะพาเจาะลึกภัยคุกคามและการป้องกันเป็นขั้น ๆ อ่านตั้งแต่บทที่ 2 ถึง 8 จะทำให้คุณเข้าใจภาพรวมของ AI Security อย่างเป็นระบบ และสามารถประยุกต์ใช้ตามระดับความสนใจ:

- บทที่ 2 : ภูมิทัศน์ภัยคุกคามไซเบอร์ยุค AI — ใครคือผู้โจมตี และเขาใช้ AI เป็นอาวุธอย่างไร
- บทที่ 3 : Prompt Injection และการหลอกระบบ AI — จุดอ่อนที่อันตรายที่สุดของ LLM

- บทที่ 4 : ข้อมูลรั่วไหลจากการใช้ AI — เมื่อสิ่งที่คุณถามกลายเป็นสมบัติสาธารณะ
- บทที่ 5 : AI-Generated Phishing และ Deepfake — การโจมตีที่วิศวกรรมสังคมมนุษย์ด้วย AI
- บทที่ 6 : ความเสี่ยงของโมเดล Machine Learning — Adversarial Attack, Data Poisoning, Model Stealing
- บทที่ 7 : การใช้งาน ChatGPT/LLM อย่างปลอดภัย — คู่มือสำหรับทั้งบุคคลและองค์กร
- บทที่ 8 : AI Governance และ Risk Management — กฎหมาย กรอบมาตรฐาน และการสร้างนโยบาย
- บทที่ 9 : Checklist สำหรับองค์กรและนักศึกษา — นำไปใช้ได้จริงทันที
- บทที่ 10 : กรณีศึกษาและแนวทางต่อยอดงานวิจัย — สรุปภาพรวมและเส้นทางในสายอาชีพ

แต่ละบทเขียนให้อ่านง่าย ใช้ภาษาไทยเป็นหลัก แต่ยังคงศัพท์เทคนิคภาษาอังกฤษกำกับไว้เพื่อประโยชน์ในการค้นคว้าต่อ ยกตัวอย่างประกอบจากเหตุการณ์จริง และปิดท้ายด้วย “สิ่งที่คุณจะได้เรียนรู้” และ “คำถามทบทวน”

1.7 เริ่มต้นเดินทางสู่โลกของ AI Security

คุณไม่จำเป็นต้องเป็นนักเขียนโค้ดหรือนักคณิตศาสตร์ก็เข้าใจ AI Security ได้ เพราะหัวใจสำคัญของมันคือ **การตั้งคำถามที่ถูกต้อง**

- “ไอ้ที่ AI ตอบมานี่ มันจริงหรือเปล่า”
- “ข้อมูลของฉันจะไปอยู่ที่ไหน”
- “ทำไมจู่ ๆ แชตบอตถึงยอมบอกวิธีแฮกเว็บไซต์”

หนังสือเล่มนี้ตั้งใจเชิญชวนให้คุณเป็น “ผู้ใช้ AI ที่รู้เท่าทัน” และในเวลาเดียวกันก็เปิดประตูสู่การเป็นผู้ปกป้องระบบ AI ได้ในอนาคต ไม่ว่าคุณจะอยู่ในบทบาทใด — นักศึกษา, อาจารย์, พนักงานออฟฟิศ, ผู้บริหาร หรือประชาชนทั่วไป — การทำความเข้าใจ AI Security คือการติดอาวุธทางปัญญาที่จำเป็นสำหรับทศวรรษนี้

พร้อมหรือยัง? พลิกไปบทที่ 2 เพื่อสำรวจว่า **ใครคือผู้คุกคามในโลกไซเบอร์ยุค AI** และพวกเขาเปลี่ยนอาวุธดิจิทัลให้ฉลาดขึ้นได้อย่างไร

เอกสารอ้างอิงประจำบท

1. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. — ตำรามาตรฐานที่อธิบายนิยามของ AI และการเรียนรู้ของเครื่องอย่างครบถ้วน
2. OpenAI. (2022, November 30). Introducing ChatGPT. https://openai.com/blog/chatgpt
3. Reuters. (2023, May 2). Samsung bans use of generative AI tools like ChatGPT after April data leak. Reuters Technology News. — เหตุการณ์จริงที่สะท้อนความเสี่ยงด้านข้อมูลขององค์กร
4. The Guardian. (2024, February 4). Deepfake CFO scams Hong Kong firm out of $25m. The Guardian.
5. Financial Times. (2024, May). Arup lost $25mn in Hong Kong deepfake video conference scam.
6. OWASP Foundation. (2023). OWASP Top 10 for LLM Applications. https://owasp.org/www-project-top-10-for-large-language-model-applications/ — กรอบช่องโหว่สำคัญของ LLM รวมถึง Prompt Injection
7. National Institute of Standards and Technology (NIST). (2023). *AI Risk* Management Framework (AI RMF 1.0). NIST AI 100-1. — กรอบบริหารความเสี่ยง AI ที่ได้รับการยอมรับในระดับสากล
8. European Commission. (2024). The EU Artificial Intelligence Act. — กฎหมายสำคัญที่กำหนดทิศทาง AI Governance โลก
9. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and Harnessing Adversarial Examples. arXiv preprint arXiv:1412.6572. — งานวิจัยต้นแบบที่เปิดประเด็น Adversarial Attack ใน Deep Learning