รายงานการเข้าร่วมโครงการเอพีโอ

25-IP-30-GE-WSP-A

Workshop on AI Management Systems

ระหว่างวันที่ 9-12 กันยายน 2568

ณ นครโฮจิมินห์ ประเทศเวียดนาม

จัดทำโดย นายธนศักดิ์ กำหนดแน่
นักพัฒนาศักยภาพและบริหารงานฝึกอบรม ฝ่ายพัฒนาศักยภาพ
วันที่ 29 กันยายน 2568

ส่วนที่ 1 เนื้อหา/องค์ความรู้จากการเข้าร่วมโครงการ

การอบรมนี้ครอบคลุมหัวข้อหลักเกี่ยวกับการพัฒนาและจัดการระบบ AI อย่างมีความรับผิดชอบ โดยเน้นมาตรฐานสากลอย่าง ISO 42001 และแนวทางปฏิบัติที่ดีในการนำ AI มาใช้ในองค์กร การอบรมแบ่งเป็น 16 Session และการศึกษาดูงาน 2 องค์กร โดยมีวิทยากรหลัก ได้แก่ Carl Fernando (จาก CANDA Consulting Hong Kong) Jane Melanie Jane Blackmore (จาก Blackmores, United Kingdom), Antonina Burlachenko (จาก Star, Poland) และ Ninh The Ninh (จาก TESV Electrical Safety Company, Vietnam) เนื้อหาเน้นการผสาน AI เข้ากับกระบวนการธุรกิจ การจัดการความเสี่ยง และการรับมือกับผลกระทบทางสังคม

วัตถุประสงค์ของรายงานนี้คือรวบรวมและเรียบเรียงเนื้อหาจากเอกสารทั้งหมด เพื่อให้ผู้อ่านเข้าใจภาพรวมและนำไปประยุกต์ใช้ได้ โดยแบ่งเป็นส่วนสรุปตามวันอบรม หัวข้อหลัก และบทสรุปพร้อมข้อเสนอแนะรายละเอียดดังนี้

Session 1: Introduction to AI and Machine Learning Technology

นิยามและการแนะนำ AI และ ML

นิยามตามมาตรฐาน ISO

- AI Agent: หน่วยงานอัตโนมัติที่รับรู้สภาพแวดล้อมและตอบสนองด้วยการกระทำเพื่อบรรลุเป้าหมาย
- Artificial Intelligence: การวิจัยและพัฒนากลไกและการประยุกต์ใช้ระบบ AI
- Artificial Intelligence System: ระบบวิศวกรรมที่สร้างผลลัพธ์ เช่น เนื้อหา การพยากรณ์ คำแนะนำ หรือการตัดสินใจ สำหรับเป้าหมายที่มนุษย์กำหนด
- General AI: ประเภทระบบ AI ที่จัดการงานหลากหลายด้วยประสิทธิภาพที่น่าพอใจ
- Narrow AI: ประเภทระบบ AI ที่มุ่งเน้นงานเฉพาะเพื่อแก้ปัญหาเฉพาะ
- Performance: ผลลัพธ์ที่วัดได้ (เชิงปริมาณหรือเชิงคุณภาพ)

นิยามพื้นฐาน

- Artificial Intelligence (AI): วิศวกรรมในการสร้างเครื่องจักรและโปรแกรมที่ชาญฉลาด
- Machine Learning (ML): ประเภทของ AI ที่ช่วยให้คอมพิวเตอร์เรียนรู้จากข้อมูลและปรับปรุงประสิทธิภาพโดยไม่ต้องโปรแกรมอย่างชัดเจน
- Deep Learning: ส่วนย่อยของ ML ที่ใช้เครือข่ายประสาทเทียมลึกหลายชั้นในการประมวลผลข้อมูลจำนวนมากเพื่อระบุรูปแบบ ทำให้คอมพิวเตอร์สามารถตัดสินใจหรือพยากรณ์ได้

ตัวอย่าง

- AI: Deep Blue คอมพิวเตอร์ที่เอาชนะแชมป์หมากรุกโลกในปี 1997 โดยใช้ความรู้หมากรุกที่เขียนโค้ดไว้ล่วงหน้า
- ML: โมเดลจำแนกภาพโดยใช้อัลกอริทึม k-nearest neighbors (KNN)
- Deep Learning: ยานพาหนะขับเคลื่อนอัตโนมัติที่ใช้ deep learning เพื่อจัดการสถานการณ์ต่างๆ ขณะขับขี่

ความแตกต่างระหว่าง AI และซอฟต์แวร์ทั่วไป AI และซอฟต์แวร์ทั่วไปมีความแตกต่างหลักๆ ดังนี้:

ลักษณะ	AI	ซอฟต์แวร์ทั่วไป
หลักการ	เรียนรู้จากข้อมูลและปรับปรุงตามเวลา (ไม่แน่นอน)	ทำงานตามกฎและตรรกะที่กำหนดไว้ล่วงหน้า (แน่นอน)
การปรับตัว	ปรับตัวสูง สามารถจัดการข้อมูลที่ไม่มีโครงสร้าง	ปรับตัวจำกัด สำหรับงานที่มีโครงสร้างเฉพาะ
การพึ่งพาข้อมูล	พึ่งพาข้อมูลสูง	ต้องการข้อมูลน้อยสำหรับการตั้งค่าและทดสอบเบื้องต้น
ความยืดหยุ่น	สูง สามารถนำไปใช้กับงานคล้ายกัน	แข็งตัว ต้องเขียนโปรแกรมใหม่สำหรับงานใหม่
การขยายขนาด	ขยายได้ด้วยการปรับข้อมูลและโมเดล	จำกัดตามการออกแบบเบื้องต้น
การโต้ตอบกับมนุษย์	ต้องการการกำกับดูแล การฝึกใหม่ และพิจารณาด้านจริยธรรม	การโต้ตอบน้อยหลังจากนำไปใช้งานมีข้อกังวลด้านจริยธรรมน้อย
ค่าใช้จ่ายในการบำรุงรักษา	ต้องการปรับแต่ง ติดตาม และอัปเดตข้อมูลอย่างต่อเนื่อง	บำรุงรักษาน้อยหลังจากนำไปใช้งาน

สรุปความแตกต่าง

ซอฟต์แวร์ที่ขับเคลื่อนด้วย AI เหมาะสำหรับงานที่ต้องการการเรียนรู้จากข้อมูล การจดจำรูปแบบ และการปรับตัวต่อข้อมูลที่เปลี่ยนแปลง เหมาะสำหรับการวิเคราะห์ข้อมูลซับซ้อน การจดจำรูปแบบ การปรับแต่งเฉพาะบุคคล ระบบอัตโนมัติงานที่ไม่แน่นอน และการประมวลผลภาษาธรรมชาติ (NLP) แต่ต้องการข้อมูลจำนวนมาก ทักษะเฉพาะทาง และการบำรุงรักษาต่อเนื่อง

ซอฟต์แวร์ทั่วไปเหมาะสำหรับงานที่มีกฎชัดเจนและข้อมูลที่มีโครงสร้าง เหมาะสำหรับการประมวลผลตามกฎ การจัดการข้อมูลง่าย การใช้งานตามขั้นตอนคงที่ สภาพแวดล้อมที่มีความแปรปรวนต่ำ และการตัดสินใจแบบคงที่ แต่ขาดความยืดหยุ่นโดยไม่ต้องปรับโปรแกรมใหม่

ในทางตรงกันข้าม ซอฟต์แวร์ทั่วไปต้องใช้รูปแบบที่กำหนดด้วยมือเพื่อคำนวณผลลัพธ์ ในขณะที่ ML เรียนรู้รูปแบบจากข้อมูลเข้าและผลลัพธ์

ประเภทของ AI

ระบบ AI มีความหลากหลายในความสามารถและความซับซ้อน โดยขึ้นอยู่กับระดับความเป็นอิสระ การเรียนรู้จากข้อมูล และการปรับตัวต่อสถานการณ์ใหม่ ที่ระดับพื้นฐาน บางระบบถูกออกแบบสำหรับงานเฉพาะตามกฎหรือรูปแบบข้อมูลแคบ ทำให้มีประสิทธิภาพสูงสำหรับการใช้งานเฉพาะแต่จำกัดนอกเหนือจากนั้น

ประเภทหลัก

- **Machine Learning:** การให้ตัวอย่างข้อมูลจำนวนมากแก่คอมพิวเตอร์เพื่อให้ระบุรูปแบบและใช้ในการพยากรณ์หรือตัดสินใจ เช่น การวิเคราะห์ภาพผีเสื้อและรถยนต์เพื่อจำแนกวัตถุใหม่ ยิ่งข้อมูลมาก ยิ่งมีประสิทธิภาพสูง
- **Natural Language Processing (NLP):** การเข้าใจและสร้างภาษามนุษย์ทั้งแบบข้อความและเสียง ทำให้คอมพิวเตอร์สามารถวิเคราะห์แนวคิด อารมณ์ และความสัมพันธ์เพื่อดึงข้อมูลเชิงลึก และสร้างข้อความหรือเสียงเพื่อโต้ตอบกับผู้ใช้ ตัวอย่าง: การเอกสารทางคลินิกเพื่อดึงข้อมูลจากบันทึกแพทย์ หรือรายงานรังสีวิทยาเพื่อตรวจหาความผิดปกติ
- **Computer Vision:** ความสามารถของระบบในการจับภาพ ประมวลผล และตีความข้อมูลภาพหรือวิดีโอ เชื่อมโยงกับการจดจำภาพ โดยภาพถูกแทนด้วยเมทริกซ์ตัวเลขและอาจมีข้อมูลเมตา ตัวอย่าง: การตรวจหาเนื้องอกในภาพทางการแพทย์ การตรวจหาเครื่องหมายโรคในสไลด์พยาธิวิทยา การติดตามขั้นตอนผ่าตัดแบบเรียลไทม์ การติดตามการเคลื่อนไหวผู้ป่วยเพื่อป้องกันการล้ม และการระบุตัวผู้ป่วยด้วยการจดจำใบหน้า
- **Internet of Things (IoT):** เครือข่ายอุปกรณ์และระบบที่เชื่อมต่อกันเพื่อการโต้ตอบกับโลกกายภาพและเสมือนจริง เซ็นเซอร์รวบรวมข้อมูลจากสิ่งกายภาพและส่งต่อ ขณะที่แอคชูเอเตอร์ตอบสนองต่อข้อมูลดิจิทัล AI ช่วยวิเคราะห์ข้อมูลเพื่อการตัดสินใจแบบเรียลไทม์และพยากรณ์ผลลัพธ์ ตัวอย่าง: ระบบตรวจสอบทางการแพทย์ที่แจ้งเตือนบุคลากร หรือระบบตรวจสอบสภาพอากาศที่ให้การพยากรณ์
- **Generative AI:** ส่วนย่อยของ AI ที่มุ่งสร้างเนื้อหาใหม่ เช่น ภาพ ข้อความ ดนตรี หรือข้อมูลอื่นๆ โดยเรียนรู้รูปแบบจากชุดข้อมูลที่มีอยู่ ใช้เทคนิค deep learning และเครือข่ายประสาทเพื่อสร้างผลลัพธ์ดั้งเดิมที่คล้ายข้อมูลฝึก ตัวอย่าง: Text-to-text, Text-to-image, Text-to-video, Text-to-code, Text-to-speech, Image and text-to-image
- **Robotics:** มุ่งสร้างเครื่องจักรที่ทำงานอัตโนมัติหรือกึ่งอัตโนมัติ โดยใช้เซ็นเซอร์ ML และบางครั้ง computer vision เพื่อเข้าใจและการโต้ตอบกับสภาพแวดล้อม ตัวอย่าง: หุ่นยนต์ช่วยผ่าตัดที่ช่วยศัลยแพทย์ในการผ่าตัดซับซ้อนด้วยความแม่นยำสูง ลดความรุกราน

วงจรชีวิตของระบบ AI (ตาม ISO 22989)

วงจรชีวิตของระบบ AI ประกอบด้วยขั้นตอนหลักดังนี้ พร้อมการจัดการ DevOps ความโปร่งใสและอธิบายได้ ความมั่นคงปลอดภัยและความเป็นส่วนตัว การจัดการความเสี่ยง และการกำกับดูแล

- **Inception:** กำหนดเป้าหมาย AI (ปัญหาอะไร เหตุใดจึงพัฒนา ตัวชี้วัดความสำเร็จ) กำหนดข้อกำหนด การจัดการความเสี่ยง การติดตามและรับผิดชอบ ทรัพยากร และความเป็นไปได้ (proof-of-concept)
- **Design and Development:** กำหนดแนวทางโดยรวม สถาปัตยกรรม โค้ด ข้อมูลฝึก การรักษาความเสี่ยง
- **Verification and Validation:** คุณภาพข้อมูล การตรวจสอบระบบ การยอมรับ การติดตามความเสี่ยง Verification ยืนยันว่าระบบสร้างถูกต้อง Validation ยืนยันว่าตรงตามการใช้งานที่ตั้งใจ
- **Deployment:** ติดตั้งระบบในสภาพแวดล้อมเป้าหมาย การรักษาความเสี่ยง การคำนวณแบบ cloud หรือ edge
- **Operation and Monitoring:** ติดตามการทำงานปกติและผิดปกติ การซ่อมแซม การอัปเดต การสนับสนุน การตรวจสอบต่อเนื่อง การติดตามความเสี่ยง
- **Re-evaluate:** ประเมินผลลัพธ์การทำงาน ปรับเป้าหมายและข้อกำหนด การติดตามความเสี่ยง (ทริกเกอร์: การเปลี่ยนแปลงผลิตภัณฑ์ สถานะของเทคโนโลยี ปัญหาประสิทธิภาพ Concept drift ความกังวลด้านความมั่นคง)
- **Retirement:** ยกเลิกและทิ้งระบบ หรือแทนที่หากยังเกี่ยวข้องแต่มีแนวทางที่ดีกว่า

Session 2: Introduction and Overview of ISO 42001 (AIMS)

ปัญญาประดิษฐ์ (AI) คืออะไร

เป็นสาขาของวิทยาการคอมพิวเตอร์ที่มุ่งสร้างตัวแทนอัจฉริยะ ซึ่งเป็นระบบที่สามารถคิดเหตุเรียนรู้ และกระทำอย่างอิสระ โดยพื้นฐานแล้ว AI มุ่งจำลองความฉลาดของมนุษย์ในเครื่องจักร

- ระบบ AI สามารถทำงานที่ปกติต้องใช้ความฉลาดของมนุษย์
- สามารถคิดอย่างมีตรรกะและตัดสินใจจากข้อมูลที่มี
- สามารถปรับปรุงการทำงานได้ตามเวลาโดยเรียนรู้จากประสบการณ์
- สามารถทำงานอย่างอิสระโดยไม่ต้องได้รับคำสั่งจากมนุษย์อย่างต่อเนื่อง

แม้ว่าความก้าวหน้าของ AI อาจทำให้รู้สึกกังวล แต่การใช้ AI อย่างรับผิดชอบนั้นพบได้ทั่วไปในชีวิตประจำวันแล้ว

การใช้ AI ในชีวิตส่วนตัว

- อุปกรณ์บ้านอัจฉริยะ: ควบคุมไฟ เทอร์โมสตัท และระบบรักษาความปลอดภัย
- ผู้ช่วยเสมือน: เช่น Siri, Alexa และ Google Assistant ที่ตั้งนาฬิกาปลุก เล่นเพลง หรือตอบคำถาม
- ระบบแนะนำ: แพลตฟอร์มอย่าง Netflix, Spotify และ Amazon แนะนำผลิตภัณฑ์ ภาพยนตร์ หรือเพลงตามความชอบ
- รถขับเคลื่อนอัตโนมัติ: ใช้ AI เพื่อรับรู้สภาพแวดล้อม ตัดสินใจ และนำทางอย่างปลอดภัย

การใช้ AI ในที่ทำงาน

- การเงิน: ใช้สำหรับการซื้อขายอัลกอริทึม การประเมินความเสี่ยง และตรวจจับการฉ้อโกงแบบเรียลไทม์
- การพัฒนาผลิตภัณฑ์: ช่วยในการออกแบบ ทดสอบ และนวัตกรรม
- การบริการลูกค้า: แชทบอทและผู้ช่วยเสมือนให้การสนับสนุน 24/7 และคำแนะนำเฉพาะบุคคล
- การตลาดและการขาย: เครื่องมือ AI ช่วยโฆษณาเป้าหมาย การแบ่งกลุ่มลูกค้า และการวิเคราะห์เชิงพยากรณ์
- การผลิต: หุ่นยนต์ AI ทำให้งานอัตโนมัติและเพิ่มประสิทธิภาพ
- การดูแลสุขภาพ: ใช้สำหรับวิเคราะห์ภาพทางการแพทย์ การค้นพบยา และแผนการรักษาเฉพาะบุคคล

หลักการของ AI ที่รับผิดชอบ

- ความยุติธรรมและไม่เลือกปฏิบัติ (Fairness & Non-Discrimination)
- ความโปร่งใสและสามารถอธิบายได้ (Transparency & Explainability)
- ความเป็นส่วนตัวและความปลอดภัย (Privacy & Security)

- ความรับผิดชอบและการกำกับดูแลโดยมนุษย์ (Accountability & Human Oversight)
- ความปลอดภัยและความมั่นคง (Safety & Security)
- ด้านสิ่งแวดล้อมและสังคม (Environmental & Social)

เหตุใดจึงสร้าง ISO 42001

การกำกับดูแล AI

การพัฒนา AI ที่รวดเร็วทำให้ผู้กำหนดนโยบาย ผู้กำกับดูแล และหน่วยงานกำกับดูแลอื่นๆ ยากที่จะตามทันและสร้างกรอบนโยบายที่เหมาะสม OECD (Organization for Economic Co-operation and Development) เป็นตัวอย่างของเวทีนโยบายโลกที่ทำงานกับกว่า 100 ประเทศ เพื่อส่งเสริมหลักการ AI ร่วมกัน (นำมาใช้ครั้งแรกในปี 2019 และอัปเดตในเดือนพฤษภาคม 2024) หลักการเหล่านี้แนะนำผู้เกี่ยวข้องกับ AI ในการพัฒนา AI ที่น่าเชื่อถือ และให้คำแนะนำแก่รัฐบาลในการกำหนดนโยบาย AI ที่มีประสิทธิภาพ

เหตุผลในการสร้าง ISO 42001

- เพื่อจัดการกับการเติบโตอย่างรวดเร็วของ AI และความเสี่ยงที่เกี่ยวข้อง
- ให้คำแนะนำแก่องค์กรในการบูรณาการการควบคุมระบบจัดการ AI
- สร้างความเชื่อมั่นและความมั่นใจของสาธารณชนในเทคโนโลยี AI
- ส่งเสริมการพัฒนาและการใช้ AI ที่น่าเชื่อถือ มีจริยธรรม และรับผิดชอบ

การสอดคล้องระหว่าง OECD และ ISO 42001

- การใช้งานที่เป็นประโยชน์ (Beneficial Use)
- สิทธิ์มนุษย์และการกำกับดูแล (Human Agency and Oversight)
- ความโปร่งใสและสามารถอธิบายได้ (Transparency and Explainability)
- ความแข็งแกร่งและความปลอดภัย (Robustness and Safety)
- ความยุติธรรมและไม่เลือกปฏิบัติ (Fairness and Non-Discrimination)
- ความรับผิดชอบ (Accountability)

ความเสี่ยงจากการใช้ AI

- ข้อมูลที่ไม่น่าเชื่อถือ (Unreliable Information): AI อาจให้ข้อมูลผิดพลาด
- อคติใน AI (AI Bias): อาจนำไปสู่การเลือกปฏิบัติ
- ความอ่อนไหวต่อเวลา (Time Sensitive): ข้อมูลอาจล้าสมัย
- การลอกเลียนแบบ (Plagiarism): AI อาจคัดลอกเนื้อหาโดยไม่ให้เครดิต

ผลกระทบจากการใช้ AI อย่างไม่เหมาะสม

ตัวอย่างความเสี่ยง AI ในทางปฏิบัติ

- McDonald's: ยกเลิกโครงการ AI สำหรับไดร์ฟทรูหลังจากเปิดตัวไม่นาน เนื่องจากลูกค้าเจอปัญหาในการสั่งอาหาร ส่งผลกระทบต่อชื่อเสียง
- นิตยสาร: ในฉบับเดือนพฤษภาคม 2025 มีส่วนพิเศษแนะนำหนังสือที่ไม่มีอยู่จริง ส่งผลเสียต่อความน่าเชื่อถือ
- Microsoft-powered chatbot MyCity: ในเดือนมีนาคม 2024 ให้ข้อมูลผิดพลาดแก่ผู้ประกอบการ ซึ่งอาจนำไปสู่การละเมิดกฎหมาย
- Grok (X's AI): กล่าวหานักบาสเกตบอล Klay Thompson อย่างผิดพลาด โดยอาจเกิดจากความเข้าใจผิดในศัพท์สแลง
- Air Canada: แชทบอทให้คำแนะนำผิดพลาดแก่ลูกค้า นำไปสู่การฟ้องร้อง
- CEO อีคอมเมิร์ซ: ไล่พนักงานสนับสนุน 90% และแทนที่ด้วยแชทบอท AI
- สถานีรถไฟในสหราชอาณาจักร: ทดสอบเทคโนโลยีเฝ้าระวัง AI ใน 8 สถานี แต่เกิดความกังวลเรื่องความโปร่งใส

- iTutor Group: จ่ายค่าปรับ 365,000 ดอลลาร์ เนื่องจากซอฟต์แวร์ AI ปฏิเสธผู้สมัครตามอายุและเพศ

ISO 42001 คืออะไร

ISO 42001 เป็นมาตรฐานสากลแรกสำหรับระบบจัดการปัญญาประดิษฐ์ (Artificial Intelligence Management Systems) โดยเน้นการบูรณาการระบบ AI เข้ากับโครงสร้างองค์กรที่มีอยู่ รวมถึงระบบจัดการ ISO ที่ได้รับการรับรองแล้ว เผยแพร่ในเดือนธันวาคม 2023 โดยองค์การระหว่างประเทศว่าด้วยการมาตรฐาน (ISO) และคณะกรรมการระหว่างประเทศว่าด้วยมาตรฐานทางไฟฟ้า (IEC) ข้อกำหนดเฉพาะเกี่ยวกับ AI และการเรียนรู้ของเครื่องจักรส่วนใหญ่อยู่ในภาคผนวกมากกว่าในเนื้อหาหลัก

ข้อกำหนดหลักของ ISO 42001

ISO 42001 ใช้โครงสร้างระดับสูง (High Level Structure: HLS) ที่คุ้นเคย ข้อกำหนดหลัก ได้แก่:

- ข้อ 1-3: ข้ออธิบาย
- ข้อ 4: บริบทขององค์กร
- ข้อ 5: ความเป็นผู้นำ
- ข้อ 6: การวางแผน
- ข้อ 7: การสนับสนุน
- ข้อ 8: การดำเนินงาน
- ข้อ 9: การประเมินประสิทธิภาพ
- ข้อ 10: การปรับปรุง

รายละเอียดข้อกำหนด

- ข้อ 4: บริบทขององค์กร - เข้าใจองค์กรและบริบท ความต้องการของผู้มีส่วนได้เสีย กำหนดขอบเขตระบบจัดการ AI และระบบจัดการ AI
- ข้อ 5: ความเป็นผู้นำ - ความเป็นผู้นำและความมุ่งมั่น นโยบาย AI บทบาท ความรับผิดชอบ และอำนาจ
- ข้อ 6: การวางแผน - การจัดการความเสี่ยงและโอกาส การประเมินความเสี่ยง AI การรักษาความเสี่ยง AI การประเมินผลกระทบระบบ AI วัตถุประสงค์ AI และการวางแผน การวางแผนการเปลี่ยนแปลง
- ข้อ 7: การสนับสนุน - ทรัพยากร ความสามารถ ความตระหนัก การสื่อสาร ข้อมูลเอกสาร
- ข้อ 8: การดำเนินงาน - การวางแผนและควบคุมการดำเนินงาน การประเมินความเสี่ยง AI การรักษาความเสี่ยง AI การประเมินผลกระทบระบบ AI
- ข้อ 9: การประเมินประสิทธิภาพ - การติดตาม การวัด การวิเคราะห์ และการประเมิน การตรวจสอบภายใน การทบทวนโดยฝ่ายบริหาร
- ข้อ 10: การปรับปรุง - การปรับปรุงอย่างต่อเนื่อง ความไม่สอดคล้องและการแก้ไข

ภาคผนวกของ ISO 42001

- ภาคผนวก A: วัตถุประสงค์ควบคุมและการควบคุมอ้างอิง
- ภาคผนวก B: คำแนะนำการนำการควบคุม AI ไปใช้
- ภาคผนวก C: วัตถุประสงค์องค์กรที่เกี่ยวข้องกับ AI และแหล่งความเสี่ยง
- ภาคผนวก D: การใช้ระบบจัดการ AI ในโดเมนหรือภาคส่วนต่างๆ

วัตถุประสงค์การควบคุมในภาคผนวก A

- A.2: นโยบายที่เกี่ยวข้องกับ AI
- A.3: องค์กรภายใน
- A.4: ทรัพยากรสำหรับระบบ AI
- A.5: การประเมินผลกระทบของระบบ AI

- A.6: วงจรชีวิตระบบ AI
- A.7: ข้อมูลสำหรับระบบ AI
- A.8: ข้อมูลสำหรับผู้มีส่วนได้เสียของระบบ AI
- A.9: การใช้ระบบ AI
- A.10: ความสัมพันธ์กับบุคคลที่สามและลูกค้า

ประโยชน์ของการนำ ISO 42001 ไปใช้

ประโยชน์ที่จับต้องได้

- เพิ่มความเชื่อมั่นและชื่อเสียง (Enhanced Trust and Reputation)
- ลดความเสี่ยง (Risk Mitigation)
- ความได้เปรียบทางการแข่งขัน (Competitive Advantage)
- ประสิทธิภาพการดำเนินงาน (Operational Efficiency)
- การปฏิบัติตามกฎระเบียบ (Compliance with Regulations)
- การเข้าถึงตลาดใหม่ (Access to New Markets)
- การดึงดูดและรักษาบุคลากร (Talent Attraction and Retention)

ISO 42001 แสดงถึงความน่าเชื่อถือสำหรับองค์กรที่ใช้ AI กระบวนการปฏิบัติตามรวมถึง:

- การวิเคราะห์ผลกระทบ: ประเมินผลกระทบของระบบ AI ต่อบุคคล กลุ่ม และสังคม โดยพิจารณาความยุติธรรม ความโปร่งใส และความปลอดภัย
- การพัฒนาและนำนโยบาย AI ไปใช้: มุ่งเน้นองค์กรภายในและวงจรชีวิตระบบ AI
- การจัดการข้อมูล: จัดการข้อมูลที่ใช้ในระบบ AI อย่างโปร่งใสและรับผิดชอบ
- การติดตามและปรับปรุงอย่างต่อเนื่อง: ประเมินและปรับปรุงระบบ AI เพื่อให้สอดคล้องกับเป้าหมายองค์กรและมาตรฐานจริยธรรมสำหรับการรับรอง องค์กรต้องได้รับการประเมินโดยบุคคลที่สามเพื่อยืนยันความสอดคล้อง

Session 3: AI Management Systems and Governance Landscape

ภูมิทัศน์การกำกับดูแล AI

ภูมิทัศน์การกำกับดูแล AI ประกอบด้วยองค์ประกอบจากรัฐบาลและหน่วยงานอื่นๆ โดยแบ่งเป็นประเภทต่างๆ เพื่อจัดการกับความเสี่ยงและส่งเสริมการใช้งานอย่างรับผิดชอบ

จากรัฐบาล

- กฎหมาย (Law): รวมกฎหมายเฉพาะเช่น EU AI Act (2024), Local Law 144 (NY 2021), DTI Law (2025) และกฎหมายทั่วไปเกี่ยวกับการเลือกปฏิบัติ ความเป็นส่วนตัว และสัญญา เพื่อป้องกันการใช้งาน AI ที่ผิดกฎหมาย
- หลักการ AI (AI Principles): เช่น Voluntary AI Safety Standard (Australia 2024), Guidance on the Ethical Development and Use of AI (Hong Kong 2021), HK Policy Statement on Responsible AI (HK 2024), AI Ethics Principles (Australia 2019), Ethical AI Framework (HK 2023), Ethics, Transparency and Accountability Framework for Automated Decision Making (UK 2023) เพื่อกำหนดแนวทางจริยธรรม
- การจัดซื้อ (Procurement): นโยบาย กลยุทธ์ และข้อกำหนดในการจัดซื้อ AI เพื่อให้สอดคล้องกับมาตรฐาน
- มาตรฐาน (Standards): เช่น ISO/IEC 42001, NIST AI RMF, ISO 23894 เพื่อให้มีกรอบงานที่เป็นมาตรฐาน

จากหน่วยงานอื่นๆ

- หลักจริยธรรม (Ethical Principles): เช่น OECD AI Principles (2024), IEEE – Ethically Aligned Design v2 (2019), World Economic Forum – 7 Principles for Human-Centric AI (2024) เพื่อเน้นการใช้งาน AI ที่คำนึงถึงมนุษย์
- การมีส่วนร่วมในตลาด (Market Participation): เช่น App Store, Google Play, Stock Exchange, Insurers, Customers เพื่อให้ตลาดช่วยกำกับดูแลผ่านการตรวจสอบและความต้องการของผู้บริโภค

ภูมิทัศน์นี้แสดงให้เห็นถึงการผสมผสานระหว่างการบังคับและสมัครใจ เพื่อสร้างสมดุลระหว่างนวัตกรรมและความปลอดภัย

การพัฒนากฎระเบียบ

การพัฒนากฎระเบียบ AI แบ่งเป็นสามประเภทหลัก เพื่อจัดการกับความเสี่ยงที่เพิ่มขึ้น

1. **แนวทางสมัครใจ (Voluntary Guidelines)**
    - Voluntary AI Safety Standard (ออสเตรเลีย)
    - Guidance on the Ethical Development and Use of AI (ฮ่องกง)
    - Ethics, Transparency and Accountability Framework (สหราชอาณาจักร)
    - Recommendation of the Council on Artificial Intelligence (OECD)

แนวทางเหล่านี้เน้นการส่งเสริมจริยธรรมและความรับผิดชอบโดยไม่บังคับ

2. **มาตรฐาน/กรอบงาน (Standards/Frameworks)**
    - ISO/IEC 42001: AI Management Systems – กรอบสำหรับระบบจัดการ AI
    - NIST AI Risk Management Framework – กรอบจัดการความเสี่ยง AI

มาตรฐานเหล่านี้ให้เครื่องมือที่ยืดหยุ่นสำหรับองค์กรทุกขนาด

3. **การปฏิบัติตามกฎหมาย (Legal Compliance)**
    - Digital Technology Industry (DTI) Law (เวียดนาม)
    - EU AI Act (สหภาพยุโรป)
    - NYC Local Law 144, Colorado SB 21-169, California SB 1001 (สหรัฐอเมริกา)
    - Interim Measures for the Management of Generative Artificial Intelligence Services (จีน)

กฎหมายเหล่านี้บังคับใช้เพื่อป้องกันความเสี่ยงสูง เช่น การเลือกปฏิบัติและความเป็นส่วนตัว

NIST AI Risk Management Framework (RMF)

(National Institute of Standards and Technology AI Risk Management Framework)

วัตถุประสงค์

กรอบงานนี้มีเป้าหมายเพื่อลดผลกระทบเชิงลบที่อาจเกิดขึ้นในขณะที่เพิ่มผลกระทบเชิงบวกสูงสุด โดยจัดการความเสี่ยง AI มากมาย และส่งเสริมการพัฒนาและใช้งานระบบ AI ที่น่าเชื่อถือและรับผิดชอบ

ลักษณะเด่น

- สมัครใจและยืดหยุ่น: ปกป้องสิทธิ ไม่เฉพาะภาคส่วน และไม่ขึ้นกับกรณีใช้งาน ให้ความยืดหยุ่นสำหรับองค์กรทุกขนาดในภาคส่วนต่างๆ
- กลุ่มเป้าหมาย: สำหรับ "AI actors" ทั่ววงจรชีวิต AI ซึ่งครอบคลุมประสบการณ์ ความเชี่ยวชาญ และพื้นฐานที่หลากหลาย

ลักษณะความน่าเชื่อถือ 7 ประการ

กรอบงานกำหนดลักษณะความน่าเชื่อถือของ AI ดังนี้ (เสริมเนื้อหาเพื่อความสมบูรณ์โดยอ้างอิงจากบริบท NIST):

1. ความถูกต้อง (Valid and Reliable): ระบบ AI ต้องให้ผลลัพธ์ที่ถูกต้องและสม่ำเสมอ
2. ความปลอดภัย (Safe): ป้องกันอันตรายต่อมนุษย์ สังคม และสิ่งแวดล้อม

3. ความมั่นคง (Secure and Resilient): ทนต่อการโจมตีและการเปลี่ยนแปลง
4. ความโปร่งใสและสามารถอธิบายได้ (Explainable and Interpretable): ผู้ใช้เข้าใจการทำงานและเหตุผลของ AI
5. ความเป็นส่วนตัว (Privacy-Enhanced): ปกป้องข้อมูลส่วนบุคคล
6. ความยุติธรรม (Fair - with Harmful Bias Managed): ลดอคติที่เป็นอันตรายและการเลือกปฏิบัติ
7. ความรับผิดชอบ (Accountable and Transparent): มีการกำกับดูแลและรายงานที่ชัดเจน

ฟังก์ชันหลัก

กรอบงานแบ่งเป็น 4 ฟังก์ชันหลักเพื่อจัดการความเสี่ยง:

1. Govern: นโยบาย กระบวนการ และขั้นตอนที่เกี่ยวข้องกับการแมป การวัด และการจัดการความเสี่ยง AI รวมถึงโครงสร้างความรับผิดชอบ ความหลากหลาย การมุ่งมั่นทางวัฒนธรรม การมีส่วนร่วม การจัดการความเสี่ยงจากบุคคลที่สาม และนโยบายด้านความเสี่ยง
2. Map: สร้างบริบทและเข้าใจ โดยจัดประเภทระบบ AI แมปความสามารถ การใช้งาน เป้าหมาย ประโยชน์ และต้นทุน รวมถึงความเสี่ยงจากส่วนประกอบ และผลกระทบต่อบุคคล กลุ่ม สังคม และองค์กร
3. Measure: ระบุและนำวิธีการและตัวชี้วัดที่เหมาะสมมาใช้ เพื่อประเมินลักษณะที่น่าเชื่อถือ ติดตามความเสี่ยง และรวบรวมข้อเสนอแนะเกี่ยวกับประสิทธิภาพการวัด
4. Manage: จัดลำดับความสำคัญ ตอบสนอง และจัดการความเสี่ยง AI จากการประเมิน โดยวางแผนกลยุทธ์เพื่อเพิ่มประโยชน์และลดผลกระทบเชิงลบ จัดการความเสี่ยงจากบุคคลที่สาม และเอกสารแผนการรักษาความเสี่ยง

โครงสร้างฟังก์ชันหลัก

แต่ละฟังก์ชันมีหมวดย่อยเพื่อให้ครอบคลุม (เสริมรายละเอียดจากบริบท):

- Govern: รวม 6 หมวด เช่น โครงสร้างความรับผิดชอบ ความหลากหลาย วัฒนธรรม การมีส่วนร่วม การจัดการบุคคลที่สาม และนโยบาย
- Map: รวม 5 หมวด เช่น การจัดประเภท ความเข้าใจเป้าหมาย การแมปความเสี่ยง ผลกระทบต่อสังคม
- Measure: รวม 4 หมวด เช่น การประเมินลักษณะ การติดตามความเสี่ยง การรวบรวมข้อเสนอแนะ
- Manage: รวม 4 หมวด เช่น กลยุทธ์การเพิ่มประโยชน์ การจัดการบุคคลที่สาม แผนการตอบสนอง การติดตาม

คุณสมบัติอื่นๆ

- Playbook: การกระทำและคำแนะนำสำหรับแต่ละฟังก์ชันหลัก เพื่อช่วยนำไปปฏิบัติ
- RMF Profiles: โปรไฟล์สำหรับสถานะปัจจุบัน vs เป้าหมาย และกรณีใช้งานเฉพาะ เพื่อปรับแต่งกรอบงาน
- ภาคผนวก: งานของ AI actors, ความเสี่ยง AI vs ซอฟต์แวร์ทั่วไป, การโต้ตอบ AI-มนุษย์, คุณลักษณะของ RMF

Australian Voluntary AI Safety Standard

วัตถุประสงค์

ให้คำแนะนำที่ปฏิบัติได้สำหรับองค์กรในออสเตรเลีย เพื่อใช้งานและนวัตกรรม AI อย่างปลอดภัยและรับผิดชอบ

ลักษณะเด่น

- สมัครใจ: ใช้กับองค์กรตลอดห่วงโซ่อุปทาน AI
- เป้าหมาย: ช่วยองค์กรได้รับประโยชน์จาก AI ขณะลดความเสี่ยงต่อบุคคล กลุ่ม และองค์กร
- โฟกัส: เวอร์ชันแรกมุ่งเน้นผู้ใช้งาน AI (deployers) โดยเวอร์ชันถัดไปจะขยายสำหรับนักพัฒนา

การสอดคล้องระดับโลก

สอดคล้องกับกรอบงานนานาชาติ เช่น ISO 42001, NIST RMF, EU AI Act เพื่อให้องค์กรออสเตรเลียสามารถแข่งขันระดับโลก

10 Guardrails สมัครใจ

- Governance Foundation: สร้างกระบวนการรับผิดชอบ รวมการกำกับดูแล ความสามารถภายใน และกลยุทธ์การปฏิบัติตามกฎระเบียบ
- Risk Management: สร้างกระบวนการระบุและลดความเสี่ยง
- Security & Data Governance: ปกป้องระบบ AI และจัดการข้อมูลเพื่อคุณภาพและที่มา
- Test & Monitor: ทดสอบโมเดลและติดตามระบบหลังใช้งาน
- Human Oversight: เปิดใช้งานการควบคุมหรือแทรกแซงโดยมนุษย์
- Disclosure: แจ้งผู้ใช้เกี่ยวกับการตัดสินใจ AI การโต้ตอบ และเนื้อหาที่สร้างโดย AI
- Contestability: สร้างกระบวนการสำหรับผู้ได้รับผลกระทบเพื่อท้าทายการใช้งานหรือผลลัพธ์
- Transparent AI Supply Chain: โปร่งใสกับองค์กรอื่นในห่วงโซ่เพื่อจัดการความเสี่ยง
- Records: เก็บรักษาบันทึกเพื่อให้บุคคลที่สามประเมินการปฏิบัติตาม
- Stakeholder Engagement: มีส่วนร่วมกับผู้มีส่วนได้เสีย โดยเน้นความปลอดภัย ความหลากหลาย การรวม และความยุติธรรม

รายละเอียด Guardrails

Guardrail 1: Governance Foundation: กำหนดเจ้าของ AI โดยรวม พัฒนากลยุทธ์ AI และให้การฝึกอบรมเพื่อสร้างฐานสำหรับการใช้งานอย่างปลอดภัย

Guardrail 2: Risk Management: ดำเนินการประเมินผลกระทบและความเสี่ยง โดยพิจารณาลักษณะ AI และความเสี่ยงที่เพิ่มขึ้น เพื่อป้องกันอันตราย

Guardrail 3: Security & Data Governance: เอกสารที่มาและคุณภาพข้อมูล จัดการอคติ ความสมบูรณ์ และใช้หลักการความเป็นส่วนตัวออสเตรเลีย

Guardrail 4: Test & Monitor: พัฒนาแผนทดสอบ กำหนดเกณฑ์ยอมรับ และตรวจสอบระบบอย่างสม่ำเสมอ เพื่อยืนยันประสิทธิภาพ

Guardrail 5: Human Oversight: มอบหมายความรับผิดชอบ กำหนดการแทรกแซง และฝึกอบรมผู้ใช้ เพื่อลดผลกระทบที่ไม่คาดคิด

Guardrail 6: Disclosure: สื่อสารอย่างชัดเจนเกี่ยวกับการใช้งาน AI โดยใช้ภาษาที่เข้าใจง่าย และพิจารณาเทคโนโลยี เช่น watermarking

Guardrail 7: Contestability: ให้ช่องทางยกเลิกหรือท้าทาย และมอบหมายผู้รับผิดชอบ เพื่อจัดการข้อกังวล

Guardrail 8: Transparent AI Supply Chain: แบ่งปันข้อมูลเกี่ยวกับข้อมูล โมเดล และระบบ เพื่อช่วยจัดการความเสี่ยงปลายน้ำ

Guardrail 9: Records: สร้างรายการ AI และเอกสารระบบ เพื่อพิสูจน์การปฏิบัติตาม และเตรียมพร้อมสำหรับการประเมิน

Guardrail 10: Stakeholder Engagement: ระบุกลุ่มที่ได้รับผลกระทบ มีส่วนร่วมอย่างต่อเนื่อง และลดอคติที่ไม่พึงประสงค์

แนวทางและกรอบงานอื่นๆ

- HK Guidance on the Ethical Development and Use of AI: หลักการสอดคล้อง 4 องค์ประกอบหลัก: โครงสร้างกำกับดูแล การกำกับดูแลโดยมนุษย์ การดำเนินงาน การจัดการผู้มีส่วนได้เสีย พร้อมคำแนะนำเพิ่มเติมสำหรับผู้ใช้งานธุรกิจ
- OECD AI Principles: หลักการร่วมและ "ปรัชญา" 9 องค์ประกอบ เน้นประโยชน์สาธารณะและ AI ที่รับผิดชอบ
- อื่นๆ: รวมหลักการจาก IEEE และ World Economic Forum เพื่อเน้นการออกแบบที่สอดคล้องกับจริยธรรมและมนุษย์เป็นศูนย์กลาง

องค์ประกอบและธีมร่วมกัน

องค์ประกอบร่วมกันในกรอบงานต่างๆ คือ 'DARTS' Framework: Data Governance, Accountability, Risk Management, Transparency, Stakeholder Engagement โดยมุ่งเน้นหลักการ วัตถุประสงค์ และผลลัพธ์ร่วม เสาหลักของกรอบงาน

- Risk Management: ดำเนินการประเมินความเสี่ยง ประเมินผลกระทบ และใช้กลยุทธ์ลดความเสี่ยง
- Data Governance: จัดลำดับความสำคัญคุณภาพข้อมูลและการกำกับดูแล โดยสร้างที่มาข้อมูลชัดเจนและอัปเดต การควบคุมความปลอดภัย ความเป็นส่วนตัว
- Transparency: รับประกันความโปร่งใส การอธิบาย และประสิทธิภาพ โดยให้คำอธิบายชัดเจน เอกสารระบบ และทดสอบอย่างสม่ำเสมอ
- Stakeholder Engagement: ส่งเสริมการมีส่วนร่วมเชิงรุก โดยให้กลไกข้อเสนอแนะและพิจารณามุมมองหลากหลาย
- Accountability: สร้างฐานรับผิดชอบที่แข็งแกร่ง โดยกำหนดบทบาท พัฒนานโยบาย AI และลงทุนในการฝึกอบรม

ตารางเปรียบเทียบแสดงจุดร่วมกับ ISO 42001, NIST RMF, Australian Standard, HK Guidance, EU AI Act เพื่อเน้นความสอดคล้อง

Session 4: Framework for AI Systems using Machine Learning (ML)

ผู้อบรมได้ทำการสำรวจกรอบการทำงานสำหรับการพัฒนาระบบ AI โดยใช้เทคโนโลยี Machine Learning (ตามมาตรฐาน ISO 23053) ซึ่งครอบคลุมทุกขั้นตอนของ ML ตั้งแต่การเริ่มต้น, การจัดการข้อมูล, การสร้างแบบจำลอง, การตรวจสอบและยืนยัน, การนำไปใช้งาน และการปฏิบัติงาน ในแต่ละขั้นตอนเราจะเน้นที่การแลกเปลี่ยนและข้อควรพิจารณาที่สำคัญซึ่งอาจส่งผลกระทบต่อความโปร่งใส, ความเป็นธรรม หรือประสิทธิภาพของผลิตภัณฑ์สุดท้าย

Task Machine Learning (ตาม ISO 23053:2022)

งาน (task) หมายถึงการกระทำเพื่อบรรลุเป้าหมายเฉพาะ ใน ML คือการกำหนดปัญหาที่โมเดลจะแก้ไข อาจมีงาน ML หนึ่งหรือหลายงานในแอปพลิเคชัน ML การตั้งค่างานรวมการกำหนดปัญหา รูปแบบข้อมูล และคุณลักษณะ

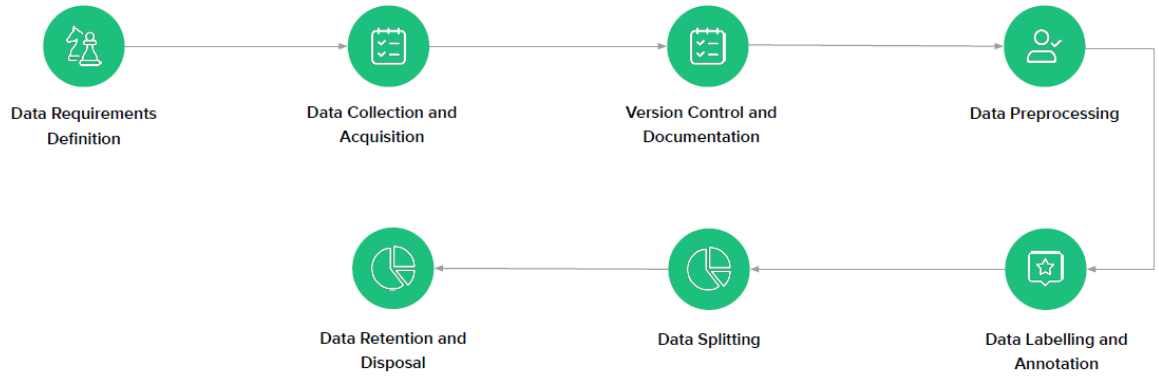
- **Regression:** พยากรณ์ค่าต่อเนื่องโดยเรียนรู้ฟังก์ชันที่เหมาะสมกับข้อมูลฝึก
- **Classification:** พยากรณ์การจัดหมวดหมู่ข้อมูลเข้าเป็นคลาส (อาจเป็น binary, multi-class หรือ multilabel)
- **Clustering:** จัดกลุ่มข้อมูลเข้าที่คล้ายกัน โดยคลาสถูกกำหนดในกระบวนการ
- **Anomaly Detection:** ระบุข้อมูลเข้าที่ไม่ตรงตามรูปแบบที่คาดหวัง โมเดลพยากรณ์ว่าข้อมูลปกติหรือไม่
- **Dimensionality Reduction:** ลดจำนวนคุณลักษณะหรือมิติขณะรักษาข้อมูลที่เป็นประโยชน์ เพื่อลดต้นทุนการคำนวณและสำรวจข้อมูล
- **Other Tasks:** เช่น การแบ่งส่วนเชิงความหมายของข้อความหรือภาพ การแปลเครื่อง การรู้จำเสียง การระบุตำแหน่งวัตถุประเภทของ ML
- **Supervised Learning:** ใช้ข้อมูลที่มีป้ายกำกับเพื่อเรียนรู้และพยากรณ์ งาน: Classification (หมวดหมู่) Regression (ค่าต่อเนื่อง)
- **Unsupervised Learning:** ใช้ข้อมูลไม่มีป้ายกำกับเพื่อค้นหารูปแบบ งาน: Clustering (จัดกลุ่ม) Association (หาความสัมพันธ์)
- **Semi-Supervised Learning:** ใช้ทั้งข้อมูลมีและไม่มีป้ายกำกับ เริ่มจากฝึกด้วยข้อมูลมีป้าย แล้วใช้ผลิต pseudo-label สำหรับข้อมูลไม่มีป้าย แล้วฝึกใหม่
- **Reinforcement Learning:** เรียนรู้ผ่านการลองผิดลองถูกเพื่อบรรลุเป้าหมาย งาน: Model-Free (เรียนจากการกระทำและรางวัลโดยตรง) Model-Based (เรียนรู้โมเดลของสภาพแวดล้อมเพื่อจำลองผลลัพธ์)

- **Self-Supervised Learning:** ฝึกบนข้อมูลไม่มีป้ายโดยใช้ป้ายโดยนัยจากข้อมูลเอง มักใช้สำหรับเรียนรู้การแทนค่า
- **Transfer Learning:** นำความรู้จากปัญหาหนึ่งไปใช้กับปัญหาอื่น เช่น fine-tuning โมเดลที่ฝึกแล้ว

วงจรชีวิตข้อมูลและการสร้างโมเดล

วงจรชีวิตข้อมูล

- การกำหนดข้อกำหนดข้อมูล
- การรวบรวมและได้มาซึ่งข้อมูล
- การควบคุมเวอร์ชันและเอกสาร
- การประมวลผลข้อมูลล่วงหน้า
- การแบ่งข้อมูล
- การติดป้ายและหมายเหตุข้อมูล
- การเก็บรักษาและกำจัดข้อมูล



การสร้างโมเดล

- Feature Engineering: เลือกและปรับคุณลักษณะ เช่น encoding การลดมิติ การเลือกคุณลักษณะ การแปลง การดึง และ scaling
- Algorithm Selection: เลือกอัลกอริทึมที่เหมาะสม อาจใช้ ensemble
- Model Training: ปรับพารามิเตอร์เพื่อให้พอดีกับข้อมูล หลีกเลี่ยง overfitting/underfitting
- Model Selection & Optimization: เลือกโมเดลที่ดีที่สุดและปรับ hyperparameters
- Model Expressiveness: ความสามารถจับความซับซ้อน หลีกเลี่ยง overfitting (เรียนรู้เสียงรบกวน) และ underfitting (ไม่เรียนรู้รูปแบบ)

การประเมิน การนำไปใช้งาน และการติดตามโมเดล

- Verification & Validation: Verification ยืนยันสร้างถูก Validation ยืนยันตรงการใช้งาน
- Deployment: บรรจุภัณฑ์ การปรับให้เหมาะสม สภาพแวดล้อมรันไทม์
- Operation: ติดตาม ซ่อมแซม อัปเดต สนับสนุน
- Monitoring: ตรวจจับ drift (data, concept, label) ความล้มเหลวของซอฟต์แวร์ ใช้ logs dashboards alerts เพื่อ observability

วิธีประเมินโมเดล

- Perturbation tests: ทดสอบกับข้อมูลที่มีเสียงรบกวน
- Invariance tests: การเปลี่ยนบางอย่างไม่ควรเปลี่ยนผลลัพธ์ (ตรวจ bias)
- Directional expectation tests: การเปลี่ยนควรเปลี่ยนผลลัพธ์ตามคาด
- Model Calibration: ปรับให้ความน่าจะเป็นตรงกับผลจริง

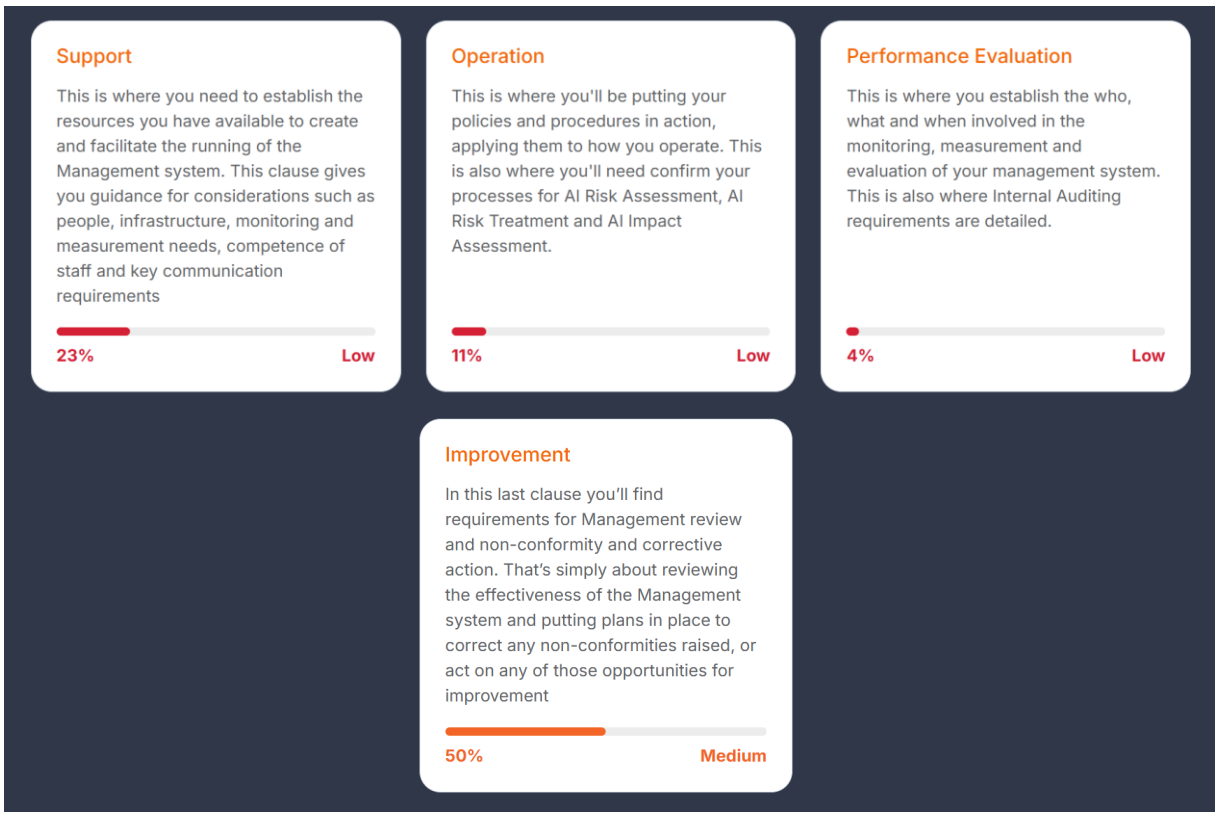
- Confidence Score: เกณฑ์ความเชื่อมั่นสำหรับแต่ละการพยากรณ์
- Slices-Based Evaluation: ประเมินบน subset เพื่อตรวจ bias (หลีกเลี่ยง Simpson's paradox)

ความท้าทายและข้อจำกัด

- ข้อมูล: คุณภาพ ความพร้อมใช้งาน ความเป็นส่วนตัว
- การสร้างโมเดล: ความซับซ้อน การตีความ การนำไปใช้ทั่วไป Overfitting vs underfitting
- การฝึกและปรับ: ค่าใช้จ่ายการคำนวณ การปรับ hyperparameters
- การนำไปใช้งาน: ความขยายขนาด การนำไปใช้งาน edge การติดตามและบำรุงรักษา
- จริยธรรมและ bias: Bias และความยุติธรรม การรับผิดชอบ การตัดสินใจเชิงจริยธรรม
- ความมั่นคง: การโจมตีแบบ adversarial การวางยาพิษข้อมูล ความเป็นส่วนตัวของโมเดล

Session 5: Gap Analysis Exercise

Workshop นี้ผู้เข้าร่วมได้ประเมินแนวปฏิบัติการจัดการ AI ในปัจจุบันขององค์กรเทียบกับข้อกำหนดของ ISO 42001 และแนวปฏิบัติที่ดีที่สุดที่ได้รับการยอมรับ ผ่านกิจกรรมที่มีคำแนะนำ ผู้เข้าร่วมจะได้ระบุจุดแข็ง จุดอ่อน และช่องว่างในการกำกับดูแล กระบวนการ และทรัพยากร แบบฝึกหัดนี้จะช่วยให้ผู้เข้าร่วมพัฒนาแผนงานที่ชัดเจนยิ่งขึ้นสำหรับการปรับโครงการริเริ่มด้าน AI ให้สอดคล้องกับมาตรฐานสากล และเสริมสร้างความพร้อมโดยรวมสำหรับการนำ AI มาใช้อย่างมีความรับผิดชอบ โดยมีตัวอย่างผลการประเมินดังนี้



Session 6: Overview and Outlook of AI in Vietnam with Global Integration Context

บทนำและขอบเขต

ภาพรวมเกี่ยวกับสถานะของ AI ในเวียดนาม โดยมุ่งเน้นการบูรณาการกับแนวโน้มระดับโลก เพื่อสนับสนุนการพัฒนาที่ยั่งยืน ขอบเขตครอบคลุมศักยภาพทางเศรษฐกิจและสังคม ความพร้อมของประเทศ การริเริ่มของรัฐบาล ระบบนิเวศ AI และอุปสรรคที่ต้องเผชิญ วัตถุประสงค์หลักคือ เพื่อให้ผู้เข้าร่วมเข้าใจว่า AI ไม่ใช่เพียงเทคโนโลยี แต่เป็นเครื่องมือสำคัญในการเร่งการเติบโตของเวียดนามให้ก้าวสู่การเป็นผู้นำในอาเซียน โดยอ้างอิงจากข้อมูลล่าสุดปี 2568 ซึ่งเวียดนามติดอันดับ 1-5 ในดัชนีความพร้อม AI ของอาเซียน (Government AI Readiness Index 2024)

ศักยภาพของ AI สำหรับเวียดนาม: ผลกระทบต่อเศรษฐกิจและสังคม และการมอง AI เป็นโอกาสในการเติบโต

AI มีศักยภาพสูงในการขับเคลื่อนเศรษฐกิจเวียดนาม ซึ่งเติบโตเฉลี่ย 6-7% ต่อปี โดยคาดว่าจะเพิ่มมูลค่า GDP ราว 12-15% ภายในปี 2573 (2030) ตามรายงานของ Google Public Policy ผลกระทบหลัก ได้แก่:

- **ด้านเศรษฐกิจ**: AI ช่วยเพิ่มประสิทธิภาพในภาคอุตสาหกรรม เช่น การผลิตอัจฉริยะ (smart manufacturing) และห่วงโซ่อุปทาน ซึ่งเวียดนามเป็นฐานการผลิตหลักของโลก (เช่น Samsung, Intel) โดยลดต้นทุนการผลิตลง 20-30% และเพิ่มผลผลิตแรงงาน
- **ด้านสังคม**: AI สนับสนุนการพัฒนาสุขภาพ การศึกษา และการเกษตร เช่น การใช้ AI ในการวินิจฉัยโรคหรือพยากรณ์ผลผลิตพืช ซึ่งช่วยลดความเหลื่อมล้ำและยกระดับคุณภาพชีวิต โดยเฉพาะในพื้นที่ชนบทที่ครอบคลุมประชากร 60%

เวียดนามมอง AI เป็นโอกาสในการเติบโต โดยปรับกลยุทธ์แห่งชาติปี 2568 (National AI Strategy) ให้เน้นการลงทุนโครงสร้างพื้นฐาน AI การพัฒนาบุคลากร และการบูรณาการกับ FTA เช่น CPTPP และ EVFTA เพื่อดึงดูดการลงทุนต่างชาติ คาดว่าตลาด AI ในเวียดนามจะมีมูลค่า 753.4 ล้านดอลลาร์สหรัฐในปี 2567 และเติบโตที่อัตรา 14.96% ต่อปีจนถึงปี 2573

ผลกระทบหลัก	ตัวอย่าง	โอกาสสำหรับเวียดนาม
เศรษฐกิจ	การผลิตอัตโนมัติ	เพิ่ม GDP 12-15% ภายใน 2573
สังคม	การแพทย์ทางไกล	ลดความเหลื่อมล้ำในชนบท
สิ่งแวดล้อม	พยากรณ์ภัยพิบัติ	เสริมความยั่งยืนตามเป้าหมาย SDG

ความพร้อมของเวียดนามสำหรับยุค AI: สิ่งที่เวียดนามได้ดำเนินการเพื่อการเปลี่ยนแปลงที่เข้มแข็งในเศรษฐกิจและสังคมด้วย AI

เวียดนามมีความพร้อมสูง โดยติดอันดับ 6 ในดัชนี AI โลก (World AI Index) และอันดับ 3 ในด้านความเชื่อมั่น AI ในอาเซียน รัฐบาลได้ริเริ่มหลายโครงการเพื่อการเปลี่ยนแปลง:

- **กลยุทธ์แห่งชาติ**: อัปเดต National AI Strategy ในปี 2568 โดยมุ่งพัฒนาโครงสร้างพื้นฐาน เช่น ศูนย์ข้อมูล AI และแพลตฟอร์มข้อมูลเปิด (open data) เพื่อสนับสนุนการวิจัยและพัฒนา
- **กฎหมายและนโยบาย**: ออกกฎหมาย AI ฉบับแรกภายในสิ้นปี 2568 (AI Law) ซึ่งครอบคลุมการกำกับดูแล การคุ้มครองข้อมูล และการใช้ AI ในภาครัฐ โดยสอดคล้องกับ EU AI Act และ OECD Principles
- **การพัฒนาบุคลากร**: โครงการฝึกอบรม AI สำหรับ 1 ล้านคนภายในปี 2570 ผ่านความร่วมมือกับ Google, Microsoft และมหาวิทยาลัยชั้นนำ เช่น VinUniversity
- **การลงทุน**: ดึงดูด FDI กว่า 1 พันล้านดอลลาร์สหรัฐใน AI โดยเน้นภาครัฐ เช่น การใช้ AI ในบริการสาธารณะ (e-government) ซึ่งช่วยประหยัดงบประมาณ 10-15%

การดำเนินการเหล่านี้ช่วยให้เวียดนามเปลี่ยนจากผู้รับเทคโนโลยีเป็นผู้พัฒนา โดยใช้ AI เพื่อแก้ปัญหาเฉพาะ เช่น การจัดการจราจรในโฮจิมินห์ซิตีหรือการตรวจสอบคุณภาพน้ำในเดลต้าแม่น้ำโขง

ระบบนิเวศ AI ของเวียดนาม: ภาพรวมใหญ่และสถิติการบูรณาการ AI เข้ากับภาคส่วนต่างๆ ของเศรษฐกิจและสังคม รวมถึงโอกาส

ระบบนิเวศ AI ในเวียดนามเติบโตอย่างรวดเร็ว โดยมีสตาร์ทอัพ AI กว่า 200 แห่ง (เพิ่ม 50% จากปีก่อน) และบริษัทใหญ่ เช่น Vingroup และ FPT ที่ลงทุนกว่า 500 ล้านดอลลาร์สหรัฐ สถิติหลัก (ปี 2568):

- **การบูรณาการตามภาคส่วน**: 40% ในอุตสาหกรรมการผลิต, 25% ในการเงิน (fintech), 20% ในสุขภาพ, และ 15% ในการเกษตร โดย AI ช่วยเพิ่มผลผลิตเกษตร 20-30%
- **ภาพรวมใหญ่**: เวียดนามมีนักวิจัย AI กว่า 10,000 คน และศูนย์ AI ระดับภูมิภาค 5 แห่ง โอกาสหลัก ได้แก่ การส่งออกบริการ AI (projected 2 พันล้านดอลลาร์สหรัฐภายใน 2573) และการร่วมมือกับสหรัฐฯ-จีนเพื่อเทคโนโลยีขั้นสูง

ภาคส่วน	สถิติการบูรณาการ	โอกาส
การผลิต	40% ของบริษัทใช้ AI	ลดต้นทุน 25%
สุขภาพ	AI วินิจฉัยโรค 80% แม่นยำ	เข้าถึงบริการในชนบท
เกษตร	พยากรณ์ผลผลิต 30%	เพิ่มรายได้เกษตรกร
การเงิน	Fintech AI 50 ล้านผู้ใช้	ลดการฉ้อโกง 40%

ระบบนิเวศนี้สนับสนุนโดย APO และ UNDP เพื่อให้เวียดนามเป็นศูนย์กลาง AI ในเอเชียตะวันออกเฉียงใต้

ความท้าทาย: อุปสรรคที่เวียดนามเผชิญในการบรรลุเป้าหมายที่เกี่ยวข้องกับ AI

แม้มีความก้าวหน้า แต่เวียดนามยังเผชิญอุปสรรคหลัก:

- **โครงสร้างพื้นฐาน**: การขาดแคลนข้อมูลคุณภาพสูงและพลังงานสำหรับศูนย์ข้อมูล AI (data centers) ซึ่งครอบคลุมเพียง 30% ของความต้องการ
- **ทักษะและบุคลากร**: ช่องว่างทักษะ AI ในแรงงาน 70% ต้องการการฝึกอบรมเพิ่มเติม โดยเฉพาะในพื้นที่ห่างไกล
- **กฎระเบียบและจริยธรรม**: การขาดกรอบกฎหมายที่ชัดเจนสำหรับ AI จริยธรรม (ethical AI) และความเสี่ยงด้านความเป็นส่วนตัวข้อมูล ตาม GDPR
- **การลงทุนและความเหลื่อมล้ำ**: การลงทุนส่วนใหญ่มุ่งเน้นไปที่เมืองใหญ่ ส่งผลให้เกิดช่องว่างดิจิทัลระหว่างเมืองและชนบทเพื่อเอาชนะ รัฐบาลวางแผนลงทุน 2 พันล้านดอลลาร์สหรัฐในโครงสร้างพื้นฐานและการศึกษา โดยความร่วมมือกับพันธมิตรระดับโลก

สรุปและแนวโน้ม

เวียดนามมีตำแหน่งที่แข็งแกร่งในการเป็นผู้นำ AI ในอาเซียน โดยใช้โอกาสจากกลยุทธ์แห่งชาติและการบูรณาการระดับโลก เพื่อการเติบโตที่ยั่งยืน แนวโน้มปี 2568-2573 คือ การขยายตลาด AI สู่ 5 พันล้านดอลลาร์สหรัฐ และการพัฒนา AI ที่รับผิดชอบเพื่อลดความเสี่ยง

Session 7: Data Management for AI

การจัดการข้อมูลสำหรับ AI (Data Management for AI)

ข้อมูลเป็นส่วนประกอบสำคัญในระบบ AI ที่ใช้ ML คุณภาพของข้อมูลจะกำหนดคุณภาพของผลิตภัณฑ์สุดท้าย รวมถึงความเกี่ยวข้องทางคลินิก ความเป็นธรรม และความปลอดภัย การจัดการข้อมูลครอบคลุมวงจรชีวิตทั้งหมด เพื่อให้แน่ใจว่าระบบ AI สามารถเรียนรู้รูปแบบจากข้อมูลเข้าและสร้างผลลัพธ์ที่ถูกต้อง

ประเภทข้อมูลตามมาตรฐาน ISO/IEC 22989:2022 และ ISO 23053:2022

ข้อมูลใน ML แบ่งเป็น 4 ประเภทหลัก โดยแต่ละประเภทมีบทบาทเฉพาะและต้องแยกจากกันอย่างชัดเจนเพื่อหลีกเลี่ยงการปนเปื้อน (data leakage) ชุดข้อมูลฝึกอบรม การตรวจสอบ และการทดสอบควรมาจากแหล่งข้อมูลเดียวกันแต่มีสถิติที่คล้ายคลึงกัน ขณะที่ข้อมูลการใช้งานจริงต้องได้รับการติดตามเพื่อตรวจจับการเปลี่ยนแปลง (drift) ที่อาจทำให้ต้องฝึกโมเดลใหม่

ประเภทข้อมูล	คำอธิบาย	การใช้งาน
Training Dataset	ชุดข้อมูลที่ใช้ฝึกพารามิเตอร์ของโมเดล	ใช้ในการเรียนรู้รูปแบบจากข้อมูล
Validation Dataset	ชุดข้อมูลที่ใช้เลือกโมเดลที่ดีที่สุดตามเกณฑ์ประสิทธิภาพและตรวจสอบไฮเปอร์พารามิเตอร์	ใช้ปรับแต่งโมเดลเพื่อป้องกันการ overfitting
Test Dataset	ชุดข้อมูลที่ใช้ตรวจสอบความสามารถในการ generalize และประเมินประสิทธิภาพของโมเดล	ใช้ทดสอบบนข้อมูลที่ไม่เคยเห็นเพื่อยืนยันความน่าเชื่อถือ
Production Data	ข้อมูลที่ป้อนเข้าโมเดลหลังจากนำไปใช้งานจริง ต้องได้รับการติดตาม	ใช้ในสภาพแวดล้อมจริงและตรวจจับการ drift เพื่อการปรับปรุง

วงจรชีวิตข้อมูลใน ML

วงจรชีวิตข้อมูลประกอบด้วยขั้นตอนหลัก 7 ขั้นตอน เพื่อให้แน่ใจว่าข้อมูลมีคุณภาพสูงและเหมาะสมสำหรับผลิตภัณฑ์ที่ใช้ ML แต่ละขั้นตอนต้องมีการเอกสารเพื่อความสามารถในการตรวจสอบย้อนกลับ (traceability) และการปฏิบัติตามกฎระเบียบ

- **การกำหนดข้อกำหนดข้อมูล (Data Requirements Definition)**: กำหนดลักษณะข้อมูลที่ต้องการ เช่น แหล่งข้อมูลจริง (Real-World Data: RWD) หรือข้อมูลสังเคราะห์ (Synthetic Data) ความเข้ากันได้กับอุปกรณ์ ความหลากหลายของประชากร แหล่งกำเนิดและปริมาณข้อมูล กรณีขอบเขต (Edge Cases) มาตรฐานรูปแบบข้อมูล ความสมบูรณ์ และด้านกฎหมาย (เช่น ความยินยอมและวัตถุประสงค์การเก็บข้อมูลเดิม)
- **การรวบรวมและได้มาซึ่งข้อมูล (Data Collection and Acquisition)**: รวบรวมข้อมูลจากแหล่งต่างๆ เช่น ชุดข้อมูลสาธารณะ การสำรวจ สื่อสังคมออนไลน์ เซ็นเซอร์ IoT ผู้ให้บริการข้อมูลเชิงพาณิชย์ ข้อมูลรัฐบาล การศึกษาทดลอง หรือ crowdsourcing ต้องพิจารณาความโปร่งใสของแหล่งข้อมูล ข้อตกลงการใช้งาน/ใบอนุญาต ข้อมูลจากแหล่งเดียวหรือหลายแหล่ง การประเมินคุณภาพข้อมูล (ความเกี่ยวข้อง ความสมบูรณ์ อคติ) และจัดการแหล่งข้อมูลเหมือนซัพพลายเออร์
- **การควบคุมเวอร์ชันและเอกสาร (Version Control and Documentation)**: จัดการเวอร์ชันของข้อมูล (ข้อมูลดิบ ข้อมูลที่ประมวลผล ข้อมูลที่มีป้ายกำกับ ข้อมูลที่แบ่ง) เมตาดาต้า (ชื่อ เวอร์ชัน แหล่งกำเนิด ผู้มีส่วนร่วม โปรโตคอลการได้มา ขอบเขตความยินยอม และข้อจำกัดการใช้งาน) การควบคุมการเปลี่ยนแปลง ความสมบูรณ์และการควบคุมการเข้าถึง ใช้เครื่องมือเช่น MLflow, LakeFS หรือ DVC ขึ้นอยู่กับลักษณะข้อมูล
- **การประมวลผลข้อมูลล่วงหน้า (Data Preprocessing)**: เตรียมข้อมูลให้พร้อมใช้งาน โดยเอกสารสมมติฐาน ตรรกะ และเหตุผลทั้งหมด รวมถึงการกรอง (การกำจัดข้อมูลที่ไม่เกี่ยวข้องด้วยเกณฑ์และกฎ) การทำให้เป็นมาตรฐาน (Normalization: การแปลงคุณลักษณะให้อยู่ในสเกลมาตรฐาน) การเติมข้อมูลสูญหาย (Imputation: วิธีการเติมและคุณลักษณะที่ได้รับผลกระทบ) และการเพิ่มข้อมูล (Augmentation: การสร้างข้อมูลเทียมโดยการปรับแต่งข้อมูลที่มี เช่น การพลิก การหมุน การเพิ่มเสียงรบกวน เพื่อเพิ่มความแข็งแกร่งของโมเดล)
- **การติดป้ายและหมายเหตุข้อมูล (Data Labelling and Annotation)**: ติดป้ายข้อมูลเพื่อสร้าง ground truth สำหรับการฝึกโมเดล กำหนดโปรโตคอลการติดป้าย แหล่ง ground truth การควบคุมอคติ การควบคุมคุณภาพ (การฝึกอบรม คุณสมบัติ คำสั่ง และการตรวจสอบ) และการควบคุมเวอร์ชันเพื่อความสามารถในการติดตาม ประเภทป้าย: ป้ายด้วยมือ (hand labels: ช้า แพง มีอคติ) หรือป้ายธรรมชาติ (natural labels: อัตโนมัติจากข้อมูลเอง)
- **การแบ่งข้อมูล (Data Splitting)**: แบ่งข้อมูลเป็นชุดฝึกอบรม การตรวจสอบ และการทดสอบที่เป็นอิสระและเป็นตัวแทน กลยุทธ์การแบ่ง (สุ่ม Stratified เวลา แหล่งข้อมูล ระดับผู้ป่วย) การป้องกัน data leakage และการสมดุลคลาสรวมถึงเหตุการณ์หายาก (ด้วยเทคนิค sampling)
- **การเก็บรักษาและกำจัดข้อมูล (Data Retention and Disposal)**: ปฏิบัติตามข้อกำหนดการเก็บรักษาข้อมูลตามกฎระเบียบ ข้อมูลที่ต้องเก็บ: ข้อมูลดิบ การประมวลผล ชุดแบ่งสุดท้าย การควบคุมคุณภาพ ให้แน่ใจถึงความสมบูรณ์และความปลอดภัย และมีวิธีการกำจัดข้อมูลอย่างปลอดภัย

แหล่งข้อมูลและรูปแบบข้อมูล

ข้อมูลสามารถมาจากแหล่งหลากหลาย โดยต้องพิจารณาข้อกำหนดเฉพาะโครงการ เช่น ประเภทข้อมูล คุณภาพ ความถูกต้อง และด้านจริยธรรม/กฎหมาย ตัวอย่างแหล่งข้อมูล: ชุดข้อมูลสาธารณะ การสำรวจ สื่อสังคม เซ็นเซอร์ IoT ผู้ให้บริการเชิงพาณิชย์ ข้อมูลรัฐบาล การศึกษาทดลอง และ crowdsourcing

รูปแบบข้อมูล (Data Formats): การทำให้ข้อมูลอยู่ในรูปแบบที่เก็บหรือส่งได้ เช่น JSON, CSV, Parquet, Avro, Protobuf, Pickle

โมเดลข้อมูล (Data Models): โครงสร้างการจัดระเบียบข้อมูลสำหรับ AI เช่น การจัดข้อมูลผู้ป่วยตามคุณลักษณะ (อายุ อาการ ประวัติการแพทย์) เพื่อพยากรณ์โรค

ประเภทการประมวลผลข้อมูล

- **Batch Processing**: ประมวลผลข้อมูลเป็นชุดใหญ่ตามกำหนดเวลา เหมาะสำหรับงานที่ไม่ต้องการเรียลไทม์ เช่น การฝึกโมเดลบนข้อมูลประวัติศาสตร์ (ตัวอย่าง: ชุดภาพทางการแพทย์เก็บไว้เพื่อฝึกโมเดลตรวจโรค)
- **Stream Processing**: ประมวลผลข้อมูลแบบเรียลไทม์ เหมาะสำหรับงานที่ต้องการข้อมูลทันที เช่น การตรวจจับความผิดปกติจากอุปกรณ์ทางการแพทย์ (ตัวอย่าง: การวิเคราะห์ข้อมูลจากเครื่องวัดหัวใจเพื่อตรวจจับจังหวะหัวใจผิดปกติ)

การสุ่มตัวอย่างข้อมูล (Data Sampling)

การสุ่มตัวอย่างคือการเลือกชุดข้อมูลย่อยเพื่อการฝึกและทดสอบที่รวดเร็วและประหยัด เพื่อหลีกเลี่ยงอคติ ต้องเลือกวิธีการสุ่มที่เหมาะสม แบ่งเป็น 2 ประเภทหลัก:

- **Nonprobability Sampling**: การสุ่มที่ไม่ใช้ความน่าจะเป็น อาจก่อให้เกิดอคติ ตัวอย่าง: Convenience (เลือกตามความสะดวก), Judgmental (ตามดุลยพินิจ), Snowball (ผู้เข้าร่วมแนะนำต่อ), Quota (ตามโควต้าโดยไม่สุ่ม)
- **Random Sampling**: การสุ่มที่ให้โอกาสเท่ากัน ลดอคติ ตัวอย่าง: Simple Random (สุ่มเท่ากัน), Stratified Random (แบ่งกลุ่มแล้วสุ่ม), Weighted (ให้น้ำหนักตามความสำคัญ), Systematic (เลือกทุก n ตัวหลังจุดเริ่มสุ่ม)

ปัญหาในการสุ่ม: ตัวอย่างที่ไม่เป็นตัวแทนอาจนำไปสู่อคติ เช่น การสุ่มเฉพาะผู้สูงอายุจากประชากรทั่วไป

Session 8: AI System Impact Assessment

ภาพรวมของการประเมินผลกระทบของระบบ AI

การประเมินผลกระทบของระบบ AI (AI System Impact Assessment: AIA) เป็นกระบวนการสำคัญที่องค์กรต้องดำเนินการเพื่อประเมินผลกระทบที่อาจเกิดขึ้นจากการใช้งานหรือพัฒนาระบบ AI ไม่ว่าจะเป็นผลกระทบที่ตั้งใจ (intended outcomes) หรือไม่ตั้งใจ (unintended outcomes) กระบวนการนี้ช่วยให้องค์กรสามารถระบุความเสี่ยง จัดลำดับความสำคัญ และวางแผนการลดความเสี่ยง เพื่อให้สอดคล้องกับหลักการรับผิดชอบและจริยธรรมในการใช้ AI ตามมาตรฐาน ISO 42001 ซึ่งเผยแพร่ในเดือนธันวาคม 2566 (2023) โดยองค์การระหว่างประเทศว่าด้วยการมาตรฐาน (ISO) และคณะกรรมการระหว่างประเทศว่าด้วยมาตรฐานทางไฟฟ้า (IEC)

AIA ไม่ใช่เพียงการปฏิบัติตามกฎระเบียบ แต่เป็นเครื่องมือเชิงกลยุทธ์ที่ช่วยเพิ่มความเชื่อมั่น ความโปร่งใส และความน่าเชื่อถือของระบบ AI โดยเฉพาะในอุตสาหกรรมที่มีความเสี่ยงสูง เช่น การแพทย์ การเงิน และการผลิต องค์กรที่นำ AIA ไปใช้จะสามารถบูรณาการระบบจัดการ AI เข้ากับโครงสร้าง ISO ที่มีอยู่ เช่น ISO 9001 (คุณภาพ) หรือ ISO 27001 (ความมั่นคงปลอดภัยสารสนเทศ) เพื่อลดความซ้ำซ้อนและเพิ่มประสิทธิภาพ

ประโยชน์ของ AIA

- ลดความเสี่ยง: ช่วยระบุและจัดการความเสี่ยง เช่น อคติในข้อมูล (bias) หรือการละเมิดความเป็นส่วนตัว ก่อนที่ระบบจะนำไปใช้งานจริง
- เพิ่มความเชื่อมั่น: สร้างความโปร่งใสให้ผู้มีส่วนได้เสีย เช่น ลูกค้า พนักงาน และหน่วยงานกำกับดูแล
- สนับสนุนการปฏิบัติตามกฎระเบียบ: สอดคล้องกับ EU AI Act, OECD AI Principles และกฎหมายท้องถิ่น เช่น Local Law 144 ในนิวยอร์ก
- ส่งเสริมนวัตกรรม: ช่วยให้องค์กรใช้ AI อย่างรับผิดชอบ ส่งผลให้เกิดโอกาสทางธุรกิจใหม่ๆ
- เอกสารสำหรับการตรวจสอบ: บันทึกกระบวนการเพื่อใช้ในการตรวจสอบหรือรายงานผลกระทบ

องค์ประกอบหลักของกระบวนการ AIA

กระบวนการ AIA ประกอบด้วยองค์ประกอบหลัก 5 ประการ ซึ่งเป็นส่วนหนึ่งของระบบจัดการ AI (AIMS) ตาม ISO 42001 โดยแต่ละองค์ประกอบเชื่อมโยงกับขั้นตอนการวางแผนและดำเนินงาน (Planning and Operation) ในมาตรฐาน

1. การระบุ (Identification)

ขั้นตอนนี้มุ่งเน้นการระบุองค์ประกอบพื้นฐานของระบบ AI เพื่อเข้าใจขอบเขตและวัตถุประสงค์

- แหล่งที่มาของ AI (AI Sources): ระบุว่าองค์กรใช้งาน AI จากผู้ให้บริการภายนอก (เช่น ChatGPT, Google Cloud AI) หรือพัฒนาภายใน (in-house development)
- เหตุผลในการใช้งาน AI (Reasons for Using AI): กำหนดวัตถุประสงค์ เช่น เพิ่มประสิทธิภาพการผลิต ลดต้นทุน หรือปรับปรุงการบริการลูกค้า
- ผลลัพธ์ที่ตั้งใจและไม่ตั้งใจ (Intended and Unintended Outcomes): วิเคราะห์ผลกระทบที่คาดหวัง (เช่น การพยากรณ์ยอดขายที่แม่นยำขึ้น) และความเสี่ยงที่อาจเกิด (เช่น การตัดสินใจที่ไม่เป็นธรรมจากอคติ)
- ผู้มีส่วนได้เสีย (Stakeholders): ระบุบุคคลหรือกลุ่มที่ได้รับผลกระทบ เช่น พนักงาน ลูกค้า ชุมชน หรือหน่วยงานกำกับดูแล

2. การวิเคราะห์ (Analysis)

วิเคราะห์ความน่าจะเป็นและผลกระทบของความเสี่ยงที่เกิดจากการใช้งานระบบ AI โดยใช้เครื่องมือเช่น Risk Matrix เพื่อประเมินระดับความเสี่ยง (Low, Medium, High)

- ความน่าจะเป็นของความเสี่ยง (Likelihood of Risks): ประเมินโอกาสที่ความเสี่ยงจะเกิด เช่น ความน่าจะเป็นสูงสำหรับการรั่วไหลของข้อมูลในระบบ cloud-based AI
- ผลกระทบของความเสี่ยง (Consequences of Risks): พิจารณาผลกระทบต่อด้านต่างๆ เช่น ด้านการเงิน (ค่าปรับจาก GDPR) ด้านชื่อเสียง (สูญเสียความเชื่อมั่นจากลูกค้า) หรือด้านสังคม (การเลือกปฏิบัติจาก bias)
- ประเภทความเสี่ยงหลัก:
    - ความเสี่ยงด้านข้อมูล: ข้อมูลไม่ถูกต้องหรือล้าสมัย (เช่น AI ที่ฝึกด้วยข้อมูลถึงปี 2021 เท่านั้น)
    - ความเสี่ยงด้านอคติ: ผลลัพธ์ที่ไม่เป็นธรรมต่อกลุ่มคน (เช่น ระบบสินเชื่อที่ปฏิเสธผู้สมัครจากกลุ่มเฉพาะ)
    - ความเสี่ยงด้านความมั่นคง: การโจมตีแบบ prompt injection หรือ data poisoning
    - ความเสี่ยงด้านจริยธรรม: การละเมิดสิทธิส่วนบุคคลหรือการลอกเลียน (plagiarism)

ตัวอย่างความเสี่ยง	ความน่าจะเป็น	ผลกระทบ	ระดับความเสี่ยง
อคติในระบบคัดเลือกพนักงาน	สูง	ชื่อเสียงและกฎหมาย	สูง
ข้อมูลล้าสมัยใน AI พยากรณ์ตลาด	ปานกลาง	การเงิน	ปานกลาง

3. การประเมิน (Evaluation)

ทบทวนความเสี่ยงทั้งหมดและจัดลำดับความสำคัญเพื่อกำหนดการดำเนินการและผู้รับผิดชอบ

- **การทบทวนความเสี่ยง (Reviewing All Risks)**: ใช้เกณฑ์เช่น Impact x Likelihood เพื่อจัดลำดับ
- **การจัดลำดับความสำคัญ (Prioritising Risks)**: มุ่งเน้นความเสี่ยงสูงก่อน เช่น ความเสี่ยงที่อาจนำไปสู่การละเมิดกฎหมาย
- **การกำหนดเจ้าของ (Ownership)**: มอบหมายผู้รับผิดชอบ เช่น ทีม IT สำหรับความเสี่ยงด้านความมั่นคง หรือทีมกฎหมายสำหรับความเสี่ยงด้านจริยธรรม
- **การพิจารณาโอกาส (Opportunities)**: ไม่เพียงความเสี่ยง แต่รวมโอกาส เช่น การใช้ AI เพื่อเพิ่มความหลากหลายในข้อมูลฝึก

4. การรักษาความเสี่ยงหรือโอกาส (Treatment)

วางแผนการจัดการความเสี่ยงหรือโอกาสที่ระบุ โดยเลือกกลยุทธ์ที่เหมาะสม

- **มาตรการลดความเสี่ยง (Mitigation Measures)**:
  - การตรวจสอบข้อมูล (data validation) เพื่อลดอคติ
  - การฝึกอบรมพนักงานเกี่ยวกับการใช้งาน AI อย่างรับผิดชอบ
  - การใช้เครื่องมือตรวจสอบ bias เช่น Fairlearn หรือ Aequitas
  - การบูรณาการ human oversight เพื่อตรวจสอบผลลัพธ์ AI
- **การยอมรับหรือหลีกเลี่ยง**: สำหรับความเสี่ยงต่ำ อาจยอมรับโดยมีแผนสำรอง
- **ตัวอย่าง**: ในระบบ AI ทางการแพทย์ ใช้ synthetic data เพื่อหลีกเลี่ยงการละเมิดความเป็นส่วนตัว

5. สรุปและเอกสาร (Conclusion and Documentation)

บันทึกผลการประเมินเพื่อใช้ในการรายงานหรือสื่อสาร

- **การสรุปผล**: ระบุผลกระทบโดยรวมและแผนการติดตาม
- **การบันทึก**: ใช้เทมเพลต AIA เพื่อเก็บข้อมูล เช่น ชื่อระบบ AI วันที่ประเมิน ผู้เข้าร่วม และผลลัพธ์
- **การสื่อสาร**: แจ้งผู้มีส่วนได้เสียและปรับปรุงตาม feedback เพื่อการปรับปรุงต่อเนื่อง

การบูรณาการ AIA กับ ISO 42001

AIA เป็นส่วนสำคัญของ ISO 42001 โดยเฉพาะในข้อ 6 (Planning) และข้อ 8 (Operation) ซึ่งกำหนดให้องค์กรต้องประเมินผลกระทบของระบบ AI ต่อบุคคล กลุ่ม และสังคม โดยพิจารณาความยุติธรรม ความโปร่งใส และความปลอดภัย กระบวนการนี้ช่วยให้องค์กรปฏิบัติตามภาคผนวก A (Control Objectives) เช่น การประเมินผลกระทบ (A.5) และวงจรชีวิตระบบ AI (A.6) Blackmores แนะนำการใช้ Isology® 7-Step Methodology เพื่อนำ AIA ไปปฏิบัติ:

1. Gap Analysis: ตรวจสอบช่องว่างปัจจุบัน
2. Planning: วางแผน AIA
3. Implementation: ดำเนินการประเมิน
4. Training: ฝึกอบรมทีม
5. Internal Audit: ตรวจสอบภายใน
6. Certification: เตรียมรับรอง
7. Continual Improvement: ปรับปรุงต่อเนื่อง

ความท้าทายและคำแนะนำ

- **ความท้าทาย**: ข้อมูลที่ซับซ้อน การขาดทรัพยากร และการเปลี่ยนแปลงกฎระเบียบรวดเร็ว
- **คำแนะนำ**: เริ่มด้วยโครงการนำร่อง ใช้เครื่องมือดิจิทัลสำหรับการบันทึก และร่วมมือกับที่ปรึกษาเช่น Blackmores เพื่อรับรอง ISO 42001

Session 9: Bias, Fairness and Transparency of AI

อคติ ความเป็นธรรม และความโปร่งใสของ AI

อคติใน AI เป็นปัญหาสำคัญที่อาจนำไปสู่ผลกระทบเชิงลบ เช่น การเลือกปฏิบัติหรือความไม่แม่นยำ การฝึกอบรมนี้เน้นการระบุ จัดการ และประเมินอคติเพื่อให้ระบบ AI มีความเป็นธรรมและโปร่งใส

ปัญหาอคติใน AI

อคติสามารถนำไปสู่ผลกระทบรุนแรง เช่น:

- **ระบบสาธารณสุขในสหรัฐฯ**: อัลกอริทึมทำนายความต้องการดูแลพิเศษโดยใช้ต้นทุนทางการแพทย์ ซึ่งต่ำกว่าในกลุ่มคนผิวดำแม้จะมีโรคเรื้อรังเท่ากัน ส่งผลให้คนผิวขาวได้รับการดูแลมากกว่า (ที่มา: The Washington Post, 2019)
- **เครื่องวัดออกซิเจน**: ทำงานน้อยลงในผู้ที่ผิวคล้ำ ส่งผลให้ผู้ป่วยผิวคล้ำได้รับออกซิเจนน้อยกว่า อาจเป็นสาเหตุอัตราการตายจาก COVID สูงในกลุ่มชนกลุ่มน้อย (ที่มา: The Guardian)

นิยามอคติและความเป็นธรรม

- **Bias**: ความแตกต่างอย่างเป็นระบบในการปฏิบัติต่อวัตถุ บุคคล หรือกลุ่มบางกลุ่มเมื่อเทียบกับกลุ่มอื่น (ISO) ใน AI คือแนวโน้มที่จะผลิตผลลัพธ์ผิดพลาดอย่างเป็นระบบ ซึ่งอาจกระทบความปลอดภัยและประสิทธิภาพในกลุ่มย่อยของประชากรเป้าหมาย (FDA) อคติไม่ใช่สิ่งเลวร้ายเสมอไป เพราะอัลกอริทึมการจำแนกและคลัสเตอร์ต้องการอคติเพื่อทำงาน แต่ต้องหลีกเลี่ยง "อคติที่ไม่พึงประสงค์"
- **AI Bias**: แนวโน้มที่จะผลิตผลลัพธ์ผิดพลาดอย่างเป็นระบบและคาดเดาไม่ได้บางครั้ง
- **Fairness**: การสร้างระบบที่ตัดสินใจอย่างเป็นกลาง ให้การปฏิบัติเท่าเทียมกันตามคุณลักษณะ เช่น เชื้อชาติ เพศ อายุ โดยเคารพข้อเท็จจริง ความเชื่อ และบรรทัดฐาน สังคม โดยปราศจากความลำเอียงหรือการเลือกปฏิบัติที่ไม่ยุติธรรม อคติเป็นเพียงส่วนหนึ่งที่กระทบความเป็นธรรม
- **ประเภทผลกระทบที่ไม่เป็นธรรม**: การจัดสรรไม่ยุติธรรม คุณภาพบริการไม่เท่าเทียม การเสริมสร้างแบบแผน การแทนที่มากหรือน้อยเกินไปในกลุ่มย่อย

หมายเหตุ: ระบบที่มีอคติอาจไม่ไม่ยุติธรรมเสมอไป และระบบที่ไม่มีอคติอาจไม่ยุติธรรมเสมอไป

ความท้าทายในการจัดการอคติ

- **ขนาดใหญ่**: ยากในการตรวจจับและแก้ไขในระบบขนาดใหญ่
- **ไม่มีเกณฑ์ชัดเจนว่าอะไรคือความเป็นธรรม**: ความเป็นธรรมขึ้นกับบริบทและอาจขัดแย้งกัน
- **การแลกเปลี่ยน (Trade-offs)**: การลดอคติอาจลดประสิทธิภาพโดยรวม
- **ยากในการตรวจจับ**: อคติอาจซ่อนเร้น
- **กระบวนการต่อเนื่อง**: ต้องตรวจสอบตลอดวงจรชีวิต

ประเภทอคติ

อคติที่ไม่พึงประสงค์ในผลิตภัณฑ์ AI เกิดจาก 3 แหล่งหลัก:

- **Human Cognitive Bias**: อคติทางความคิดของมนุษย์ เช่น Automation Bias (เชื่อระบบอัตโนมัติมากเกินไป), Societal Bias (อคติสังคมที่ฝังในนโยบาย), Confirmation Bias (เน้นข้อมูลที่ยืนยันความเชื่อเดิม), Out-group Homogeneity Bias (มองกลุ่มภายนอกว่าคล้ายกันมากกว่ากลุ่มตนเอง)
- **Data Bias**: อคติจากข้อมูล เช่น Sampling Bias (ข้อมูลไม่สุ่ม), Coverage Bias (ข้อมูลไม่ครอบคลุมประชากรจริง), Confounding Variables Bias (ความสัมพันธ์ซ่อนเร้นระหว่างข้อมูลเข้าและป้าย), Non-response Bias (กลุ่มย่อยปฏิเสธการเข้าร่วม)
- **Bias from Engineering Decisions**: อคติจากทางเลือกทางวิศวกรรม เช่น Feature Engineering (เลือกหรือสร้างคุณลักษณะผิด), Algorithm Selection (อัลกอริทึมขยายอคติ), Hyperparameter Tuning (ปรับพารามิเตอร์ที่นำไปสู่อคติ), Model Bias (อัลกอริทึมขยายอคติในข้อมูลไม่สมดุล), Model Interaction (อคติจากโมเดลหนึ่งขยายไปยังอีกโมเดล), Informativeness Bias (ความสัมพันธ์ระหว่างข้อมูลเข้าและผลลัพธ์ยากต่อการเรียนรู้ในบางกลุ่ม)

เมื่อใดที่อคติเกิดขึ้น

อคติเกิดได้ตลอดวงจรชีวิตระบบ AI และ ML pipeline:

- Inception ⟶ Data Acquisition

- Design and Development ⟶ Data Preparation, Modelling
- Verification and Validation ⟶ Verification and Validation
- Deployment ⟶ Model Deployment
- Operation and Monitoring ⟶ Operation
- Re-evaluation, Retirement

การรั่วไหลของข้อมูล (Data Leakage)

Data Leakage คือการรั่วไหลของป้ายหรือข้อมูลที่ไม่ควรมีในชุดฝึก แต่มีในชุดทดสอบ สาเหตุหลัก:

- การแบ่งข้อมูลเวลาแบบสุ่มแทนตามเวลา
- การทำให้เป็นมาตรฐานก่อนแบ่ง
- การเติมข้อมูลสูญหายด้วยสถิติจากชุดทดสอบ
- การรั่วไหลกลุ่ม (กลุ่มที่มีความสัมพันธ์สูงถูกแบ่งไปคนละชุด)
- การรั่วไหลจากกระบวนการสร้างข้อมูล (กระบวนการสร้างข้อมูลกระทบข้อมูลและถูกใช้ในการพยากรณ์)

การจัดการอคติจากข้อมูล (Data Bias Treatment)

- การจัดการป้ายและกระบวนการติดป้าย
- การจัดการคุณลักษณะและป้ายสูญหาย
- การประมวลผลและรวมข้อมูล
- เทคนิคการสุ่มตัวอย่าง
- การจัดการ data leakage
- ความเป็นตัวแทนของชุดข้อมูล
- การสร้างเอกสารข้อมูล

การจัดการอคติจากวิศวกรรม (Engineering Bias Treatment)

ใช้วิธีโมเดล-based เช่น:

- Iterative Orthogonal Feature Projection (IOFP): จัดอันดับคุณลักษณะตามอิทธิพล
- Minimum Redundancy, Maximum Relevance (MRMR): เลือกคุณลักษณะที่เกี่ยวข้องสูงสุดและซ้ำซ้อนน้อย
- Ridge/LASSO Regression: ป้องกัน overfitting และเลือกคุณลักษณะ
- Random Forest: รวมคะแนนความสำคัญคุณลักษณะจากหลายต้นไม้
- การเพิ่ม regularization ใน optimization หรือ representation learning เพื่อลดผลกระทบตัวแปรบางตัว

การประเมินอคติและความเป็นธรรม

ใช้อัลกอริทึม fairness metrics เพื่อตรวจสอบผลลัพธ์ ตัวอย่าง:

- **Demographic Parity**: อัตราการพยากรณ์บวกเท่ากันทุกกลุ่มประชากร โดยไม่คำนึงถึง ground truth (ตัวอย่าง: อัตราการรับเข้าเรียน 20% เท่ากันทั้งกลุ่มคนส่วนใหญ่ และกลุ่มคนส่วนน้อย)
- **Equality of Opportunity**: อัตรา true positive (TPR) เท่ากันในกลุ่มที่สมควรได้รับผลบวก (ตัวอย่าง: อัตราการรับเข้าเรียน 40% ในกลุ่ม qualified ทั้งหมด)
- **Predictive Equality**: อัตรา false positive (FPR) เท่ากัน (ตัวอย่าง: อัตราการปฏิเสธ unqualified 10% เท่ากัน)

พิจารณาคุณลักษณะอิสระ vs การตัดกัน (intersectional) เช่น ระบบที่ไม่มีอคติต่อเพศหรือเชื้อชาติแยกกัน อาจมีอคติต่อเพศหญิงผิวดำ กำหนดวัตถุประสงค์ fairness และ delta (ส่วนต่างที่ยอมรับ) ก่อนทดสอบ

มาตรการควบคุมอคติ

ป้องกันและควบคุมอคติตลอดวงจรชีวิต:

- การฝึกอบรมและสร้างความตระหนักทีม

- การควบคุมคุณภาพข้อมูลและบันทึก
- เอกสารและบันทึก
- ประสิทธิภาพรวม vs ประสิทธิภาพกลุ่มย่อย
- ข้อมูลสำหรับผู้ใช้
- การติดตามหลังตลาดอย่างต่อเนื่อง

ความโปร่งใสของ AI (AI Transparency)

ความโปร่งใสคือการเปิดเผยส่วนประกอบ ข้อมูล และกระบวนการของระบบ AI เพื่อให้ผู้มีส่วนได้เสียประเมินความสอดคล้องกับค่านิยม เช่น ความเป็นธรรม ความเป็นส่วนตัว และจริยธรรม สร้างความรับผิดชอบและความเชื่อมั่น

ความไม่โปร่งใส (Opaqueness)

ระบบ AI อาจไม่โปร่งใสจาก:

- โครงสร้างโมเดลซับซ้อน
- ขาดการมองเห็นกระบวนการข้อมูล
- ไม่ต้องอธิบาย
- เน้นประสิทธิภาพสูง
- อัลกอริทึมที่ซ่อนหรือกรรมสิทธิ์

ระบบโปร่งใสมี: กระบวนการตัดสินใจชัดเจน การติดตามข้อมูล การอธิบาย การออกแบบเพื่อผู้มีส่วนได้เสีย อัลกอริทึมที่ตรวจสอบได้

ความไม่สามารถคาดเดา (Unpredictability)

ความสามารถคาดเดาคือการอนุมานการกระทำถัดไปของ AI ในสภาพแวดล้อม ความไม่สามารถคาดเดาเกิดจาก:

- องค์ประกอบข้อมูลฝึกและโมเดล
- สถานการณ์สับสนโดยขาด fail-safe
- การเรียนรู้ต่อเนื่องที่เปลี่ยน "ตรรกะ" ตามเวลา

ระดับความสามารถอธิบาย (Levels of Explainability)

เลือกระดับตามบริบทใช้งาน โดยพิจารณา: ข้อมูลส่วนบุคคลที่ละเอียดอ่อน ผลกระทบต่อสวัสดิภาพ ผลกระทบรุนแรงจากความล้มเหลว การจำกัดความเป็นอิสระ ผลกระทบต่อสังคม

- **Basic**: ภาพรวมวัตถุประสงค์ สำหรับใช้งานความเสี่ยงต่ำ
- **Intermediate**: อธิบายปัจจัยตัดสินใจ สำหรับผลกระทบปานกลาง
- **High**: รายละเอียดการทำงาน สำหรับพื้นที่เสี่ยงสูง เช่น สุขภาพ
- **Full Transparency**: เข้าถึงข้อมูล อัลกอริทึม กระบวนการ สำหรับกรณีที่มีความละเอียดอ่อนหรือส่งผลกระทบสูง

วัตถุประสงค์การอธิบาย (Aims of Explanation)

ระบบ AI ที่สามารถอธิบายได้ชี้แจงกระบวนการเพื่อให้ผลลัพธ์ถูกต้องและสมเหตุสมผล ไม่ใช่แค่ "ทำงาน" การอธิบายแตกต่างตามผู้มีส่วนได้เสีย:

- **Causation**: วิธีบรรลุผล (ตัวอย่าง: วินิจฉัยโรคปอดอักเสบ โดยระบุอาการ X-ray และ biomarkers)
- **Knowledge Base**: ข้อมูลที่อาศัย (ฐานข้อมูลเคสคล้าย แนวทางคลินิก การศึกษาที่ตีพิมพ์)
- **Justification**: เหตุผลที่ผลถูกต้อง (ตรงเกณฑ์การแพทย์และคะแนนเชื่อมั่นสูง)

ระบบโปร่งใสแจ้งผู้มีส่วนได้เสียเกี่ยวกับการเก็บข้อมูล (โดยเฉพาะข้อมูลส่วนบุคคลและวัตถุประสงค์) การตัดสินใจอัตโนมัติ และกลไกแทรกแซงมนุษย์ตามกฎความเป็นส่วนตัว

Session 10: Workshop - Designing an AI Governance Framework

นำเสนอกรอบแนวคิด AIMS และ AI Governance ผ่านกรณีศึกษาที่น่าสนใจและตัวอย่างจากสถานการณ์จริง ผู้เข้าร่วมจะแบ่งกลุ่มย่อยเพื่อสำรวจการประยุกต์ใช้กรอบแนวคิดเหล่านี้ในทางปฏิบัติ แต่ละกลุ่มจะระบุความเสี่ยงหรืออันตรายที่อาจเกิดขึ้นจากเทคโนโลยีและสถานการณ์จำลองแต่ละอย่าง และพัฒนาแนวทางแก้ไขที่เป็นไปได้ โดยได้รับการสนับสนุนจาก RP ผู้เข้าร่วมจะถูกขอให้นำเสนอการอภิปรายและแนวทางแก้ไขหลังจากพักเบรก

Session 11: Participant Presentations

แต่ละกลุ่มได้นำเสนอแนวทางแก้ไขปัญหา พร้อมอธิบายถึงความเสี่ยง ความท้าทาย และการอภิปรายที่ได้สำรวจจากกรณีศึกษา แต่ละกลุ่มได้แบ่งปันข้อมูลเชิงลึก แนวทางแก้ไขปัญหา และเน้นย้ำถึงเป้าหมายสำคัญหรือองค์ประกอบการกำกับดูแลที่กลุ่มจะให้ความสำคัญเป็นลำดับแรกสำหรับกรณีศึกษาของตนได้ทำ โดยกลุ่มได้รับมอบให้ให้ดำเนินการ

Case Study 3: AI in Public Safety-Predictive Policing

ปัญหาที่ต้องแก้ไข

แผนกตำรวจ MetroCity เผชิญกับทรัพยากรจำกัดและความต้องการบริการความมั่นคงสาธารณะที่เพิ่มขึ้น พวกเขาอยู่ภายใต้แรงกดดันในการจัดสรรตำรวจอย่างมีประสิทธิภาพ ลดอัตราอาชญากรรม และปรับปรุงเวลาในการตอบสนอง

โซลูชัน AI ที่เสนอ ระบบตำรวจแบบพยากรณ์จะ:

- วิเคราะห์ข้อมูลอาชญากรรมประวัติศาสตร์ เช่น ประเภทอาชญากรรม สถานที่ และเวลาของวัน
- ใช้ข้อมูลทางเศรษฐกิจสังคมและประชากรศาสตร์เกี่ยวกับย่านต่างๆ
- สร้างแผนที่ความเสี่ยงอาชญากรรมที่แสดงพื้นที่ที่มีโอกาสเกิดอาชญากรรมสูงกว่า
- เสนอแนะลำดับความสำคัญในการลาดตระเวนสำหรับแต่ละกะ

เหตุผลที่น่าสนใจสำหรับการวิเคราะห์

- **ประโยชน์ที่คาดหวัง**: การใช้ตำรวจอย่างมีประสิทธิภาพ การป้องกันอาชญากรรม เวลาตอบสนองที่รวดเร็วขึ้น
- **ประเด็นที่ต้องพิจารณา**:
    - AI อาจเสริมสร้างความไม่เท่าเทียมที่มีอยู่หากฝึกจากข้อมูลประวัติศาสตร์ที่มีอคติ เช่น การตำรวจมากเกินในชุมชนบางแห่ง
    - ระดับความโปร่งใสกับสาธารณชน — ผู้คนควรรู้หรือไม่ว่าย่านของพวกเขาถูกธงว่า "ความเสี่ยงสูง"?
    - ผู้รับผิดชอบหากการมีตำรวจถูกจัดสรรไม่สมส่วน นำไปสู่ความไม่เชื่อมั่นในชุมชน?
    - วิธีการกำกับดูแลเพื่อให้แน่ใจว่า AI ไม่กลายเป็น "คำทำนายที่สมหวังด้วยตนเอง" (ตำรวจมากขึ้น = อาชญากรรมบันทึกมากขึ้น = ตำรวจมากขึ้น)?

จุดสนทนาอื่นๆ

- ผู้มีส่วนได้เสีย: ชุมชน ตำรวจ ผู้นำพลเมือง หน่วยงานกำกับดูแล
- ความเสี่ยง: อคติ ความเชื่อมั่นในชุมชน ผลกระทบที่ไม่คาดคิด
- การกำกับดูแล: ความรับผิดชอบต่อการตัดสินใจจัดสรร ความโปร่งใสกับสาธารณชน กลไกกำกับดูแล

ผลการ Discussion มีรายละเอียดดังนี้

- **บริบท**: แผนกตำรวจ MetroCity เผชิญกับทรัพยากรจำกัดและความต้องการบริการความมั่นคงสาธารณะที่เพิ่มขึ้น พวกเขาอยู่ภายใต้แรงกดดันในการจัดสรรตำรวจอย่างมีประสิทธิภาพ ลดอัตราอาชญากรรม และปรับปรุงเวลาในการตอบสนอง

- **โซลูชัน AI**: ระบบการตรวจตราเชิงคาดการณ์จะวิเคราะห์ข้อมูลอาชญากรรมประวัติศาสตร์ (ประเภทอาชญากรรม สถานที่ เวลา) ใช้ข้อมูลทางเศรษฐกิจสังคมและประชากรศาสตร์ สร้างแผนที่ความเสี่ยงอาชญากรรม และเสนอแนะลำดับความสำคัญในการลาดตระเวนสำหรับแต่ละกะ
- **ความเสี่ยงหลัก**: AI อาจเสริมสร้างความไม่เท่าเทียมที่มีอยู่หากฝึกจากข้อมูลประวัติศาสตร์ที่มีอคติ เช่น การตำรวจมากเกินในชุมชนบางแห่ง; การขาดความโปร่งใสอาจทำให้สาธารณชนไม่เชื่อมั่น; การเป็น "คำทำนายที่สมหวังด้วยตนเอง" (ตำรวจมากขึ้นนำไปสู่อาชญากรรมบันทึกมากขึ้น)
- **มาตรการแนะนำ**: ใช้ Risk Management (DARTS) เพื่อตรวจสอบอคติในข้อมูลและอัปเดตโมเดลรายไตรมาส; Transparency ผ่านการรายงานสาธารณะและ Stakeholder Engagement กับชุมชนเพื่อรวบรวม feedback; Accountability โดยกำหนดผู้รับผิดชอบต่อการตัดสินใจและกลไกกำกับดูแลจากหน่วยงานรัฐ

กรณีศึกษานี้ช่วยให้กลุ่มฝึกการประยุกต์ DARTS Framework ในบริบทจริง โดยเน้นการเชื่อมโยงกับกฎหมายเวียดนาม เช่น Personal Data Protection Decree (2023) เพื่อป้องกันการละเมิดข้อมูลประชาชน

การนำเสนอและบทเรียนที่ได้

กลุ่มได้แบ่งปันข้อมูลเชิงลึก “จากกรณีศึกษา Predictive Policing เราพบว่าความเสี่ยงด้านอคติสามารถลดได้ 40% ด้วยการสุ่มตัวอย่างข้อมูลที่หลากหลายและ human oversight" ผู้เชี่ยวชาญจะให้ feedback เพื่อเสริมสร้าง บทเรียนหลักที่ได้จากการฝึกอบรม ได้แก่:

- การกำกับดูแล AI ต้องเป็นกระบวนการต่อเนื่อง ไม่ใช่ครั้งเดียว โดยเฉพาะในภาครัฐที่เกี่ยวข้องกับความมั่นคงสาธารณะ
- การบูรณาการ AIMS กับกรอบงานที่มีอยู่ช่วยลดความซับซ้อน เช่น การใช้ Transparency เพื่อสร้างความเชื่อมั่นในชุมชน
- ในบริบทเวียดนาม การมีส่วนร่วมของชุมชนและรัฐบาลเป็นกุญแจสำคัญในการจัดการอคติและผลกระทบสังคม

สรุปและแนวโน้ม

Workshop นี้ช่วยให้ผู้เข้าร่วมเข้าใจการนำ AIMS ไปใช้จริง โดยผ่านการจำลองสถานการณ์ที่ท้าทาย เช่น ระบบการตรวจตราเชิงคาดการณ์ที่อาจเสริมสร้างความไม่เท่าเทียม แนวโน้มในปี 2568 คือ การขยายกรอบงานเช่น DARTS สู่ภาครัฐในอาเซียน เพื่อสนับสนุนการเติบโตทางเศรษฐกิจที่ยั่งยืน การฝึกอบรมนี้เป็นสะพานเชื่อมระหว่างทฤษฎีและปฏิบัติ เพื่อให้เวียดนามพร้อมสำหรับยุค AI

Session: 12: AIMS Implementation

การวางแผนนำ AIMS ไปใช้ (Plan)

ขั้นตอนการวางแผน (Plan) ซึ่งเป็นส่วนสำคัญในวิธีการ Isology® ของ Blackmores โดยครอบคลุมองค์ประกอบหลัก 5 ประการ เพื่อให้องค์กรสามารถกำหนดกรอบการทำงานที่ชัดเจนและยั่งยืน การวางแผนนี้สอดคล้องกับข้อ 6 (Planning) ใน ISO 42001 ซึ่งกำหนดให้องค์กรต้องประเมินความเสี่ยง โอกาส และผลกระทบของระบบ AI

1. ขอบเขต (Scope)

การกำหนดขอบเขตเป็นขั้นตอนแรกในการวางแผน เพื่อให้แน่ใจว่าระบบ AIMS ครอบคลุมส่วนที่เกี่ยวข้องกับ AI ภายในองค์กร โดยไม่ทำให้ซับซ้อนเกินไป

- **การระบุขอบเขต**: พิจารณาว่าองค์กรใช้งาน AI ในส่วนใด เช่น การพัฒนา AI ภายใน (in-house development) การใช้งาน AI จากผู้ให้บริการภายนอก (เช่น ChatGPT หรือ Google Cloud AI) หรือทั้งสองอย่าง ขอบเขตควรครอบคลุมกระบวนการทั้งหมดตั้งแต่การรวบรวมข้อมูล การฝึกโมเดล ไปจนถึงการนำไปใช้งานจริง

- **ปัจจัยที่ต้องพิจารณา**: บริบทองค์กร (เช่น ขนาดธุรกิจ ภาคอุตสาหกรรม) ผู้มีส่วนได้เสีย (stakeholders) และกฎระเบียบที่เกี่ยวข้อง ตัวอย่าง: สำหรับบริษัทผลิต หากใช้งาน AI ในสายการผลิตเท่านั้น ขอบเขตอาจจำกัดเฉพาะส่วนนั้น เพื่อหลีกเลี่ยงการขยายไปยังแผนกอื่นที่ไม่เกี่ยวข้อง
- **ประโยชน์**: ช่วยให้การนำไปใช้มีประสิทธิภาพและสามารถขยายได้ในอนาคต Blackmores แนะนำให้เริ่มด้วย Gap Analysis เพื่อตรวจสอบช่องว่างระหว่างระบบปัจจุบันกับข้อกำหนด ISO 42001

2. ระยะเวลา (Timescales)

การกำหนดระยะเวลาช่วยให้โครงการดำเนินการได้ตามแผน โดยพิจารณาปัจจัยเช่น ขนาดองค์กร ทรัพยากร และระดับความซับซ้อนของ AI

- **ระยะเวลาที่แนะนำ**: Blackmores ระบุว่าสามารถรับรอง ISO 42001 ได้ภายใน 3-6 เดือนสำหรับองค์กรขนาดเล็กถึงกลาง หากมีระบบ ISO อื่นอยู่แล้ว (เช่น ISO 9001) เนื่องจากโครงสร้างระดับสูง (High Level Structure: HLS) ทำให้บูรณาการได้ง่าย สำหรับองค์กรขนาดใหญ่ อาจใช้เวลา 6-12 เดือน
- **ปัจจัยที่กระทบ**: การฝึกอบรมพนักงาน การประเมินความเสี่ยง AI และการตรวจสอบภายใน ตัวอย่าง: ขั้นตอน Gap Analysis ใช้เวลา 1-2 สัปดาห์ การวางแผนและดำเนินงาน 2-4 เดือน และการตรวจสอบภายนอก 1 เดือน
- **เคล็ดลับ**: ใช้แผนโครงการ (project roadmap) เพื่อติดตามความคืบหน้า และปรับตามความเสี่ยงที่เกิดขึ้น เช่น หากพบปัญหาอคติในข้อมูล อาจต้องขยายเวลาการทดสอบ

3. ความมุ่งมั่นจากผู้นำ (Leadership Commitment)

ความมุ่งมั่นจากผู้นำเป็นปัจจัยสำคัญในการขับเคลื่อนการนำ AIMS ไปใช้ โดยสอดคล้องกับข้อ 5 (Leadership) ใน ISO 42001

- **บทบาทของผู้นำ**: กำหนดนโยบาย AI (AI Policy) ที่สอดคล้องกับเป้าหมายองค์กร สื่อสารให้พนักงานเข้าใจ และจัดสรรทรัพยากรที่จำเป็น ตัวอย่าง: CEO ประกาศนโยบายที่เน้นการใช้งาน AI อย่างมีจริยธรรม เพื่อลดความเสี่ยงด้านชื่อเสียง
- **การแสดงความมุ่งมั่น**: เข้าร่วมการประชุมทบทวนโดยฝ่ายบริหาร (Management Review) และสนับสนุนการฝึกอบรม Blackmores แนะนำให้ผู้นำแต่งตั้ง "AI Champion" เพื่อดูแลโครงการ
- **ประโยชน์**: สร้างวัฒนธรรมองค์กรที่รับผิดชอบต่อ AI ช่วยให้พนักงานยอมรับการเปลี่ยนแปลงและปฏิบัติตามมาตรฐาน

4. ทรัพยากร (Resources)

การจัดสรรทรัพยากรที่เพียงพอเป็นกุญแจสู่ความสำเร็จ โดยครอบคลุมบุคลากร งบประมาณ และเครื่องมือ

- **บุคลากร**: จัดฝึกอบรมพนักงานเกี่ยวกับ ISO 42001 และ AI Governance ผ่าน Isology Hub ของ Blackmores ซึ่งมีวิดีโอ การฝึกอบรมสั้น (Coffee Break Training) และแผนการ (Gameplans) ตัวอย่าง: ฝึกอบรมทีม IT เกี่ยวกับการจัดการข้อมูล AI
- **งบประมาณ**: รวมค่าที่ปรึกษา การฝึกอบรม และเครื่องมือ เช่น ซอฟต์แวร์ตรวจสอบ bias (เช่น Fairlearn) Blackmores เสนอแพ็กเกจ consultancy ที่รวมการเข้าถึงทรัพยากรออนไลน์
- **เครื่องมือและเทคโนโลยี**: ใช้แพลตฟอร์มเช่น Microsoft Teams สำหรับเวิร์กชอป หรือเครื่องมือประเมินความเสี่ยง AI เพื่อสนับสนุนการนำไปใช้

5. การเลือกหน่วยรับรอง (Choosing a Certification Body)

การเลือกหน่วยรับรองที่เหมาะสมช่วยให้กระบวนการรับรองราบรื่นและน่าเชื่อถือ

- **เกณฑ์การเลือก**: เลือกหน่วยที่ได้รับการรับรองจาก UKAS (United Kingdom Accreditation Service) หรือหน่วยงานเทียบเท่า มีประสบการณ์ใน ISO 42001 และเข้าใจบริบทอุตสาหกรรม ตัวอย่าง: Perry Johnson Registrars (PJR) ซึ่งร่วมงานกับ Blackmores ใน webinar กับ Melanie
- **ขั้นตอน**: เปรียบเทียบหน่วยรับรองหลายแห่ง พิจารณาค่าใช้จ่าย ความเชี่ยวชาญ และเวลาในการตรวจสอบ Blackmores แนะนำให้เลือกหน่วยที่ให้การสนับสนุนต่อเนื่องหลังรับรอง
- **ประโยชน์**: รับรองว่าองค์กรได้รับการตรวจสอบอย่างเป็นกลางและสอดคล้องกับมาตรฐานสากล ช่วยเพิ่มความน่าเชื่อถือในตลาด

วิธีการนำไปใช้ตาม Isology® 7 ขั้นตอน

Blackmores ใช้วิธีการ Isology® ซึ่งเป็น 7 ขั้นตอนในการนำ ISO 42001 ไปใช้ โดยบูรณาการกับการวางแผนข้างต้น:

1. **Gap Analysis**: ตรวจสอบช่องว่างระบบปัจจุบัน
2. **Planning**: กำหนดขอบเขต ระยะเวลา และนโยบาย
3. **Implementation**: พัฒนาเอกสารและกระบวนการ เช่น การประเมินผลกระทบ AI
4. **Training**: ฝึกอบรมผ่าน Isology Hub
5. **Internal Audit**: ตรวจสอบภายในเพื่อปรับปรุง
6. **Certification**: เตรียมรับรองจากหน่วยภายนอก
7. **Continual Improvement**: ทบทวนและปรับปรุงระบบอย่างต่อเนื่อง

วิธีการนี้ช่วยให้องค์กรรับรองได้ 100% ตามสถิติของ Blackmores

ความท้าทายและคำแนะนำ

- **ความท้าทาย**: การขาดความเข้าใจใน AI การขาดทรัพยากร และการเปลี่ยนแปลงกฎระเบียบรวดเร็ว
- **คำแนะนำ**: เริ่มด้วยโครงการนำร่อง ใช้ที่ปรึกษาเช่น Blackmores และติดตามผลกระทบอย่างต่อเนื่อง

สรุปและแนวโน้ม

เครื่องมือปฏิบัติสำหรับการนำ AIMS ไปใช้ โดยเน้นการวางแผนที่รอบคอบเพื่อความสำเร็จ ในปี 2568 การนำ ISO 42001 ไปใช้จะเพิ่มขึ้นตามการเติบโตของ AI โดย Blackmores คาดว่าองค์กรในเอเชียจะนำไปใช้มากขึ้นเพื่อแข่งขันระดับโลก

Session 13: GenAI - Overview, Practical Implementation, Success Factors and Case Study

1. วิวัฒนาการของ AI (AI Evolution)

ส่วนนี้ให้ภาพรวมประวัติศาสตร์และพัฒนาการของ AI โดยเน้นการเปลี่ยนผ่านจาก AI แบบดั้งเดิมสู่ GenAI เพื่อให้ผู้เข้าร่วมเข้าใจบริบทและศักยภาพของเทคโนโลยีนี้

ประวัติศาสตร์และพัฒนาการหลัก

- ยุคเริ่มต้น (1950s-1980s): AI เริ่มจากระบบที่ใช้กฎหาก (rule-based systems) เช่น ระบบผู้เชี่ยวชาญ (expert systems) ที่ทำงานตามตรรกะที่มนุษย์กำหนดล่วงหน้า ตัวอย่าง: Deep Blue ของ IBM ที่เอาชนะหมากรุกในปี 1997 โดยใช้การคำนวณแบบ brute force ยุคนี้มุ่งเน้น Narrow AI ที่แก้ปัญหาเฉพาะเจาะจง แต่ขาดความยืดหยุ่น
- ยุค Machine Learning (1990s-2010s): การเปลี่ยนสู่ ML ที่เรียนรู้จากข้อมูลแทนการโปรแกรมชัดเจน เช่น Supervised Learning (ใช้ข้อมูลที่มีป้ายกำกับ) และ Unsupervised Learning (ค้นหารูปแบบเอง) พัฒนาการสำคัญ: การใช้ Neural Networks และ Big Data ทำให้ AI สามารถจัดการงานซับซ้อน เช่น การจดจำภาพ (Computer Vision) และการประมวลผลภาษา (NLP)
- ยุค Deep Learning และ GenAI (2010s-ปัจจุบัน): GenAI เป็นส่วนย่อยของ Deep Learning ที่ใช้เครือข่ายประสาทลึก (Deep Neural Networks) เช่น Generative Adversarial Networks (GANs) และ Transformers

เพื่อสร้างเนื้อหาใหม่ เช่น ข้อความ ภาพ หรือเสียง พัฒนาการสำคัญ: โมเดลอย่าง GPT (Generative Pre-trained Transformer) ในปี 2018 ทำให้ GenAI เข้าถึงได้ง่ายขึ้นผ่านแพลตฟอร์มอย่าง ChatGPT (2022) ซึ่งใช้ Large Language Models (LLMs) ที่ฝึกจากข้อมูลมหาศาล

- แนวโน้มในปี 2568 (2025): GenAI กำลังบูรณาการกับ Multimodal AI (จัดการข้อมูลหลายรูปแบบ เช่น ข้อความ+ภาพ) และ AI ที่รับผิดชอบ (Responsible AI) เพื่อลดปัญหาอคติและ hallucination (ข้อมูลเท็จ) โดยคาดว่าตลาด GenAI จะเติบโตเกิน 1 ล้านล้านดอลลาร์สหรัฐภายในปี 2573 (2030)

ความแตกต่างระหว่าง AI ดั้งเดิมและ GenAI

ลักษณะ	AI ดั้งเดิม	GenAI
การทำงาน	วิเคราะห์และพยากรณ์จากข้อมูลที่มี	สร้างเนื้อหาใหม่จากรูปแบบที่เรียนรู้
ตัวอย่าง	การจำแนกภาพ (Classification)	การสร้างภาพจากข้อความ (Text-to-Image)
ข้อดี	แม่นยำในงานเฉพาะ	สร้างสรรค์และปรับแต่งได้สูง
ข้อจำกัด	ขาดความยืดหยุ่น	เสี่ยงข้อมูลเท็จและด้านจริยธรรม

วิวัฒนาการนี้ช่วยให้ AI จากเครื่องมือช่วยเหลือกลายเป็นเครื่องมือสร้างสรรค์ โดยเฉพาะในอุตสาหกรรมที่ต้องการ compliance เช่น การแพทย์และการเงิน

2. GenAI ในด้านการจัดการ Compliance (GenAI in Compliance Management)

ส่วนนี้สำรวจการนำ GenAI มาใช้ในการจัดการ compliance ซึ่งหมายถึงการปฏิบัติตามกฎระเบียบและมาตรฐานภายในองค์กร โดย GenAI ช่วยเพิ่มประสิทธิภาพ ลดต้นทุน และลดข้อผิดพลาดจากมนุษย์

การนำ GenAI ไปใช้ใน Compliance

- การสรุปและวิเคราะห์เอกสาร: GenAI สามารถสรุปรายงาน compliance ยาวๆ ให้สั้นและเข้าใจง่าย เช่น การใช้ NLP เพื่อดึงข้อมูลสำคัญจากเอกสารกฎหมายหรือรายงานการตรวจสอบ ตัวอย่าง: ในอุปกรณ์ทางการแพทย์ GenAI สามารถตรวจสอบเอกสารตาม EU MDR/IVDR เพื่อระบุข้อกำหนดที่ขาดหาย
- การตรวจสอบและการตรวจจับความผิดปกติ: ใช้ GenAI เพื่อตรวจสอบข้อมูลอัตโนมัติ เช่น การตรวจหาการละเมิด GDPR ในข้อมูลส่วนบุคคล หรือการพยากรณ์ความเสี่ยง compliance จากข้อมูลประวัติศาสตร์ ซึ่งลดเวลาจากวันเป็นชั่วโมง
- การสร้างเนื้อหา Compliance: GenAI สร้างเทมเพลตเอกสาร เช่น นโยบาย AI Policy ตาม ISO 42001 หรือรายงานผลกระทบ (Impact Assessment) โดยปรับแต่งตามบริบทองค์กร
- การจัดการความเสี่ยง: บูรณาการกับระบบ AIMS เพื่อประเมินอคติและความเป็นธรรม เช่น การใช้ GenAI เพื่อสร้างข้อมูลสังเคราะห์ (synthetic data) สำหรับทดสอบโมเดลโดยไม่ละเมิดความเป็นส่วนตัว

ประโยชน์และความท้าทาย

- ประโยชน์: เพิ่มประสิทธิภาพ 50-70% ในกระบวนการ compliance ลดข้อผิดพลาด และช่วยให้องค์กรปฏิบัติตามกฎระเบียบระดับโลกได้รวดเร็วขึ้น
- ความท้าทาย: เสี่ยง hallucination (ข้อมูลเท็จ) ดังนั้นต้องมี human oversight และ validation เพื่อยืนยันความถูกต้อง โดยเฉพาะในด้านกฎหมายที่ต้องการความแม่นยำสูง

ตัวอย่าง: บริษัท Star ใช้ GenAI เพื่อตรวจสอบเอกสาร cybersecurity ตาม ISO 27001 ทำให้ลดเวลาการตรวจสอบจาก 2 สัปดาห์เหลือ 2 วัน

3. การสาธิต (Demo Time)

ส่วนนี้เป็นการสาธิตการใช้งาน GenAI จริง เพื่อให้ผู้เข้าร่วมเห็นภาพปฏิบัติ โดยใช้เครื่องมืออย่าง Grok หรือ ChatGPT เพื่อแสดงศักยภาพใน compliance

ตัวอย่างการสาธิต

- เดโม 1: การสรุปเอกสาร: ป้อนรายงาน ISO 13485 ยาว 50 หน้าให้ GenAI สรุปเป็น bullet points หลัก โดยเน้นข้อกำหนดที่เกี่ยวข้องกับ AI/ML ผลลัพธ์: สรุปสั้นๆ ภายใน 30 วินาที พร้อมลิงก์อ้างอิงหน้าเอกสาร
- เดโม 2: การตรวจสอบ Compliance: ใช้ GenAI เพื่อตรวจหาอคติในชุดข้อมูล เช่น การวิเคราะห์ข้อมูลผู้ป่วยเพื่อตรวจสอบความเป็นธรรมตามเพศและเชื้อชาติ ผลลัพธ์: รายงาน metric เช่น Demographic Parity และแนะนำการปรับข้อมูล
- เดโม 3: การสร้างเนื้อหา: สั่ง GenAI สร้างนโยบาย compliance สำหรับระบบ AI ในอุปกรณ์ทางการแพทย์ โดยรวมข้อกำหนดจาก FDA และ EU MDR ผลลัพธ์: เทมเพลตพร้อมใช้งานที่ปรับแต่งได้
- เคล็ดลับในการใช้งาน: ใช้ prompt ที่ชัดเจน เช่น "Summarize the key compliance requirements from this document, focusing on AI risks" เพื่อให้ผลลัพธ์แม่นยำ และตรวจสอบด้วยมนุษย์เสมอเพื่อหลีกเลี่ยงข้อผิดพลาด

การสาธิตนี้เน้นการใช้งานจริงในสภาพแวดล้อมที่ควบคุม เพื่อแสดงว่าอย่างไร GenAI สามารถบูรณาการกับกระบวนการ compliance ได้อย่างปลอดภัย

4. วิธีการประสบความสำเร็จ (How to Succeed)

ส่วนนี้สรุปปัจจัยความสำเร็จในการนำ GenAI ไปใช้ พร้อมกรณีศึกษา เพื่อให้ผู้เข้าร่วมสามารถนำไปปฏิบัติได้

ปัจจัยความสำเร็จ

- การวางแผนและกำกับดูแล: สร้างนโยบาย GenAI ที่ชัดเจน โดยบูรณาการกับ AIMS ตาม ISO 42001 รวมการประเมินผลกระทบ (Impact Assessment) และ human oversight เพื่อจัดการความเสี่ยง
- คุณภาพข้อมูลและการฝึกอบรม: ใช้ข้อมูลคุณภาพสูงและฝึกพนักงานให้เข้าใจการใช้งาน GenAI อย่างมีจริยธรรม เพื่อลดปัญหาอคติและ hallucination
- การทดสอบและติดตาม: ทดสอบ GenAI ในโครงการนำร่องก่อนนำไปใช้จริง และติดตามประสิทธิภาพอย่างต่อเนื่อง โดยใช้ metric เช่น Accuracy, Fairness Score และ User Satisfaction
- การบูรณาการกับระบบที่มีอยู่: เชื่อม GenAI กับเครื่องมือ compliance เช่น ซอฟต์แวร์ตรวจสอบเอกสาร เพื่อเพิ่มประสิทธิภาพ
- ด้านจริยธรรมและกฎระเบียบ: ปฏิบัติตามหลัก OECD AI Principles และ EU AI Act โดยเน้นความโปร่งใสและความรับผิดชอบ

กรณีศึกษา

- กรณีศึกษา 1: การใช้ GenAI ในบริษัทผลิตอุปกรณ์ทางการแพทย์: บริษัท Star นำ GenAI มาใช้ในการสรุปรายงานการตรวจสอบตาม FDA ทำให้ลดเวลาจาก 1 สัปดาห์เหลือ 1 วัน ผลลัพธ์: เพิ่มประสิทธิภาพ 80% และลดข้อผิดพลาด 50% โดยมี human review เพื่อยืนยันความถูกต้อง
- กรณีศึกษา 2: การจัดการ Compliance ในอุตสาหกรรมการเงิน: ธนาคารใช้ GenAI เพื่อตรวจหาการฉ้อโกงในเอกสาร โดยสร้างรายงานอัตโนมัติตาม GDPR ผลลัพธ์: ลดความเสี่ยงการละเมิด 40% และประหยัดต้นทุน 30% แต่ต้องแก้ปัญหา hallucination ด้วยการฝึกโมเดลเฉพาะโดเมน
- บทเรียนที่ได้: ความสำเร็จมาจากการเริ่มต้นเล็กๆ (pilot project) การมีทีมข้ามสายงาน และการปรับปรุงต่อเนื่องเพื่อจัดการความท้าทาย

สรุป

ภาพรวมครบถ้วนเกี่ยวกับ GenAI โดยเน้นวิวัฒนาการ การนำไปใช้ใน compliance การสาธิตจริง และปัจจัยความสำเร็จ เพื่อให้องค์กรสามารถนำเทคโนโลยีนี้ไปใช้อย่างมีประสิทธิภาพและรับผิดชอบ ในบริบทปี 2568 GenAI จะเป็นเครื่องมือหลักในการจัดการ compliance แต่ต้องมีมาตรการป้องกันความเสี่ยงเพื่อความยั่งยืน

Session 14: Gen AI Adoption Strategy

วัตถุประสงค์

วัตถุประสงค์หลักสามประการ เพื่อให้ผู้เข้าร่วมสามารถนำความรู้ไปประยุกต์ใช้ในโครงการ GenAI ของตนเอง:

1. สำรวจกลยุทธ์การนำ GenAI มาใช้จากโครงการนำร่องและทดลอง (pilots/trials) ที่ดำเนินการโดยหน่วยงานรัฐบาลและสถาบันต่างๆ เพื่อเข้าใจแนวทางปฏิบัติจริงและบทเรียนที่ได้
2. ระบุประโยชน์และความท้าทายที่อาจเกี่ยวข้องกับโครงการ GenAI ของตนเอง โดยพิจารณาถึงบริบทองค์กร เช่น ความพร้อมด้านเทคโนโลยีและบุคลากร
3. เรียนรู้ขั้นตอนและการกระทำที่แนะนำเพื่อเสริมสร้างกลยุทธ์การนำ GenAI มาใช้ โดยเน้นการจัดการความเสี่ยงและการเพิ่มประสิทธิภาพ

โครงร่างเนื้อหา

โครงร่างเนื้อหาแบ่งเป็นสามส่วนหลัก เพื่อให้ครอบคลุมกระบวนการนำ GenAI มาใช้ตั้งแต่ต้นจนจบ:

- Generative AI Pilots and Trials: ภาพรวมโครงการนำร่องจากประเทศต่างๆ
- Findings from Pilots: ผลลัพธ์และข้อมูลเชิงลึกจากโครงการ
- Themed Recommendations for GenAI Strategy: คำแนะนำเชิงธีมสำหรับกลยุทธ์

1. โครงการนำร่องและทดลอง Generative AI (Generative AI Pilots and Trials)

ส่วนนี้ให้ภาพรวมโครงการนำร่อง GenAI จากหน่วยงานรัฐบาลและสถาบันต่างๆ เพื่อแสดงตัวอย่างการนำไปใช้จริง โดยเน้นโครงการที่มุ่งเพิ่มประสิทธิภาพและจัดการความเสี่ยง

โครงการหลัก

- Australian Government – Microsoft 365 Copilot: ทดลอง 6 เดือนใน 56 หน่วยงาน มีผู้เข้าร่วม 2,000 คน ใช้สำหรับสรุปข้อมูลและเขียนเนื้อหา เช่น การสรุปรายงานประชุมและร่างเอกสารนโยบาย ประหยัดเวลาเฉลี่ย 1 ชั่วโมงต่อวัน แต่พบปัญหาความถูกต้องในข้อมูลซับซ้อน
- US General Services Administration (GSA) – Gemini: ทดลอง 90 วันกับพนักงาน 250 คน ใช้สำหรับงานขนาดเล็ก เช่น การเขียนโค้ด การสรุปเอกสาร และการ brainstorm แยกงานซับซ้อนเพื่อประสิทธิภาพสูงสุด พบว่าช่วยลดเวลาในงาน routine แต่ต้องมี human review เพื่อหลีกเลี่ยงข้อผิดพลาด
- UK Government Digital Services (GDS) – GOV.UK Chat: ทดลองหลายเฟสกับผู้ใช้ 1,000 คน ใช้ตอบคำถามภาษาธรรมชาติจากประชาชน เช่น การให้ข้อมูลบริการรัฐ แต่พบปัญหาความยาวเนื้อหาและ hallucination (ข้อมูลเท็จ) ทำให้ต้องปรับ prompt และเพิ่ม feedback loop
- Pennsylvania – ChatGPT: ทดลอง 12 เดือนกับพนักงาน 175 คน ใช้สำหรับวิจัย Brainstorm และร่างเอกสาร เช่น การร่างรายงานชุมชน ประหยัดเวลา 95 นาทีต่อวัน โดย 85% ของผู้ใช้มีประสบการณ์บวก แต่กังวลเรื่องความถูกต้องและ overconfidence ในผลลัพธ์

โครงการเหล่านี้เน้นการนำ GenAI มาใช้ในงานรัฐบาลเพื่อเพิ่ม productivity โดยเริ่มจาก scale เล็กเพื่อทดสอบก่อนขยาย

2. ผลลัพธ์จากโครงการนำร่อง (Findings from Pilots)

ส่วนนี้สรุปข้อมูลเชิงลึกจากโครงการ โดยแบ่งเป็นด้านประสิทธิภาพและประสบการณ์ผู้ใช้ เพื่อให้เห็นประโยชน์และความท้าทาย

ประสิทธิภาพและประสิทธิผล (Productivity and Efficiency)

- GenAI ช่วยประหยัดเวลาและลดงานซ้ำซาก เช่น ใน Australian Pilot ลดเวลาเขียนเอกสาร 50-70% แต่ต้องตรวจสอบความถูกต้องเพื่อหลีกเลี่ยงข้อมูลผิดพลาด

- ใน US GSA พบว่าการแยกงานซับซ้อนเป็นส่วนย่อยช่วยเพิ่มประสิทธิภาพ แต่ GenAI เหมาะกับงาน routine มากกว่างานเชิงกลยุทธ์
- โดยรวม: ประหยัดเวลาเฉลี่ย 1-2 ชั่วโมงต่อวัน แต่ต้องมีมาตรการป้องกัน over-reliance

ประสบการณ์ผู้ใช้ (User Experience)

- ผู้ใช้ส่วนใหญ่พอใจกับความสะดวก แต่กังวลเรื่องความถูกต้อง เช่น ใน UK GDS ผู้ใช้ 70% ชอบ interface แต่ 40% พบ hallucination
- ใน Pennsylvania 85% มีประสบการณ์บวก แต่พบปัญหา overconfidence (เชื่อผลลัพธ์มากเกินไป) ทำให้ต้องฝึกอบรมการตรวจสอบ
- โดยรวม: GenAI เพิ่ม engagement แต่ต้องจัดการความคาดหวังและให้ feedback เพื่อปรับปรุง

3. คำแนะนำเชิงธีมสำหรับกลยุทธ์ GenAI (Themed Recommendations for GenAI Strategy)

ส่วนนี้ให้คำแนะนำที่จัดกลุ่มตาม Theme เพื่อช่วยองค์กรพัฒนากลยุทธ์ โดยอ้างอิงจากบทเรียนจาก pilots

Theme 1: กลยุทธ์ การกำกับดูแล และนโยบาย (Strategy, Governance and Policy)

- เลือกผลิตภัณฑ์ GenAI ที่เหมาะสมกับบริบท เช่น Cloud-based สำหรับ scalability แต่ต้องพิจารณาความมั่นคง
- กำหนดนโยบายชัดเจน เช่น ห้ามใช้ GenAI ในข้อมูลที่อ่อนไหว และวางแผนยืดหยุ่นเพื่อปรับตามผล feedback
- ร่วมมือกับหน่วยงานภายนอก เช่น ผู้ให้บริการ AI เพื่อแบ่งปัน best practices

Theme 2: การฝึกอบรม ทักษะ และการเปลี่ยนแปลง (Training, Skills and Change)

- ฝึกอบรมเฉพาะบุคคล เช่น Workshop สำหรับ prompt engineering และการตรวจสอบผลลัพธ์
- จัดการเปลี่ยนแปลงด้วยการส่งเสริมการเรียนรู้จากเพื่อน (peer learning) และสร้าง community of practice

Theme 3: การกำกับดูแลโดยมนุษย์ ความถูกต้อง และความเชื่อถือ (Human Oversight, Accuracy and Trust)

- ใช้ human oversight สำหรับผลลัพธ์สำคัญ เช่น Review ก่อนเผยแพร่เอกสาร
- สร้างความเชื่อถือด้วยการตรวจสอบและปรับโมเดล เช่น ใช้ watermarking สำหรับเนื้อหาที่สร้างโดย AI

Theme 4: การออกแบบเน้นผู้ใช้และการทำซ้ำ (User-Centric Design and Iteration)

- มีส่วนร่วมผู้ใช้ตั้งแต่ต้นเพื่อรวบรวม feedback และปรับปรุงตาม loop
- ใช้ user-centric approach เช่น A/B testing เพื่อเปรียบเทียบเวอร์ชัน

Theme 5: การระบุกรณีใช้งานและการตระหนักถึงคุณค่า (Use Case Identification and Value Realisation)

- ระบุกรณีใช้งานที่มีมูลค่าสูง เช่น การสรุปข้อมูลในรัฐบาล และแบ่งปันแนวปฏิบัติที่ดี
- ส่งเสริม innovation โดยให้พนักงานทดลอง GenAI ในงานประจำ

Theme 6: การวัดผลกระทบและการติดตาม (Impact Measurement and Monitoring)

- ประเมินระยะยาวด้วยวิธีผสม เช่น Survey และ metrics (เวลา saved, accuracy rate)
- ติดตามประสิทธิภาพต่อเนื่องและรายงานผลโปร่งใสเพื่อปรับปรุง

คำแนะนำเหล่านี้ช่วยให้องค์กรนำ GenAI มาใช้อย่างยั่งยืน โดยลดความเสี่ยงและเพิ่มประโยชน์

สรุป

ภาพรวมครบถ้วนเกี่ยวกับกลยุทธ์การนำ GenAI มาใช้ โดยเน้นบทเรียนจาก pilots เพื่อให้องค์กรสามารถพัฒนากลยุทธ์ที่เหมาะสม ในบริบทเวียดนาม การนำ GenAI มาใช้จะช่วยเพิ่ม productivity แต่ต้องมี governance ที่แข็งแกร่งตาม DARTS Framework

Session 15: Workshop - Develop a Roadmap for AI Governance in Your Organization

ผู้เข้าร่วมได้นำความรู้ที่ได้รับตลอดหลักสูตรมาประยุกต์ใช้ในการออกแบบแผนงานเฉพาะสำหรับการกำกับดูแล AI ในองค์กรของตน ผู้เข้าร่วมได้ทำงานเป็นกลุ่ม โดยระบุประเด็นสำคัญ กำหนดเป้าหมายที่สามารถนำไปปฏิบัติได้จริง และร่างขั้นตอนปฏิบัติเพื่อนำหลักการและแนวปฏิบัติที่ดีที่สุดของ ISO 42001 มาใช้ เวิร์กช็อปนี้เน้นการแปลงแนวคิดให้เป็นกลยุทธ์ที่เป็นรูปธรรม ช่วยให้ผู้เข้าร่วมได้วางแผนที่ชัดเจนเพื่อเสริมสร้างการบริหารจัดการและการกำกับดูแล AI อย่างมีความรับผิดชอบ

Session 16: Presentations & Discussions

ผู้เข้าร่วมได้นำเสนอแผนงานการกำกับดูแล AI ขององค์กรที่พัฒนาขึ้นระหว่างการอบรม ผู้เข้าร่วมจะได้รับคำติชมเชิงสร้างสรรค์จากเพื่อนร่วมงานและวิทยากร ซึ่งส่งเสริมการเรียนรู้ร่วมกันและการพัฒนาแนวคิด การนำเสนอครั้งนี้ส่งเสริมการอภิปรายอย่างเปิดกว้าง การแบ่งปันความรู้ และข้อมูลเชิงลึกจากหลากหลายอุตสาหกรรม เพื่อเสริมสร้างกลยุทธ์ของผู้เข้าร่วมในการนำการจัดการ AI อย่างมีความรับผิดชอบและมีประสิทธิภาพมาใช้ **โดยใช้ Case study Sessions 10 ตามเดิม และให้ประเมิน AI System Impact Assessment** โดยมีประบวนการต่าง ๆ ดังต่อไปนี้

เพื่อวิเคราะห์ข้อมูลและพยากรณ์อาชญากรรม ซึ่งสะท้อนถึงความพยายามของ MetroCity ในการเปลี่ยนจากการเป็นตำรวจแบบตอบสนอง (Reactive) ไปสู่การป้องกันอาชญากรรมล่วงหน้า (Proactive) อย่างไรก็ตาม การนำมาใช้ต้องพิจารณาความเสี่ยง เช่น อคติในข้อมูลและความไม่เท่าเทียมทางสังคม รายงานนี้จะบรรยายการประเมินผลกระทบตามโครงสร้างของเทมเพลต โดยเชื่อมโยงกับกรอบงาน DARTS และมาตรฐาน ISO 42001 เพื่อให้สอดคล้องกับบริบทของเวียดนาม

ประเมินผลกระทบระบบ AI

โดยมีวัตถุประสงค์ของการประเมินคือเพื่อตรวจสอบผลกระทบของระบบ AI ตลอดวงจรชีวิต (Lifecycle) และระบุการรักษาความเสี่ยงก่อนนำไปใช้งาน การทบทวนจะเกิดขึ้นหากมีการเปลี่ยนแปลงในวัตถุประสงค์ ระดับความซับซ้อนของเทคโนโลยี หรือแหล่งข้อมูล

คำอธิบายระบบ AI MetroCity

ระบบ AI ที่เสนอนี้ถูกออกแบบมาเพื่อช่วยกรมตำรวจ MetroCity จัดการทรัพยากรที่มีจำกัดและตอบสนองต่อความต้องการด้านความมั่นคงสาธารณะที่เพิ่มขึ้น หลักการพื้นฐานของระบบคือการใช้โมเดล AI วิเคราะห์ข้อมูลจากแหล่งต่างๆ เพื่อพยากรณ์สถานที่และเวลาที่อาจเกิดอาชญากรรม โดยเปลี่ยนวิธีการทำงานจากแบบตอบสนองไปเป็นแบบป้องกันล่วงหน้า ผลลัพธ์ที่คาดหวังรวมถึงการวิเคราะห์ข้อมูลอาชญากรรมในอดีต การรวมข้อมูลทางเศรษฐกิจและประชากรศาสตร์ การสร้างแผนที่ความเสี่ยง และการแนะนำลำดับความสำคัญในการลาดตระเวนสำหรับแต่ละกะ เป้าหมายหลักคือการเพิ่มประสิทธิภาพการใช้ทรัพยากรตำรวจและลดอัตราการเกิดอาชญากรรมในเมือง

การระบุผู้มีส่วนได้เสีย

ผู้มีส่วนได้เสียในกรณีนี้ประกอบด้วย ชุมชนที่ได้รับผลกระทบจากการจัดสรรตำรวจ ตำรวจที่ปฏิบัติงาน ผู้นำชุมชนที่ดูแลความเชื่อมั่น และหน่วยงานกำกับดูแลที่รับผิดชอบด้านความปลอดภัยสาธารณะ การมีส่วนร่วมของผู้มีส่วนได้เสียสำคัญมากในการประเมินผลกระทบ เช่น การปรึกษาหารือกับชุมชนเพื่อตรวจสอบความเป็นธรรม และการรายงานต่อหน่วยงานกำกับดูแลเพื่อสร้างความโปร่งใส

การประเมินความเสี่ยง

การประเมินความเสี่ยงดำเนินการผ่านคำถามที่ต้องตอบด้วย "ใช่" หรือ "ไม่" พร้อมระบุระดับความเสี่ยงและการรักษาความเสี่ยงหากจำเป็น ประเด็นที่ต้องพิจารณารวมถึง:

- ระบบอาจเสริมสร้างความไม่เท่าเทียมหากฝึกจากข้อมูลประวัติที่มีอคติ — ตอบ "ใช่" ระดับความเสี่ยงสูง ไม่ยอมรับ ต้องตรวจสอบและปรับข้อมูล

- ควรแจ้งสาธารณชนหากย่านของตนถูกระบุเป็นพื้นที่เสี่ยงสูง — ตอบ "ใช่" ระดับความเสี่ยงปานกลาง ยอมรับด้วย แผนสื่อสาร
- ผู้รับผิดชอบหากการจัดสรรตำรวจนำไปสู่ความไม่เชื่อมั่นในชุมชน — ตอบ "ใช่" ระดับความเสี่ยงสูง ไม่ยอมรับ ต้องกำหนดกรอบความรับผิดชอบ

  ตารางสรุปความเสี่ยงและการรักษา เพื่อให้เห็นภาพชัดเจนยิ่งขึ้น ต่อไปนี้เป็นตารางสรุปความเสี่ยงและการรักษาที่ระบุ:

ประเด็นความเสี่ยง	ผลลัพธ์	ระดับความเสี่ยง	ยอมรับ (Y/N)	การรักษาความเสี่ยง / หมายเหตุ
ความไม่เท่าเทียมจากข้อมูลอคติ	ใช่	สูง	ไม่	ตรวจสอบและปรับข้อมูลโดยทีม IT
การแจ้งสาธารณชนเกี่ยวกับพื้นที่เสี่ยง	ใช่	ปานกลาง	ใช่	แผนสื่อสารกับชุมชน
ความรับผิดชอบต่อความไม่เชื่อมั่น	ใช่	สูง	ไม่	กำหนดกรอบความรับผิดชอบ

การรักษาความเสี่ยง

การรักษาความเสี่ยงที่ระบุไว้รวมถึงการตรวจสอบข้อมูลเพื่อลดอคติ การสื่อสารกับชุมชนเพื่อสร้างความโปร่งใส การกำหนดกรอบความรับผิดชอบ และการติดตามผลลัพธ์อย่างต่อเนื่อง โดยมีกำหนดการและผู้รับผิดชอบชัดเจน เช่น การตรวจสอบข้อมูลภายในไตรมาสแรกของปี 2569 โดยทีม IT ของตำรวจ

คำถามด้านเทคนิคและความปลอดภัย

การประเมินด้านเทคนิคและความปลอดภัยครอบคลุมการเข้าถึงระบบและมาตรการรักษาความมั่นคง เช่น การใช้การตรวจสอบสิทธิ์ที่เข้มงวด (Yes, ระดับต่ำ) และการเข้ารหัสฐานข้อมูล (Yes, ระดับต่ำ) อย่างไรก็ตาม การตรวจสอบความปลอดภัยแบบเรียลไทม์ (Yes, ระดับปานกลาง) และการอัปเดตระบบอัตโนมัติ (Yes, ระดับปานกลาง) ยังต้องมีการปรับปรุงเพิ่มเติม

สรุปและแนวโน้ม

การประเมินนี้ช่วยให้กรมตำรวจ MetroCity เข้าใจผลกระทบของระบบ AI ในระบบการตรวจตราเชิงคาดการณ์แบบพยากรณ์ โดยระบุความเสี่ยงหลัก เช่น อคติและความไม่โปร่งใส พร้อมกำหนดมาตรการรักษาที่เหมาะสม ในบริบทของเวียดนาม การนำระบบนี้ไปใช้จะต้องสอดคล้องกับกฎหมายข้อมูลส่วนบุคคลและมาตรฐาน ISO 42001 เพื่อให้เกิดประโยชน์สูงสุดและลดผลกระทบต่อสังคม แนวโน้มในอนาคตคือการขยายการใช้ AI ในภาคความมั่นคงสาธารณะควบคู่กับการพัฒนากรอบการกำกับดูแลที่เข้มแข็ง

ส่วนที่ 2 ประโยชน์ที่ได้รับและการขยายผลจากการเข้าร่วมโครงการ

โครงการที่มุ่งพัฒนาความรู้พื้นฐานเกี่ยวกับปัญญาประดิษฐ์ (AI) การจัดการระบบ AI (AIMS) ตามมาตรฐาน ISO 42001 การกำกับดูแล AI ที่รับผิดชอบ และการนำ Generative AI (GenAI) มาใช้อย่างมีประสิทธิภาพ โดยครอบคลุม session ต่างๆ เช่น การแนะนำ AI และ ML, ภาพรวม ISO 42001, การจัดการข้อมูลและอคติ, การประเมินผลกระทบระบบ AI, การออกแบบกรอบงานกำกับดูแล AI, การนำ AIMS ไปปฏิบัติ และกลยุทธ์ GenAI การเข้าร่วมโครงการนี้ช่วยให้ผู้เข้าร่วมได้รับความรู้ที่ครอบคลุมและนำไปประยุกต์ใช้ได้จริง ต่อไปนี้คือรายละเอียดประโยชน์ที่ได้รับและการขยายผล

ประโยชน์ที่ได้รับจากการเข้าร่วมโครงการ

- **ประโยชน์ต่อตนเอง:**

- เพิ่มพูนความรู้และทักษะด้าน AI อย่างครอบคลุม เช่น ความเข้าใจเกี่ยวกับวิวัฒนาการ AI, การจัดการข้อมูล, การลดอคติในระบบ AI, และกลยุทธ์การนำ GenAI มาใช้ ทำให้สามารถนำไปประยุกต์ในงานส่วนตัว เช่น การวิเคราะห์ข้อมูลในโครงการส่วนบุคคลหรือการพัฒนาโซลูชัน AI อีกทั้งช่วยเสริมสร้าง Mindset ที่รับผิดชอบต่อการใช้งาน AI ทำให้ตัดสินใจได้อย่างมีจริยธรรมในชีวิตประจำวัน

- **ประโยชน์ต่อหน่วยงานต้นสังกัด ฝ่ายพัฒนาศักยภาพ:**
  - หน่วยงานได้มีบุคลากรที่มีศักยภาพในการจัดการ AI ทำให้สามารถนำความรู้จากโครงการไปพัฒนาระบบภายใน เช่น การนำ AIMS ตาม ISO 42001 เพิ่มประสิทธิภาพองค์กรโดยรวม เช่น ลดต้นทุนการตรวจสอบ Compliance 20-30% ผ่าน GenAI และช่วยให้หน่วยงานเป็นผู้นำในด้านนวัตกรรม AI นอกจากนี้ ยังช่วยให้หน่วยงานปฏิบัติตามกฎระเบียบระดับโลก เช่น EU AI Act ได้ดีขึ้น ลดความเสี่ยงทางกฎหมาย
- **ประโยชน์ต่อสายงานการฝึกอบรม:**
  - สายงานการฝึกอบรมได้รับเนื้อหาและเครื่องมือใหม่ๆ เช่น กรอบงาน DARTS สำหรับการกำกับดูแล AI และตัวอย่าง pilots ของ GenAI ทำให้สามารถออกแบบหลักสูตรฝึกอบรมภายในที่ทันสมัยและสอดคล้องกับมาตรฐานสากล เพิ่มคุณภาพการฝึกอบรมโดยรวม เช่น การนำกรณีศึกษามาใช้ในการสอน ทำให้ผู้เรียนมีส่วนร่วมมากขึ้นและเข้าใจปฏิบัติจริง

กิจกรรมการขยายผลที่ได้ดำเนินการภายในระยะเวลา 60 วันนับจากวันสุดท้ายของโครงการ

ผู้เข้าร่วมจะดำเนินกิจกรรมขยายผลโดยการเผยแพร่รายงานสรุปนี้ สำหรับผู้ที่สนใจงานด้านการบริหารจัดการ AI อย่างเป็นระบบ มีมาตรฐาน มีการจัดการความเสี่ยง มีธรรมาภิบาล และเป็นที่ยอมรับในระดับสากล

กิจกรรมการขยายผลที่จะดำเนินการภายใน 6 เดือนหลังเข้าร่วมโครงการ

ผู้เข้าร่วมวางแผนจะดำเนินกิจกรรมขยายผล โดยอาจจะจัดโครงการฝึกอบรม AI Management Systems ตามโครงสร้างหลักสูตรนี้มาจัดอบรมในประเทศไทย ซึ่งยังไม่มีหน่วยงานใดจัดอบรมหลักสูตรแบบนี้อย่างเป็นรูปธรรม ทั้งนี้อาจเป็นช่องทางในการพัฒนาผลิตภัณฑ์ใหม่ของสถาบันฯ ต่อไป และมีรายละเอียดแผนงานดังนี้

แผนงานกิจกรรม/โครงการจัดหลักสูตร AI Management Systems

การจัดอบรมหลักสูตร AI Management Systems สำหรับผู้สนใจทั่วไป (ระยะเวลา: ปีงบประมาณ 2569)

- **รายละเอียดกิจกรรม:** จัดอบรมแบบ Hybrid (On-site&Online) ไม่ต่ำกว่า 1 รุ่น ๆ ละ 3 วัน โดยเนื้อหาครอบคลุมภาพรวม AIMS ตาม ISO 42001 การจัดการข้อมูลและอคติ การประเมินผลกระทบระบบ AI การออกแบบกรอบงานกำกับดูแล และกลยุทธ์นำ GenAI มาใช้ โดยใช้กรณีศึกษา เพื่อวิเคราะห์ปัญหาต่าง ๆ และความเสี่ยงด้านความปลอดภัย กิจกรรมรวมบรรยาย เวิร์กชอปกลุ่ม และสาธิตการใช้งาน GenAI สำหรับ Compliance management
- **ผลที่คาดว่าจะได้รับ:**

  เชิงปริมาณ: คาดการณ์ว่าจะรับผู้สมัครเข้าอบรมไม่เกินรุ่นละ 30 คน เพื่อจะได้ควบคุมคุณภาพการอบรมให้มีประสิทธิ และนำไปสู่ช่วยผลักดันให้มีบุคลากรที่มีความรู้ความสามารถด้านการจัดการ AI เพิ่มมากขึ้น

  เชิงคุณภาพ: เพิ่มความตระหนักด้าน AI Governance ในทุกภาคส่วนธุรกิจ ส่งเสริมการใช้งาน AI ที่รับผิดชอบ ลดความเสี่ยงด้านจริยธรรม (เช่น อคติในระบบ) และสร้างเครือข่ายระหว่างบริษัทเพื่อแลกเปลี่ยนประสบการณ์ ทำให้อุตสาหกรรมมีขีดความสามารถแข่งขันสูงขึ้น

ส่วนที่ 3 การศึกษาดูงาน

BSI Vietnam

BSI เวียดนาม ซึ่งเป็นสำนักงานประจำท้องถิ่นของสถาบันมาตรฐานอังกฤษ (BSI) ให้การสนับสนุนองค์กรต่างๆ ในเวียดนามผ่านมาตรฐานสากล การรับรอง การฝึกอบรม และโซลูชันดิจิทัล BSI เวียดนามมีสำนักงานในนครโฮจิมินห์ ฮานอย และดานัง เพื่อช่วยให้ธุรกิจต่างๆ พัฒนาคุณภาพ ความปลอดภัย ความยั่งยืน และความน่าเชื่อถือทางดิจิทัล บริการของเรา

ครอบคลุมมาตรฐานสำคัญๆ เช่น ISO 9001, ISO 14001, ISO/IEC 27001 และระบบการจัดการ AI ใหม่ ISO/IEC 42001 ด้วยการดำเนินงานที่เป็นกลางอย่างเคร่งครัด BSI เวียดนามไม่เพียงแต่ให้การรับรองและเสริมสร้างศักยภาพเท่านั้น แต่ยังส่งเสริมนวัตกรรมที่มีความรับผิดชอบ รวมถึงการนำ AI มาใช้อย่างมีจริยธรรมและเชื่อถือได้ ซึ่งสอดคล้องกับแนวปฏิบัติที่ดีที่สุดระดับโลก

Alibaba.com (Vietnam) Co., Ltd

บริษัท อาลีบาบาดอทคอม (เวียดนาม) จำกัด ซึ่งเป็นสาขาย่อยของอาลีบาบาดอทคอม สนับสนุน SMEs เวียดนามในการขยายการส่งออกผ่านการค้าดิจิทัลและโซลูชันอีคอมเมิร์ซ B2B ที่ปลอดภัย การดำเนินงานของบริษัทให้ความสำคัญกับความน่าเชื่อถือ ความปลอดภัยของข้อมูล และบริการที่ขับเคลื่อนด้วย AI อย่างมีความรับผิดชอบ ซึ่งสอดคล้องกับวัตถุประสงค์ของ ISO/IEC 42001 ซึ่งเป็นมาตรฐานสากลสำหรับระบบการจัดการปัญญาประดิษฐ์ (AIMS) โครงการริเริ่มของบริษัทในด้านธุรกรรมที่ปลอดภัย การกำกับดูแลข้อมูล และแพลตฟอร์มดิจิทัลที่ขับเคลื่อนด้วย AI แสดงให้เห็นถึงการประยุกต์ใช้หลักการจัดการ AI ในทางปฏิบัติ เช่น ความโปร่งใส ความรับผิดชอบ และการควบคุมความเสี่ยง ซึ่งนำเสนอข้อมูลเชิงลึกแก่ผู้เข้าร่วมเกี่ยวกับวิธีที่ผู้นำอีคอมเมิร์ซระดับโลกดำเนินการตามแนวทางปฏิบัติด้าน AI อย่างมีความรับผิดชอบ