



ETDA



AIGPC
AI GOVERNANCE
PRACTICE CENTER

(ร่าง)

ETHICAL IMPACT ASSESSMENT PLAYBOOK

คู่มือการประเมินผลกระทบทางจริยธรรม
ปัญญาประดิษฐ์



สารบัญ

สารบัญ.....	2
บทสรุปผู้บริหาร.....	1
บทนำ.....	2
บทที่ 1 ความเข้าใจพื้นฐานเกี่ยวกับ Ethical Impact Assessment (EIA).....	4
บทที่ 2 การกำหนดขอบเขตการประเมิน	14
บทที่ 3 ความปลอดภัย และ ความมั่นคงปลอดภัย.....	21
บทที่ 4 ความเป็นธรรม การไม่เลือกปฏิบัติ และ ความหลากหลาย.....	25
บทที่ 5 ความยั่งยืน.....	28
บทที่ 6 ความเป็นส่วนตัว และ การคุ้มครอง ข้อมูล.....	31
บทที่ 7 การกำกับดูแลและ การตัดสินใจโดยมนุษย์	36
บทที่ 8 ความรับผิดชอบ และ ภาระรับผิดชอบ.....	40
บทที่ 9 ความโปร่งใส ความสามารถในการอธิบายและความตระหนักรู้.....	44
บทที่ 10 การวิเคราะห์ผลกระทบทางจริยธรรม	47
บทที่ 11 การนำไปใช้ผ่านกรณีตัวอย่าง.....	56
อภิธานศัพท์.....	68
บรรณานุกรม.....	76

บทสรุปผู้บริหาร

องค์การการศึกษา วิทยาศาสตร์ และวัฒนธรรมแห่งสหประชาชาติ (United Nations Educational, Scientific and Cultural Organization หรือ UNESCO) ได้มีการเผยแพร่เครื่องมือ The UNESCO Ethical Impact Assessment (EIA) เพื่อเป็นเครื่องมือในการประเมินว่าองค์กรมีกระบวนการดำเนินงานเพื่อควบคุมให้ระบบหรือโครงการที่มีการนำ AI มาประยุกต์ใช้ สอดคล้องกับหลักการจริยธรรมปัญญาประดิษฐ์ (AI Ethical Principles) 10 มิติตามแนวทางที่กำหนดโดย UNESCO หรือไม่ และประเมินว่าการนำ AI มาประยุกต์ใช้นั้นส่งผลให้เกิดผลกระทบทั้งเชิงบวกและเชิงลบ (Positive and Negative Impact) อย่างไร เพื่อนำไปสู่การกำหนดมาตรการในการบรรเทาผลกระทบเชิงลบ และใช้เป็นข้อมูลประกอบการตัดสินใจในการดำเนินงาน

เพื่อให้องค์กรหรือบุคคลที่สนใจในประเทศไทยสามารถนำหลักการและเครื่องมือที่จัดทำโดย UNESCO มาปรับใช้ได้เหมาะสม สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (สพธอ.) จึงได้จัดทำคู่มือการประเมินผลกระทบทางจริยธรรมปัญญาประดิษฐ์ (Ethical Impact Assessment Playbook: EIA Playbook) เพื่ออธิบายวิธีการประเมินผลกระทบในรูปแบบภาษาไทย พร้อมยกตัวอย่างการประเมินผลกระทบ เพื่อเป็นแนวทางในการประเมินผลกระทบ ซึ่งประกอบด้วยเนื้อหา 3 ส่วนหลัก ได้แก่

1. การประเมินขอบเขตการนำ AI มาประยุกต์ใช้ (Scoping Questions)

ผู้ประเมินจะต้องตอบคำถามเพื่อกำหนดขอบเขต วัตถุประสงค์ และรายละเอียดทางเทคนิคที่นำมาใช้ในโครงการ ประเมินความพร้อมของกลไกการกำกับดูแล ความสมเหตุสมผลในการนำ AI มาใช้งานเพื่อป้องกันไม่ให้เกิดความเสียหาย และกำหนดรูปแบบการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย (Stakeholders) เพื่อรับความเห็นเกี่ยวกับผลกระทบที่อาจเกิดขึ้นได้อย่างครบถ้วนและถูกต้อง

2. การประเมินการดำเนินงานตามหลักการจริยธรรมปัญญาประดิษฐ์ (Principles-Based Questions)

ผู้ประเมินจะต้องตอบคำถามเพื่อประเมินการดำเนินงาน (Procedural Assessment) ในปัจจุบันและความพร้อมของมาตรการป้องกัน (Safeguard) ว่าเพียงพอและเหมาะสมต่อการบริหารจัดการผลกระทบที่อาจขัดต่อหลักการจริยธรรมปัญญาประดิษฐ์หรือไม่ โดยคำถามในส่วนนี้ครอบคลุมหลักการจริยธรรมปัญญาประดิษฐ์ 8 หลักการ และจัดเป็น 7 หมวดการประเมิน (ความปลอดภัยและความมั่นคงปลอดภัย ความเป็นธรรม ความยั่งยืน ความเป็นส่วนตัว การกำกับดูแลโดยมนุษย์ ความรับผิดชอบ และความโปร่งใส และความตระหนักรู้) เพื่อประเมินว่าการดำเนินงานในปัจจุบันสอดคล้องกับแต่ละหลักการเพียงใด

3. การประเมินผลกระทบจากการประยุกต์ใช้ AI (Impact Mapping)

ผู้ประเมินจะต้องตอบคำถามในส่วนที่ 1 และ 2 พร้อมทั้งนำผลลัพธ์ดังกล่าวมาประกอบการประเมินผลกระทบเชิงบวก (Positive Impact) และผลกระทบเชิงลบ (Negative Impact) ที่อาจเกิดขึ้นจากการนำ AI มาประยุกต์ใช้ โดยการประเมินผลกระทบเชิงบวกนั้นจะเป็นการวิเคราะห์คุณค่าและประโยชน์ที่ผู้ที่เกี่ยวข้องและสังคมจะได้รับตามหลักการในแต่ละมิติ สำหรับการประเมินผลกระทบเชิงลบนั้นจะเป็นการวิเคราะห์ผลลัพธ์ที่ขัดต่อหลักการจริยธรรมปัญญาประดิษฐ์และสร้างผลลัพธ์ที่ไม่พึงประสงค์ เพื่อกำหนดมาตรการควบคุมและบรรเทาผลกระทบ (Mitigation Measures) ที่เพียงพอและเหมาะสม เพื่อป้องกันไม่ให้เกิดความเสียหายต่อผู้ที่เกี่ยวข้องและสังคม

บทนำ

องค์การการศึกษา วิทยาศาสตร์ และวัฒนธรรมแห่งสหประชาชาติ (United Nations Educational, Scientific and Cultural Organization หรือ UNESCO) ได้ประกาศมาตรฐานด้านหลักการจริยธรรมปัญญาประดิษฐ์ “Recommendation on the Ethics of Artificial Intelligence” ซึ่งได้รับการรับรองจากประเทศสมาชิกทั้ง 194 ประเทศในปี 2021 ซึ่งมาตรฐานนี้มีวัตถุประสงค์เพื่อวางรากฐานการประยุกต์ใช้ปัญญาประดิษฐ์ (Artificial Intelligence หรือ AI) โดยคำนึงถึงการเคารพลักษณะมนุษย์ ศักดิ์ศรีความเป็นมนุษย์ เสรีภาพขั้นพื้นฐาน ความเท่าเทียม การปกป้องสิ่งแวดล้อมและวัฒนธรรม พร้อมทั้งเปิดโอกาสให้ทุกภาคส่วนมีส่วนร่วมและเข้าถึงองค์ความรู้และเทคโนโลยีได้อย่างเป็นธรรม



รูปที่ 1 หลักการจริยธรรมปัญญาประดิษฐ์ตามแนวทางของ UNESCO

เพื่อให้เกิดการนำหลักการจริยธรรมปัญญาประดิษฐ์ 10 มิติ ที่กำหนดในมาตรฐานดังกล่าวข้างต้น (ดังแสดงในรูปที่ 1) มาสู่การปฏิบัติอย่างเป็นรูปธรรม UNESCO จึงได้มีการเผยแพร่เครื่องมือ The UNESCO Ethical Impact Assessment (EIA) เพื่อเป็นเครื่องมือในการประเมินผลกระทบจากการนำ AI มาประยุกต์ใช้ ในการนี้เพื่อให้มีการนำหลักการและเครื่องมือดังกล่าวมาปรับใช้ในการกำกับดูแล AI ภายในประเทศ สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (สพธอ.) จึงได้มีการจัดทำคู่มือการประเมินผลกระทบทางจริยธรรมปัญญาประดิษฐ์ (Ethical Impact Assessment Playbook: EIA Playbook) รูปแบบภาษาไทย เพื่ออธิบายวิธีการและยกตัวอย่างการประเมินผลกระทบจากการนำ AI มาประยุกต์ใช้ในบริบทต่าง ๆ ภายใต้โครงสร้าง 3 ส่วนหลัก ได้แก่

1. **การประเมินขอบเขตของการนำ AI มาประยุกต์ใช้ (Scoping Questions)** มุ่งเน้นการวิเคราะห์เพื่อกำหนดขอบเขต วัตถุประสงค์ และรายละเอียดทางเทคนิค ประเมินความพร้อมของกลไกการกำกับดูแล และความสมเหตุสมผลในการนำ AI มาใช้งานเพื่อป้องกันไม่ให้เกิดความเสียหาย พร้อมทั้ง

กำหนดรูปแบบการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย (Stakeholders) เพื่อรับความเห็นเกี่ยวกับผลกระทบที่อาจเกิดขึ้นได้อย่างครบถ้วนและถูกต้อง

2. **การประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions)** มุ่งเน้นประเมินการดำเนินงาน (Procedural Assessment) ในปัจจุบันและความพร้อมของมาตรการป้องกัน (Safeguard) ว่ามีความเหมาะสม เพียงพอ และสอดคล้องกับหลักการจริยธรรมปัญญาประดิษฐ์แต่ละมิติเพียงใด
3. **การประเมินผลกระทบ (Impact Mapping)** มุ่งเน้นการประเมินผลกระทบทั้งเชิงบวกและเชิงลบที่อาจเกิดขึ้นจากการประยุกต์ใช้ AI ตามหลักการในแต่ละมิติ พร้อมทั้งกำหนดมาตรการในการบรรเทาผลกระทบเชิงลบในกรณีที่ไม่สามารถกำกับดูแลการดำเนินงานให้สอดคล้องในแต่ละหลักการ

การประเมินขอบเขตการนำ AI มาประยุกต์ใช้ ในส่วนที่ 1 มุ่งเน้นการประเมินความสอดคล้องกับหลักการของ UNESCO ใน 2 มิติ ได้แก่ หลักการความสมเหตุสมผลและการไม่ก่อให้เกิดอันตราย (Proportionality and Do No Harm) เพื่อกลั่นกรองความจำเป็นและป้องกันความเสียหายจากการใช้ AI และหลักการกำกับดูแลที่ปรับตัวได้และการมีส่วนร่วมจากหลากหลายภาคส่วน (Multi-stakeholder and Adaptive Governance and Collaboration) เพื่อเปิดรับฟังความคิดเห็นและสะท้อนมุมมองของผู้มีส่วนได้ส่วนเสียอย่างรอบด้าน

สำหรับการประเมินในส่วนที่ 2 และส่วนที่ 3 มุ่งเน้นการประเมินการดำเนินงานและผลกระทบจากการนำ AI มาประยุกต์ใช้ตามหลักการอีก 8 มิติ เช่น หลักการความปลอดภัยและความมั่นคงปลอดภัย (Safety and Security) หลักการความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย (Fairness, Non-discrimination and Diversity) หลักการความยั่งยืน (Sustainability) และหลักความเป็นส่วนตัวและการคุ้มครองข้อมูล (Privacy and Data Protection) เป็นต้น เพื่อกำหนดมาตรการในการบรรเทาผลกระทบเชิงลบที่อาจเกิดขึ้นจากการนำ AI มาประยุกต์ใช้

การบูรณาการเครื่องมือทั้ง 3 ส่วนนี้ จึงไม่เพียงช่วยประเมินความพร้อมในระดับปฏิบัติการเท่านั้น แต่ยังเป็นกลไกสำคัญที่ช่วยให้มั่นใจว่า AI จะสามารถสร้างคุณค่า ถูกนำไปใช้งานอย่างมีความรับผิดชอบ และน่าเชื่อถือ โดยเคารพและคุ้มครองสิทธิมนุษยชน ตลอดจนเสริมสร้างความไว้วางใจให้แก่ผู้มีส่วนได้ส่วนเสียและสังคมอย่างยั่งยืน

บทที่ 1

ความเข้าใจพื้นฐานเกี่ยวกับ Ethical Impact Assessment (EIA)



1.1 หลักการจริยธรรมปัญญาประดิษฐ์ตามแนวทางของ UNESCO ประกอบด้วยอะไรบ้าง

เพื่อให้การใช้งาน AI เป็นไปโดยคำนึงถึงการเคารพสิทธิมนุษยชน ศักดิ์ศรีความเป็นมนุษย์ เสรีภาพขั้นพื้นฐาน ความเท่าเทียม ตลอดจนปกป้องสิ่งแวดล้อมและวัฒนธรรม ควบคู่ไปกับการเปิดพื้นที่ให้ทุกภาคส่วนได้มีส่วนร่วมและเข้าถึงองค์ความรู้และเทคโนโลยีอย่างเป็นธรรม UNESCO จึงได้มีการกำหนดหลักการโดยครอบคลุม 10 มิติ ดังนี้

มิติที่ 1: ความสมเหตุสมผลและการไม่ก่อให้เกิดอันตราย (Proportionality and Do No Harm)

มิตินี้มุ่งเน้นการรับรองว่าการใช้ AI ต้องสมเหตุสมผลและได้สัดส่วนกับวัตถุประสงค์ โดยไม่ละเมิดสิทธิมนุษยชนหรือก่ออันตรายต่อสังคมและสิ่งแวดล้อม องค์กรต้องมีมาตรการประเมินความเสี่ยงเพื่อป้องกันความเสียหายอย่างรัดกุม และห้ามใช้ AI เพื่อการจัดอันดับทางสังคมหรือส่อส่งประชาชน และมนุษย์ต้องเป็นผู้มีอำนาจตัดสินใจขั้นสุดท้ายในกรณีที่ส่งผลกระทบที่ย้อนคืนไม่ได้หรือเกี่ยวข้องกับชีวิตและชีวิต

มิติที่ 2: ความปลอดภัยและความมั่นคงปลอดภัย (Safety and Security)

มิตินี้มุ่งเน้นการพัฒนาและใช้งาน AI ให้มีความปลอดภัย เชื่อถือได้ และสามารถป้องกันความเสียหายที่อาจเกิดขึ้นต่อบุคคล สังคม หรือระบบงาน โดยองค์กรควรพิจารณาความเสี่ยงที่อาจเกิดขึ้นจากข้อผิดพลาดของระบบ การโจมตีทางไซเบอร์ หรือการใช้งาน AI ในทางที่ไม่เหมาะสม รวมถึงกำหนดมาตรการป้องกันตั้งแต่ขั้นตอนการออกแบบและพัฒนา

มิติที่ 3: ความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย (Fairness, Non-discrimination and Diversity)

มิตินี้มุ่งเน้นการส่งเสริมให้ AI ปฏิบัติต่อบุคคลและกลุ่มคนอย่างเป็นธรรม ไม่ก่อให้เกิดการเลือกปฏิบัติหรือความเหลื่อมล้ำ โดยองค์กรควรพิจารณาความหลากหลายของข้อมูล การลดอคติของระบบ และการคำนึงถึงผู้ใช้ที่มีพื้นฐานแตกต่างกัน เพื่อให้ทุกคนสามารถเข้าถึงประโยชน์จาก AI ได้อย่างเท่าเทียม

มิติที่ 4: ความยั่งยืน (Sustainability)

มิตินี้มุ่งเน้นการพิจารณาผลกระทบระยะยาวของ AI ต่อสังคม เศรษฐกิจ และสิ่งแวดล้อม องค์กรควรคำนึงถึงการใช้ทรัพยากรอย่างเหมาะสม การใช้พลังงาน ผลกระทบต่อสิ่งแวดล้อม ตลอดจนผลกระทบต่อการจ้างงานและสังคม เพื่อให้การใช้ AI สนับสนุนการพัฒนาอย่างยั่งยืน

มิติที่ 5: ความเป็นส่วนตัวและการคุ้มครองข้อมูล (Privacy and Data Protection)

มิตินี้มุ่งเน้นการคุ้มครองข้อมูลส่วนบุคคลและสิทธิความเป็นส่วนตัวของผู้เกี่ยวข้อง โดยองค์กรควรพิจารณาการเก็บ ใช้ เผยแพร่ และจัดการข้อมูลให้สอดคล้องกับกฎหมายและหลักการคุ้มครองข้อมูลส่วนบุคคล รวมถึงการออกแบบระบบโดยคำนึงถึงความเป็นส่วนตัวตั้งแต่ต้น (Privacy by Design)

มิติที่ 6: การกำกับดูแลและการตัดสินใจโดยมนุษย์ (Human Oversight and Determination)

มิตินี้มุ่งเน้นการรักษาบทบาทของมนุษย์ในการควบคุมการทำงานของ AI โดยเฉพาะในกรณีที่การตัดสินใจอาจส่งผลกระทบต่อสิทธิ โอกาส หรือความเป็นอยู่ของบุคคล องค์กรควรกำหนดกลไกให้มนุษย์สามารถตรวจสอบ ทบทวน แก้ไข หรือยกเลิกการตัดสินใจของ AI ได้เมื่อจำเป็น

มิติที่ 7: ความโปร่งใสและความสามารถในการอธิบาย (Transparency and Explainability)

มิตินี้มุ่งเน้นการแจ้งให้ผู้ใช้ทราบเมื่อมีการใช้ AI และรับรองว่าระบบสามารถอธิบายที่มาของผลลัพธ์ได้อย่างชัดเจน โดยเฉพาะเมื่อการตัดสินใจกระทบต่อสิทธิหรือความปลอดภัย บุคคลต้องสามารถเข้าถึงเหตุผลที่ได้รับคำอธิบายที่เข้าใจง่าย และมีช่องทางร้องเรียนให้มนุษย์ทบทวนแก้ไขได้

มิติที่ 8: ความรับผิดชอบ และภาระรับผิดชอบ (Accountability and Responsibility)

มิตินี้มุ่งเน้นการรับรองว่าผลของการตัดสินใจโดย AI ต้องผูกพันกับมนุษย์ตามบทบาทในวงจรชีวิตของ AI เสมอ องค์กรต้องจัดให้มีกลไกกำกับดูแล การตรวจสอบ (Audit) และการสืบย้อนกลับได้ (Traceability) ทั้งในเชิงเทคนิคและโครงสร้างองค์กร เพื่อป้องกันและเยียวยาผลกระทบที่อาจเกิดขึ้น

มิติที่ 9: ความตระหนักรู้และความรู้ความสามารถ (Awareness and Literacy)

มิตินี้มุ่งเน้นการส่งเสริมความรู้ ความเข้าใจ และทักษะที่เกี่ยวข้องกับ AI ให้แก่ผู้พัฒนา ผู้ใช้งาน และผู้ได้รับผลกระทบ เพื่อให้สามารถใช้ AI ได้อย่างเหมาะสม เข้าใจข้อจำกัดของเทคโนโลยี และไม่พึ่งพาการตัดสินใจของ AI มากเกินไป

มิติที่ 10: การกำกับดูแลที่ปรับตัวได้และการมีส่วนร่วมจากหลากหลายภาคส่วน (Multi-stakeholder and Adaptive Governance and Collaboration)

มิตินี้มุ่งเน้นการเปิดพื้นที่ให้ผู้มีส่วนได้ส่วนเสียที่หลากหลาย ตลอดจนกลุ่มเปราะบาง ได้เข้ามามีส่วนร่วมอย่างมีความหมายตลอดวงจรชีวิตของระบบ AI เพื่อขับเคลื่อนไปสู่การพัฒนาที่ยั่งยืน ขณะเดียวกัน องค์กรต้องมีความยืดหยุ่นพร้อมปรับเปลี่ยนกลไกการดำเนินงานและมาตรการควบคุมต่าง ๆ ให้เท่าทันต่อการวิวัฒนาการของเทคโนโลยีใหม่ และบริบทของกฎหมายที่เปลี่ยนแปลงไปอยู่เสมอ

1.2 เหตุใดการประเมินผลกระทบทางจริยธรรมจึงมีความสำคัญ

ปัญญาประดิษฐ์ หรือ AI ได้เข้ามามีบทบาทสำคัญในการดำเนินงานขององค์กรทั้งภาครัฐและภาคเอกชนอย่างต่อเนื่อง เพื่อเพิ่มประสิทธิภาพ สนับสนุนการตัดสินใจ และสร้างนวัตกรรมใหม่ ๆ แต่อย่างไรก็ตาม การใช้งาน AI นั้นอาจก่อให้เกิดผลกระทบที่ไม่พึงประสงค์ในหลายมิติ ตัวอย่างเช่น การละเมิดความเป็นส่วนตัว ความไม่เป็นธรรมต่อกลุ่มเปราะบาง การตัดสินใจที่ไม่โปร่งใส และการทำงานที่ส่งผลกระทบต่อสิทธิมนุษยชน เป็นต้น ซึ่งหากขาดการพิจารณาอย่างรอบด้าน อาจนำไปสู่ความเสี่ยงและอันตรายที่ส่งผลกระทบต่อบุคคลและสังคมในระยะยาว

ดังนั้น การประเมินผลกระทบทางจริยธรรมจึงเป็นสิ่งจำเป็นอย่างยิ่ง เนื่องจากจะช่วยให้องค์กรเห็นถึงผลกระทบเชิงลบที่อาจเกิดขึ้น พร้อมทั้งกำหนดมาตรการป้องกันความเสียหาย ลดความเสี่ยงจากการตัดสินใจหรือหลีกเลี่ยงผลลัพธ์ที่ไม่พึงประสงค์ ตลอดจนเสริมสร้างความน่าเชื่อถือและความไว้วางใจให้กับระบบ AI อย่างยั่งยืน

1.3 การประเมินผลกระทบทางจริยธรรม (Ethical Impact Assessment: EIA) คืออะไร

UNESCO ได้กำหนดให้การประเมินผลกระทบทางจริยธรรมเป็นหนึ่งในกลไกสำคัญสำหรับการกำกับดูแล AI อย่างมีความรับผิดชอบ โดยมีเป้าหมายเพื่อช่วยให้องค์กรสามารถพิจารณาผลกระทบต่อสิทธิมนุษยชน เสรีภาพขั้นพื้นฐาน สิ่งแวดล้อม และผลกระทบทางสังคมที่อาจเกิดขึ้นจากการใช้ AI ได้อย่างเป็นระบบ

การประเมินผลกระทบทางจริยธรรมจะช่วยให้องค์กรเห็นถึงคุณค่าจากการนำ AI มาใช้งานหรือผลกระทบเชิงบวก (Positive Impact) และตระหนักถึงผลกระทบเชิงลบ (Negative Impact) ที่อาจเกิดขึ้นจากการนำ AI มาใช้งาน พร้อมทั้งนำผลการประเมินดังกล่าวมากำหนดเป็นมาตรการหรือกลยุทธ์ในการบรรเทาผลกระทบเชิงลบ แก้ไขผลลัพธ์ที่ไม่พึงประสงค์ และใช้เป็นข้อมูลเพื่อประกอบการตัดสินใจในการดำเนินโครงการดังกล่าว

1.4 EIA ช่วยสนับสนุนการตัดสินใจอย่างไร

EIA เป็นเครื่องมือที่ช่วยให้องค์กรเข้าใจผลกระทบทางจริยธรรมและผลกระทบที่ไม่ได้ตั้งใจ (Unintended Consequences) ได้อย่างลึกซึ้งและรอบด้าน โดยผลลัพธ์จากการประเมินจะช่วยให้เกิดการออกแบบ พัฒนา จัดหา และการใช้งานโดยมีข้อมูลเชิงลึกเพื่อประกอบการพิจารณา ซึ่งจะส่งผลให้องค์กรสามารถตัดสินใจได้อย่างรอบคอบ ลดความเสี่ยง และป้องกันความเสียหายอย่างเท่าทัน

1.5 ความเชื่อมโยงระหว่าง EIA และ AI Governance

AI Governance เป็นกรอบการกำกับดูแลที่ช่วยให้องค์กรสามารถบริหารจัดการโอกาส ความเสี่ยง และผลกระทบที่เกิดจาก AI ได้อย่างเหมาะสม ขณะที่ EIA เป็นหนึ่งในเครื่องมือสำคัญภายใต้ระบบ AI Governance ที่ช่วยให้องค์กรสามารถประเมินผลกระทบทางจริยธรรมได้อย่างเป็นระบบ

ดังนั้น EIA จึงทำหน้าที่เป็นกลไกเชื่อมโยงระหว่างหลักการด้านจริยธรรมของ AI กับการตัดสินใจในทางปฏิบัติ โดยช่วยให้องค์กรสามารถนำหลักการด้านจริยธรรมมาพิจารณาประกอบการตัดสินใจในการนำ AI มาประยุกต์ใช้ในบริบทต่างๆ และกำหนดมาตรการกำกับดูแลที่เหมาะสมเพื่อป้องกันผลกระทบเชิงลบที่อาจเกิดขึ้นกับผู้มีส่วนได้ส่วนเสียและสังคม

1.6 EIA ควรดำเนินการเมื่อใด

การประเมินผลกระทบทางจริยธรรม (Ethical Impact Assessment: EIA) ไม่ควรถูกมองว่าเป็นกิจกรรมที่ดำเนินการเพียงครั้งเดียวก่อนนำ AI ไปใช้งาน แต่ควรเป็นกระบวนการที่สามารถดำเนินการและทบทวนได้ตลอดวงจรชีวิตของระบบ AI (AI Lifecycle) ตามความเหมาะสมของบริบทและระดับความเสี่ยงของ AI Use Case โดยทั่วไป องค์กรควรพิจารณาดำเนินการ EIA ในช่วงเวลาสำคัญ ดังนี้

ก่อนนำ AI ไปใช้งาน (Before Deployment)

เป็นช่วงเวลาที่แนะนำให้ดำเนินการ EIA มากที่สุด โดยเฉพาะอย่างยิ่งในระหว่างขั้นตอนการออกแบบพัฒนา (Design and Develop) กระบวนการจัดหา (Procurement) และก่อนนำไปใช้งานจริง (Initial Deployment) เนื่องจากการประเมินในช่วงก่อนนำไปใช้งานจะช่วยให้องค์กรสามารถทำความเข้าใจบริบท วัตถุประสงค์ ผู้มีส่วนได้ส่วนเสีย และผลกระทบที่อาจเกิดขึ้นจาก AI ได้ตั้งแต่ระยะเริ่มต้น รวมถึงสามารถระบุความเสี่ยงและกำหนดมาตรการกำกับดูแลที่เหมาะสมก่อนเริ่มใช้งานจริง

ระหว่างการใช้งาน (During and Post Deployment)

องค์กรอาจพิจารณาทบทวนผลการประเมินหรือดำเนินการประเมินเพิ่มเติม เพื่อติดตามผลกระทบที่เกิดขึ้นจริง ตรวจสอบว่าสมมติฐานที่ใช้ในการประเมินยังคงสอดคล้องกับบริบทการใช้งาน และพิจารณาว่ามีประเด็นด้านจริยธรรมใหม่ที่ต้องได้รับการจัดการหรือไม่

เมื่อบริบทมีการเปลี่ยนแปลง (When Context Changes)

องค์กรควรพิจารณาทบทวนหรือประเมินซ้ำเมื่อมีการเปลี่ยนแปลงที่อาจส่งผลกระทบต่อผลกระทบทางจริยธรรมของ AI Use Case เช่น

- มีการเปลี่ยนวัตถุประสงค์ของระบบ AI
- มีการเปลี่ยนแหล่งข้อมูลหรือรูปแบบข้อมูลที่ใช้
- มีการเปลี่ยนกลุ่มผู้ใช้งานหรือผู้ได้รับผลกระทบ
- มีการปรับปรุงโมเดล (Model) หรือเทคโนโลยีที่เกี่ยวข้อง
- มีกฎหมาย ข้อกำหนด หรือบริบททางสังคมที่เปลี่ยนแปลงไป

แม้ว่า EIA จะสามารถดำเนินการได้ในหลายช่วงของวงจรชีวิต AI แต่การประเมินตั้งแต่ก่อนนำ AI ไปใช้งานจะช่วยให้องค์กรสามารถคาดการณ์และบริหารจัดการผลกระทบทางจริยธรรมได้อย่างมีประสิทธิภาพมากกว่าการแก้ไขปัญหาหลังจากผลกระทบเกิดขึ้นแล้ว

1.7 จัดทำ EIA ต้องดำเนินการอย่างไร

ในการประเมินผลกระทบจะต้องเกิดจากการทำงานร่วมกันระหว่างทีมงานที่มีความเชี่ยวชาญที่หลากหลายและผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้อง โดยในการทำงานร่วมกันดังกล่าวจำเป็นต้องมีการกำหนดผู้รับผิดชอบ รวมถึงกำหนดวิธีการในการทำงานร่วมกัน เช่น การกำหนดประเด็นในการประเมินตามความเชี่ยวชาญ การจัดประชุมระดมสมอง เป็นต้น เพื่อให้การผลการประเมินสะท้อนถึงข้อมูลในมิติต่าง ๆ ได้อย่างครบถ้วน ถูกต้อง และมีการกำหนดมาตรการหรือกลยุทธ์ที่สามารถนำไปสู่การปฏิบัติได้จริง โดยมีรายละเอียดดังต่อไปนี้

1) กำหนดผู้รับผิดชอบโครงการ (Identify a Project Manager)

กำหนดผู้รับผิดชอบหลักในการวางแผน ประสานงาน และติดตามการดำเนินกระบวนการ EIA ตลอดทั้งโครงการ เพื่อให้การประเมินเป็นไปอย่างต่อเนื่อง มีการกำหนดระยะเวลา ผู้รับผิดชอบ และสามารถประสานการทำงานระหว่างผู้ที่เกี่ยวข้องได้อย่างมีประสิทธิภาพ

2) จัดตั้งทีมประเมินที่มีความหลากหลาย (Assemble a Diverse Team)

จัดตั้งทีมที่ประกอบด้วยผู้บริหารหรือหัวหน้าทีมที่มีหน้าที่รับผิดชอบเพื่อสนับสนุนและกำหนดวิสัยทัศน์ รวมถึงมีความเชี่ยวชาญจากหลากหลายสาขาและผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้อง เช่น ด้านเทคนิค กฎหมาย ธุรกิจ สังคมศาสตร์ และการดำเนินงาน เพื่อให้การประเมินสะท้อนมุมมองที่รอบด้าน ครอบคลุมทั้งประเด็นทางเทคนิคและผลกระทบต่อสังคม

3) ประเมินในส่วนที่รับผิดชอบตามความเชี่ยวชาญ (Independently Work on the EIA)

มอบหมายให้ผู้เชี่ยวชาญแต่ละด้านร่วมตอบหรือทบทวนคำถามในส่วนที่เกี่ยวข้องกับบทบาทของตนเอง โดยสามารถดำเนินการได้อย่างอิสระก่อนการประชุมร่วมกัน เพื่อเตรียมข้อมูลและข้อคิดเห็นสำหรับการอภิปรายในขั้นตอนถัดไป

4) จัดประชุมเชิงปฏิบัติการร่วมกัน (Facilitate a Collaborative Workshop)

นำผลการประเมินจากผู้เชี่ยวชาญแต่ละด้านมาร่วมอภิปราย แลกเปลี่ยนความคิดเห็น และวิเคราะห์ผลกระทบทางจริยธรรมร่วมกัน เพื่อสร้างความเข้าใจร่วม ระบุประเด็นที่ควรปรับปรุง และกำหนดแนวทางในการลดผลกระทบที่อาจเกิดขึ้น

5) การมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย (Stakeholder Engagement)

เปิดโอกาสให้ผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้องเข้ามามีส่วนร่วมตลอดกระบวนการ EIA โดยเฉพาะในช่วงการประชุมเชิงปฏิบัติการ เพื่อให้การประเมินสะท้อนมุมมองที่หลากหลายและครอบคลุมผลกระทบที่อาจเกิดขึ้น

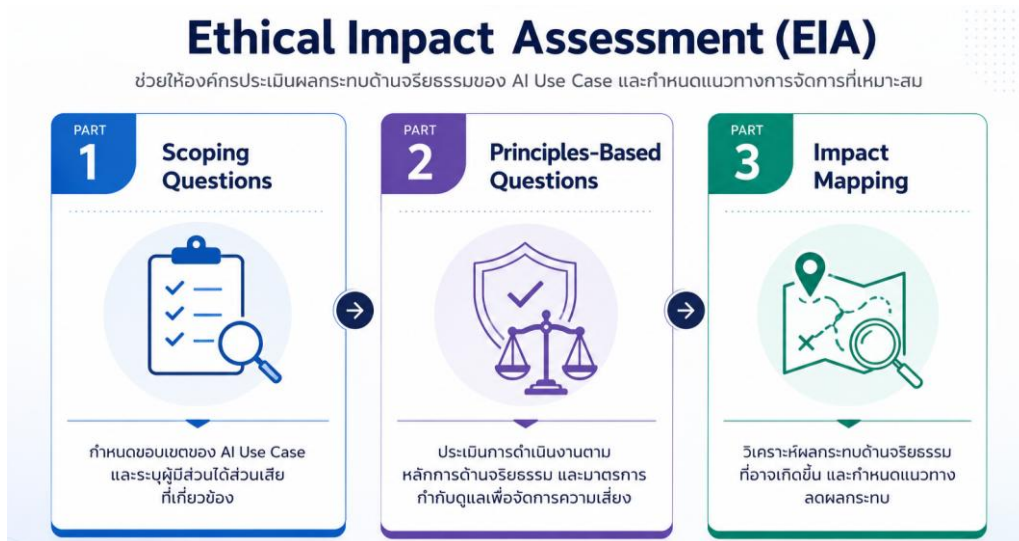
6) ผลการประเมินและกำหนดแนวทางดำเนินงาน (End of Assessment)

สรุปผลการประเมิน กำหนดแนวทางดำเนินงาน ผู้รับผิดชอบ และระยะเวลาดำเนินการสำหรับการปรับปรุงระบบ AI รวมถึงกำหนดแนวทางการติดตามและทบทวนผลการประเมินอย่างต่อเนื่อง เพื่อให้การกำกับดูแล AI สามารถปรับตัวตามการเปลี่ยนแปลงของระบบและบริบทการใช้งาน

1.8 ภาพรวมกระบวนการประเมินผลกระทบทางจริยธรรม

การประเมินผลกระทบทางจริยธรรม (Ethical Impact Assessment: EIA) เป็นกระบวนการที่ช่วยให้องค์กรสามารถทำความเข้าใจผลกระทบที่อาจเกิดขึ้นจากการประยุกต์ใช้ปัญญาประดิษฐ์ (Artificial Intelligence: AI) ได้อย่างรอบด้าน ทั้งในมิติของประโยชน์ ความเสี่ยง และผลกระทบที่อาจเกิดขึ้นต่อผู้มีส่วนได้ส่วนเสีย สังคม และสิ่งแวดล้อม

คู่มือฉบับนี้อ้างอิงหลักการด้านจริยธรรมสำหรับปัญญาประดิษฐ์ของ UNESCO และออกแบบกระบวนการประเมินให้เป็น 3 ขั้นตอน เพื่อให้สอดคล้องกับลำดับการพัฒนาและการใช้งาน AI ในทางปฏิบัติ ประกอบด้วย การกำหนดขอบเขตการประเมิน (Scoping Questions) การประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions) และการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ดังแสดงในภาพต่อไปนี้



ผลลัพธ์จากแต่ละขั้นตอนนี้จะถูกนำไปใช้ประกอบการวิเคราะห์ในขั้นตอนถัดไป จนสามารถนำไปสู่ข้อเสนอแนะมาตรการกำกับดูแล และการตัดสินใจด้าน AI ที่เหมาะสมกับบริบทขององค์กร

Part 1 การกำหนดขอบเขตการประเมิน (Scoping Questions)

แนวคิดของการกำหนดขอบเขตการประเมิน:

การประเมินผลกระทบทางจริยธรรมที่มีประสิทธิภาพจำเป็นต้องเริ่มต้นจากการทำความเข้าใจบริบทของกรณีการใช้งาน AI (AI Use Case) อย่างรอบด้าน เนื่องจากผลกระทบที่อาจเกิดขึ้นไม่ได้ขึ้นอยู่กับเทคโนโลยี AI เพียงอย่างเดียว แต่ยังขึ้นอยู่กับวัตถุประสงค์ของการใช้งาน ลักษณะของระบบ

ปัญญาประดิษฐ์ (AI System) ผู้ใช้งาน ผู้ที่อาจได้รับผลกระทบ กลไกการกำกับดูแล และบริบทแวดล้อมที่เกี่ยวข้อง ดังนั้น ก่อนเริ่มการประเมินในมิติด้านจริยธรรม องค์กรจึงควรทำความเข้าใจขอบเขตของการทำงาน เหตุผลในการนำ AI มาใช้ และผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้อง เพื่อให้สามารถกำหนดขอบเขตของการประเมินได้อย่างเหมาะสม และสะท้อนบริบทของการใช้งานจริง

ซึ่งในส่วนการกำหนดขอบเขตการประเมินนี้ จะเป็นไปตาม 2 หลักการจริยธรรมปัญญาประดิษฐ์ของ UNESCO ได้แก่ หลักการที่ 1 ความสมเหตุสมผลและการไม่ก่อให้เกิดอันตราย (Proportionality and Do No Harm) เพื่อพิจารณาถึงความจำเป็นในการใช้งานและป้องกันความเสียหายจากการใช้ AI และหลักการที่ 2 การกำกับดูแลที่ปรับตัวได้และการมีส่วนร่วมจากหลากหลายภาคส่วน (Multi-stakeholder and Adaptive Governance and Collaboration) เพื่อองค์กรสามารถเห็นถึงผลกระทบที่อาจไม่เคยพิจารณามาก่อนและสามารถรองรับการเปลี่ยนแปลงที่อาจเกิดขึ้นได้อย่างมีประสิทธิภาพ

สิ่งที่ต้องพิจารณาในขั้นตอนการกำหนดขอบเขตการประเมิน (Scoping Questions)

1) ข้อมูลพื้นฐานของระบบปัญญาประดิษฐ์ (AI System Description)

ทำความเข้าใจวัตถุประสงค์ ลักษณะการใช้งาน ระยะเวลาวงจรชีวิต ผู้ใช้งาน ผู้ที่อาจได้รับผลกระทบ ขอบเขตการนำไปใช้ และปัจจัยแวดล้อมที่เกี่ยวข้อง เพื่อใช้เป็นข้อมูลตั้งต้นในการประเมิน

2) กลไกการกำกับดูแลโครงการ AI (Project Governance)

พิจารณาทบทวน ความรับผิดชอบ อำนาจในการตัดสินใจ ความหลากหลายของทีมโครงการ และกลไกการทบทวนประเด็นด้านความรับผิดชอบ ความโปร่งใส และจริยธรรมของระบบ AI

3) การพิจารณาความสมเหตุสมผลของการใช้ AI และหลักการไม่ก่อให้เกิดอันตราย (Proportionality Screening and Do No Harm)

พิจารณาว่าการนำ AI มาใช้มีความจำเป็น เหมาะสม และได้สัดส่วนกับวัตถุประสงค์ของโครงการหรือไม่ รวมถึงประเมินว่าระบบ AI อาจก่อให้เกิดอันตรายหรือผลกระทบต่อบุคคล สังคม สิทธิมนุษยชน ศักดิ์ศรีความเป็นมนุษย์ ความปลอดภัย หรือเสรีภาพขั้นพื้นฐานหรือไม่

4) การมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย (Stakeholder Engagement)

พิจารณาว่าผู้มีส่วนได้ส่วนเสียที่อาจได้รับผลกระทบจากระบบ AI ได้รับการรับฟังหรือมีส่วนร่วมอย่างเหมาะสมหรือไม่ รวมถึงพิจารณาว่าข้อคิดเห็นหรือข้อเสนอแนะที่ได้รับถูกนำไปใช้ประกอบการออกแบบ พัฒนา ใช้งาน หรือกำกับดูแลระบบ AI อย่างไร

ผลลัพธ์ที่คาดหวังจากการกำหนดขอบเขตการประเมิน (Scoping Questions)

เมื่อดำเนินการในขั้นตอนนี้แล้ว องค์กรควรสามารถ

- เข้าใจวัตถุประสงค์ ลักษณะการใช้งาน ขอบเขต และบริบทของระบบ AI ได้อย่างชัดเจน
- ระบุผู้ใช้งาน ผู้มีส่วนได้ส่วนเสีย และผู้ที่อาจได้รับผลกระทบได้อย่างครบถ้วน
- เข้าใจบทบาท ความรับผิดชอบ และกลไกการกำกับดูแลที่เกี่ยวข้องกับระบบ AI
- พิจารณาความจำเป็น ความเหมาะสม และข้อจำกัดของการนำ AI มาใช้ในบริบทของโครงการ

- ระบุประเด็นสำคัญที่ควรนำไปประเมินตามหลักการด้านจริยธรรมในขั้นตอนถัดไป

โดยผลลัพธ์จากขั้นตอนนี้จะเป็นพื้นฐานสำหรับการประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions) และการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ในขั้นตอนถัดไป

Part 2 การประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions)

แนวคิดของการประเมินตามหลักการด้านจริยธรรม

ภายหลังจากการกำหนดขอบเขตการประเมิน (Scoping Questions) องค์กรจะมีข้อมูลพื้นฐานเกี่ยวกับระบบปัญญาประดิษฐ์ (AI System) และบริบทของการใช้งานแล้ว ขั้นตอนนี้จึงมุ่งประเมินว่าการออกแบบ พัฒนา และใช้งาน AI มีมาตรการและกลไกการกำกับดูแลที่สอดคล้องกับหลักการด้านจริยธรรมของ UNESCO หรือไม่ ซึ่งการประเมินในขั้นตอนนี้ช่วยให้องค์กรทบทวนความพร้อมของมาตรการกำกับดูแล ระบุจุดแข็ง ช่องว่าง และประเด็นที่ควรปรับปรุง ก่อนนำข้อมูลไปใช้วิเคราะห์ผลกระทบทางจริยธรรมในขั้นตอนถัดไป

สิ่งที่ต้องพิจารณาในขั้นตอนการประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions)

ในขั้นตอนนี้จะมุ่งเน้นการประเมินหลักการด้านจริยธรรมปัญญาประดิษฐ์ที่เหลืออีก 8 หลักการ เพื่อพิจารณาว่าองค์กรมีมาตรการและกลไกการกำกับดูแลที่เหมาะสมเพียงพอหรือไม่ ซึ่งครอบคลุมหลักการสำคัญดังต่อไปนี้

- ความปลอดภัยและความมั่นคงปลอดภัย (Safety and Security)
- ความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย (Fairness, Non-discrimination and Diversity)
- ความยั่งยืน (Sustainability)
- ความเป็นส่วนตัวและการคุ้มครองข้อมูล (Privacy and Data Protection)
- การกำกับดูแลและการตัดสินใจโดยมนุษย์ (Human Oversight and Determination)
- ความโปร่งใส และความสามารถในการอธิบาย (Transparency and Explainability)
- ความรับผิดชอบ และภาระรับผิดชอบ (Accountability and Responsibility)
- ความตระหนักรู้และความรู้ความสามารถ (Awareness and Literacy)

โดยหลักการแต่ละด้านจะช่วยให้องค์กรสามารถทบทวนความพร้อมของมาตรการกำกับดูแล ระบุช่องว่าง (Gap) และเตรียมข้อมูลสำหรับการวิเคราะห์ผลกระทบทางจริยธรรมในขั้นตอนถัดไป

ผลลัพธ์ที่คาดหวังจากการประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions)

เมื่อดำเนินการในขั้นตอนนี้แล้ว องค์กรควรสามารถ

- ประเมินความพร้อมของมาตรการและกลไกการกำกับดูแล AI
- ระบุจุดแข็งและประเด็นที่ควรปรับปรุง
- เตรียมข้อมูลสำหรับการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping)

ผลลัพธ์จากขั้นตอนนี้จะเป็ข้อมูลสำคัญสำหรับการวิเคราะห์ผลกระทบเชิงบวกและเชิงลบ รวมถึงการกำหนดมาตรการกำกับดูแลและข้อเสนอแนะที่เหมาะสมในขั้นตอนถัดไป

Part 3 การวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping)

แนวคิดของการวิเคราะห์ผลกระทบทางจริยธรรม

หลังจากการกำหนดขอบเขตการประเมิน (Scoping Questions) และการประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions) องค์กรจะมีข้อมูลพื้นฐานเกี่ยวกับระบบปัญญาประดิษฐ์ (AI System) รวมถึงมาตรการและกลไกการกำกับดูแลที่เกี่ยวข้องแล้ว

ขั้นตอนการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) จึงมุ่งนำข้อมูลดังกล่าวมาวิเคราะห์ร่วมกัน เพื่อระบุผลกระทบที่อาจเกิดขึ้นจากการใช้งาน AI ทั้งในเชิงบวกและเชิงลบ ตลอดจนพิจารณาผู้ที่อาจได้รับผลกระทบ ระดับของผลกระทบ และแนวทางในการป้องกันหรือบรรเทาผลกระทบที่อาจเกิดขึ้น

สิ่งที่ต้องพิจารณาในขั้นตอน Impact Mapping

การวิเคราะห์ผลกระทบด้านจริยธรรมมุ่งนำข้อมูลจากขั้นตอนก่อนหน้ามาพิจารณาร่วมกัน เพื่อระบุผลกระทบที่อาจเกิดขึ้นจากการใช้งานระบบ AI และกำหนดแนวทางในการป้องกันหรือบรรเทาผลกระทบ โดยพิจารณาประเด็นสำคัญ ได้แก่

- ผลกระทบเชิงบวกและเชิงลบที่อาจเกิดขึ้น
- ระดับของผลกระทบและผู้ที่อาจได้รับผลกระทบ
- ระยะเวลาที่ผลกระทบอาจเกิดขึ้นหรือคงอยู่
- โอกาสที่ผลกระทบจะเกิดขึ้น
- สำหรับผลกระทบเชิงลบ ให้พิจารณาความสามารถในการแก้ไขหรือเยียวยา ประสิทธิภาพของมาตรการที่มีอยู่ และมาตรการเพิ่มเติมที่จำเป็น
- เหตุผล หลักฐาน สมมติฐาน และข้อจำกัดของข้อมูลที่ใช้ประกอบการประเมิน

ผลลัพธ์ที่คาดหวังจาก Impact Mapping

เมื่อดำเนินการในขั้นตอนนี้แล้ว องค์กรควรสามารถ

- ระบุผลกระทบเชิงบวกและผลกระทบเชิงลบที่อาจเกิดขึ้นจากการใช้งาน AI ได้อย่างเป็นระบบ
- เข้าใจว่าผลกระทบแต่ละประเด็นส่งผลต่อผู้มีส่วนได้ส่วนเสียกลุ่มใด
- ระบุแนวทางในการป้องกันหรือบรรเทาผลกระทบที่อาจเกิดขึ้น
- นำผลการวิเคราะห์ไปใช้ประกอบการตัดสินใจ ออกแบบมาตรการกำกับดูแล และปรับปรุง AI System ให้สอดคล้องกับหลักการด้านจริยธรรมและบริบทของการใช้งานจริง

บทที่ 2

การกำหนดขอบเขตการประเมิน



การกำหนดขอบเขตการประเมิน (Scoping Questions) เป็นขั้นตอนแรกของกระบวนการประเมินผลกระทบทางจริยธรรมของระบบปัญญาประดิษฐ์ (Ethical Impact Assessment: EIA) โดยมีวัตถุประสงค์เพื่อให้องค์กรทำความเข้าใจบริบทของระบบปัญญาประดิษฐ์ (AI System) ที่จะนำมาประเมินอย่างรอบด้าน ก่อนเข้าสู่การประเมินตามหลักการด้านจริยธรรมและการวิเคราะห์ผลกระทบในขั้นตอนถัดไป

โดยการประเมินผลกระทบทางจริยธรรมไม่ควรเริ่มจากการพิจารณาความเสี่ยงหรือผลกระทบในทันที หากยังไม่เข้าใจว่าระบบ AI ถูกพัฒนาขึ้นเพื่อวัตถุประสงค์ใด ใครเป็นผู้ใช้งาน ใครอาจได้รับผลกระทบ ระบบถูกนำไปใช้ในบริบทใด และมีกลไกการกำกับดูแลอย่างไร ดังนั้นการกำหนดขอบเขตการประเมินจึงทำหน้าที่เป็นฐานข้อมูลตั้งต้นของการประเมิน เพื่อช่วยให้องค์กรสามารถกำหนดขอบเขตของกรณีการใช้งาน AI ได้อย่างชัดเจน และทำให้การประเมินในขั้นตอนต่อไปสะท้อนบริบทของการใช้งานจริง

2.1 หลักการของการกำหนดขอบเขตการประเมิน

หลักการสำคัญของการกำหนดขอบเขตการประเมิน คือ การทำความเข้าใจระบบ AI และบริบทของการใช้งานก่อนเริ่มประเมินผลกระทบทางจริยธรรม เนื่องจากผลกระทบของระบบ AI ไม่ได้เกิดจากเทคโนโลยีเพียงอย่างเดียว แต่เกิดจากความเชื่อมโยงระหว่างวัตถุประสงค์ของระบบ วิธีการใช้งาน ข้อมูลที่ใช้ ผู้ใช้งาน ผู้ที่อาจได้รับผลกระทบ กลไกการกำกับดูแล และสภาพแวดล้อมที่ระบบถูกนำไปใช้ ระบบ AI ที่มีลักษณะทางเทคนิคคล้ายกันอาจก่อให้เกิดระดับความเสี่ยงและผลกระทบที่แตกต่างกัน หากถูกนำไปใช้ในบริบทที่ต่างกัน เช่น ระบบ AI ที่ใช้เพื่อให้ข้อมูลทั่วไปแก่ผู้ใช้งาน อาจมีระดับความเสี่ยงแตกต่างจากระบบ AI ที่ใช้สนับสนุนการตัดสินใจเกี่ยวกับสิทธิ บริการสาธารณะ การเงิน สุขภาพ การศึกษา หรือโอกาสของบุคคล ดังนั้น การประเมินจึงควรเริ่มจากการระบุบริบทของระบบให้ชัดเจนก่อน

การกำหนดขอบเขตการประเมินในคู่มือฉบับนี้สอดคล้องกับหลักการด้านจริยธรรมของ UNESCO โดยเฉพาะหลักการ ความสมเหตุสมผลและการไม่ก่อให้เกิดอันตราย (Proportionality and Do No Harm) และหลักการ การกำกับดูแลที่ปรับตัวได้และการมีส่วนร่วมจากหลากหลายภาคส่วน (Multi-stakeholder and Adaptive Governance and Collaboration)

หลักการความสมเหตุสมผลและการไม่ก่อให้เกิดอันตรายมุ่งให้องค์กรพิจารณาว่าการนำ AI มาใช้มีความจำเป็นและเหมาะสมกับวัตถุประสงค์ของโครงการหรือไม่ มีทางเลือกอื่นที่ไม่ใช่ AI หรือไม่ และการใช้ AI อาจก่อให้เกิดผลกระทบต่อบุคคล สังคม สิทธิมนุษยชน คีร์ติศรีความเป็นมนุษย์ ความปลอดภัย หรือเสรีภาพขั้นพื้นฐานหรือไม่

ขณะที่หลักการด้านการกำกับดูแลที่ปรับตัวได้และการมีส่วนร่วมจากหลากหลายภาคส่วน มุ่งให้องค์กรพิจารณากลไกการกำกับดูแล บทบาทและความรับผิดชอบของผู้เกี่ยวข้อง รวมถึงการเปิดโอกาสให้ผู้มีส่วนได้ส่วนเสียเข้ามามีส่วนร่วม เพื่อให้การประเมินสะท้อนมุมมองที่หลากหลาย และสามารถรองรับการเปลี่ยนแปลงที่อาจเกิดขึ้นตลอดวงจรชีวิตของระบบ AI

กล่าวโดยสรุปการกำหนดขอบเขตการประเมิน เป็นขั้นตอนที่ช่วยให้องค์กรตอบคำถามพื้นฐานก่อนการประเมินได้ว่า ระบบ AI นี้คืออะไร ถูกนำมาใช้เพื่อวัตถุประสงค์ใด ใครเกี่ยวข้อง ใครได้รับผลกระทบ มีความจำเป็นเพียงใด และมีเงื่อนไขหรือปัจจัยใดที่อาจส่งผลกระทบต่อความเสี่ยงและผลกระทบทางจริยธรรม

2.2 คำถามสำหรับการกำหนดขอบเขตการประเมิน

คำถามสำหรับการกำหนดขอบเขตการประเมินแบ่งออกเป็น 4 หมวดหลัก ได้แก่ ข้อมูลพื้นฐานของระบบปัญญาประดิษฐ์ กลไกการกำกับดูแลโครงการ AI ความเหมาะสมผลของการใช้ AI และหลักการไม่ก่อให้เกิดอันตราย และการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย โดยแต่ละหมวดมีวัตถุประสงค์เพื่อช่วยให้องค์กรเตรียมข้อมูลที่จำเป็นสำหรับการประเมินในขั้นตอนต่อไป

2.2.1 ข้อมูลพื้นฐานของระบบปัญญาประดิษฐ์ (AI System Description)

ส่วนนี้มีวัตถุประสงค์เพื่อทำความเข้าใจภาพรวมของระบบปัญญาประดิษฐ์ (AI System) ที่จะนำมาประเมิน ทั้งในด้านวัตถุประสงค์ของระบบ ลักษณะการใช้งาน ระยะของวงจรชีวิต ผู้ใช้งาน ผู้ที่อาจได้รับผลกระทบ ขอบเขตการนำไปใช้ และปัจจัยแวดล้อมที่เกี่ยวข้อง ข้อมูลในส่วนนี้ไม่ใช่เพียงรายละเอียดทางเทคนิคของระบบ แต่เป็นข้อมูลตั้งต้นที่ช่วยให้ผู้ประเมินเห็นบริบทของการใช้งานจริง โดยมีประเด็นคำถามดังนี้

ตารางที่ 2-1 ประเด็นคำถามสำคัญสำหรับข้อมูลพื้นฐานของระบบปัญญาประดิษฐ์
(AI System Description)

กลุ่มประเด็น	รายการที่ควรพิจารณา
วัตถุประสงค์และลักษณะของระบบ AI	ระบบ AI คืออะไร ถูกนำมาใช้เพื่อวัตถุประสงค์ใด และมีบทบาทอย่างไรในกระบวนการทำงาน
สถานะของระบบและบริบทโครงการ	ระบบอยู่ในระยะใดของวงจรชีวิต AI เกี่ยวข้องกับการจัดซื้อจัดจ้าง หรือเป็นการขยาย ดัดแปลงจากระบบเดิมหรือไม่ หากใช่ ระบบเดิมเคยได้รับการประเมินหรือไม่ และมีส่วนใดเปลี่ยนแปลงไปจากเดิม
ผู้ใช้งานและผู้ที่เกี่ยวข้องที่ได้รับผลกระทบ	ใครเป็นผู้ใช้งานโดยตรง ใครอาจได้รับผลกระทบจากผลลัพธ์ของระบบ จำนวนผู้ใช้งานโดยตรงและผู้ที่เกี่ยวข้องได้รับผลกระทบโดยประมาณเป็นเท่าใด และผู้ใช้งานสามารถเลือกไม่ใช้ระบบได้หรือไม่ และผู้ใช้งานมีความรู้ ทักษะ หรือได้รับการฝึกอบรมที่จำเป็นต่อการใช้งานระบบอย่างเหมาะสมเพียงใด
ภาคส่วนและระดับความสำคัญของการใช้งาน	ระบบถูกนำไปใช้ในภาคส่วนและกระบวนการงานใด มีขอบเขตการใช้งานระดับใด และเกี่ยวข้องกับโครงสร้างพื้นฐานหรือภารกิจสำคัญหรือไม่
ปัจจัยภายนอกและการเชื่อมต่อกับระบบอื่น	ระบบพึ่งพาโมเดล ข้อมูล ซอฟต์แวร์ โครงสร้างพื้นฐาน หรือระบบอื่นที่อาจส่งผลกระทบต่อพฤติกรรม ความเสี่ยง หรือผลกระทบของระบบหรือไม่

แนวทางการประเมิน

ในการประเมินส่วนนี้ ควรอธิบายระบบ AI ด้วยภาษาที่เข้าใจง่าย โดยไม่จำกัดอยู่เพียงชื่อระบบหรือเทคโนโลยีที่ใช้ แต่ควรทำให้เห็นวัตถุประสงค์ บทบาทของระบบ และบริบทการใช้งานจริง ควรแยกให้ชัดเจนระหว่าง ผู้ใช้งาน ซึ่งเป็นผู้ใช้งานระบบโดยตรง และ ผู้ที่อาจได้รับผลกระทบ ซึ่งอาจไม่ได้ใช้งานระบบเอง แต่ผลลัพธ์หรือข้อเสนอแนะจากระบบอาจส่งผลกระทบต่อสิทธิ โอกาส การเข้าถึงบริการ หรือการได้รับการปฏิบัติของบุคคลเหล่านั้น นอกจากนี้ ควรพิจารณาปัจจัยแวดล้อมที่เกี่ยวข้อง เช่น ระบบเดิมที่มีอยู่ การจัดซื้อจัดจ้าง แหล่งข้อมูล โมเดลภายนอก ซอฟต์แวร์ โครงสร้างพื้นฐาน หรือระบบอื่นที่เชื่อมต่อกัน เนื่องจากปัจจัยเหล่านี้อาจส่งผลกระทบต่อความเสี่ยงและผลกระทบของระบบ AI ได้

ข้อมูลในส่วนนี้ควรช่วยให้เห็นภาพว่า ระบบ AI นี้ใช้เพื่ออะไร ใช้กับใคร ใครอาจได้รับผลกระทบ และมีขอบเขตหรือข้อจำกัดใดที่ต้องนำไปพิจารณาต่อการประเมิน

2.2.2 กลไกการกำกับดูแลโครงการ AI (Project Governance)

ส่วนนี้มีวัตถุประสงค์เพื่อพิจารณาว่าโครงการ AI มีโครงสร้างการกำกับดูแล บทบาท ความรับผิดชอบ และกลไกการตัดสินใจที่ชัดเจนหรือไม่ เนื่องจากการพัฒนาและใช้งาน AI อย่างมีความรับผิดชอบไม่ได้ขึ้นอยู่กับความสามารถของระบบเพียงอย่างเดียว แต่ยังขึ้นอยู่กับกำกับการกำกับดูแลขององค์กรตลอดวงจรชีวิตของระบบ AI

ข้อมูลในส่วนนี้จะช่วยให้เห็นว่าใครเป็นผู้มีอำนาจตัดสินใจสูงสุด ใครรับผิดชอบด้านเทคนิค การใช้งาน การกำกับดูแล ความเสี่ยง และการปฏิบัติตามข้อกำหนด รวมถึงทีมโครงการมีมุมมองที่หลากหลายเพียงพอในการระบุและเข้าใจผลกระทบที่อาจเกิดขึ้นกับผู้ใช้งานและผู้มีส่วนได้ส่วนเสียหรือไม่ โดยมีประเด็นคำถามดังนี้

ตารางที่ 2-2 ประเด็นคำถามสำคัญสำหรับกลไกการกำกับดูแลโครงการ AI (Project Governance)

กลุ่มประเด็น	รายการที่ควรพิจารณา
ผู้รับผิดชอบและอำนาจตัดสินใจ	ระบุผู้มีอำนาจตัดสินใจสูงสุดของระบบ AI รวมถึงบทบาท ขอบเขตความรับผิดชอบ และประเด็นที่แต่ละคนรับผิดชอบ เช่น ด้านเทคนิค การใช้งาน การกำกับดูแล ความเสี่ยง หรือการปฏิบัติตามข้อกำหนด
ความหลากหลายของทีมโครงการ	พิจารณาว่าทีมโครงการมีความหลากหลายเพียงพอที่จะสะท้อนความซับซ้อนและความหลากหลายของผู้ใช้งานและผู้ที่ได้รับผลกระทบหรือไม่ รวมถึงมุมมองใดที่ยังขาดหาย และอาจก่อให้เกิดอคติอย่างไร
กลไกทบทวนความรับผิดชอบและประเด็นด้านจริยธรรม	พิจารณาว่ามีบุคคล คณะกรรมการ หรือหน่วยงานที่รับผิดชอบในการทบทวนประเด็นด้านความรับผิดชอบ ความโปร่งใส และจริยธรรมของระบบ AI หรือไม่ และข้อเสนอแนะถูกนำไปดำเนินการอย่างไร

แนวทางการประเมิน

ในการประเมินส่วนนี้ ควรระบุให้ชัดเจนว่าใครเป็นผู้รับผิดชอบหลักของระบบ AI และใครมีอำนาจตัดสินใจในประเด็นสำคัญ โดยเฉพาะประเด็นที่เกี่ยวข้องกับการออกแบบระบบ การนำไปใช้งาน การบริหารความเสี่ยง การคุ้มครองข้อมูล และการกำกับดูแลด้านจริยธรรม โดยควรพิจารณาด้วยว่าความรับผิดชอบของแต่ละฝ่ายมีความชัดเจนเพียงใด หากมีผู้รับผิดชอบหลายฝ่าย ควรอธิบายว่าฝ่ายใดรับผิดชอบเรื่องใด เพื่อป้องกันช่องว่างในการกำกับดูแลหรือความไม่ชัดเจนเมื่อเกิดข้อผิดพลาดหรือผลกระทบจากระบบ AI นอกจากนี้ ควรพิจารณาความหลากหลายของทีมโครงการ ทั้งในด้านความเชี่ยวชาญ ประสบการณ์ และความเข้าใจต่อบริบทของผู้ใช้งานหรือผู้ที่อาจได้รับผลกระทบ เนื่องจากทีมที่มีมุมมองจำกัดอาจมองข้ามความเสี่ยงบางประการ หรือไม่สามารถระบุผลกระทบที่เกิดขึ้นกับกลุ่มผู้ใช้งานที่หลากหลายได้อย่างครบถ้วน

จึงสรุปได้ว่า ข้อมูลในส่วนนี้ควรช่วยให้เห็นภาพว่า ระบบ AI นี้มีใครรับผิดชอบ ใครมีอำนาจตัดสินใจ มีกลไกใดในการทบทวนประเด็นด้านจริยธรรม และโครงสร้างการกำกับดูแลของโครงการเพียงพอต่อการพัฒนาและใช้งาน AI อย่างมีความรับผิดชอบหรือไม่

2.2.3 การพิจารณาความสมเหตุสมผลของการใช้ AI และหลักการไม่ก่อให้เกิดอันตราย (Proportionality Screening and Do No Harm)

ส่วนนี้มีวัตถุประสงค์เพื่อพิจารณาว่าการนำ AI มาใช้มีความจำเป็น เหมาะสม และได้สัดส่วนกับวัตถุประสงค์ของโครงการหรือไม่ รวมถึงประเมินว่าระบบ AI อาจก่อให้เกิดอันตรายหรือผลกระทบเชิงลบต่อบุคคล กลุ่มบุคคล สังคม สิทธิมนุษยชน ศักดิ์ศรีความเป็นมนุษย์ ความปลอดภัย หรือเสรีภาพขั้นพื้นฐานหรือไม่ ข้อมูลในส่วนนี้ช่วยให้องค์กรไม่พิจารณาเพียงว่า AI สามารถทำงานได้หรือไม่ แต่ยังพิจารณาว่า AI เป็นทางเลือกที่เหมาะสมที่สุดหรือไม่ มีการจำกัดขอบเขตการใช้งานอย่างเพียงพอหรือไม่ และมีมาตรการป้องกันไม่ให้ระบบถูกนำไปใช้ในลักษณะที่อาจก่อให้เกิดความเสียหายหรือกระทบต่อสิทธิของบุคคลอย่างไม่เหมาะสมหรือไม่ โดยมีประเด็นคำถามดังนี้

ตารางที่ 2-3 ประเด็นคำถามสำคัญสำหรับการพิจารณาความสมเหตุสมผลของการใช้ AI และหลักการไม่ก่อให้เกิดอันตราย (Proportionality Screening and Do No Harm)

กลุ่มประเด็น	รายการที่ควรพิจารณา
ทางเลือกที่ไม่ใช้ AI	มีการพิจารณาทางเลือกอื่นที่ไม่ใช้ AI หรือไม่ เช่น การปรับปรุงกระบวนการเดิม การใช้ระบบสารสนเทศทั่วไป หรือการเพิ่มการพิจารณาโดยมนุษย์
ขอบเขตและข้อจำกัดของระบบ AI	มีการกำหนดขอบเขตการใช้งานของระบบอย่างไร เพื่อให้การใช้ AI สอดคล้องกับวัตถุประสงค์และไม่เกินความจำเป็น

กลุ่มประเด็น	รายการที่ควรพิจารณา
การป้องกันการใช้ AI ในลักษณะที่อาจก่อให้เกิดอันตราย	ระบบมีวัตถุประสงค์หรือความสามารถที่อาจเกี่ยวข้องกับการให้คะแนนทางสังคม การเฝ้าระวังในวงกว้าง หรือการใช้ข้อมูลในลักษณะที่กระทบต่อสิทธิของบุคคลหรือไม่
ผลกระทบต่อสิทธิมนุษยชน ศักดิ์ศรี และเสรีภาพขั้นพื้นฐาน	ระบบอาจก่อให้เกิดผลกระทบต่อความเป็นส่วนตัว ความเป็นธรรม เสรีภาพในการแสดงออก กระบวนการที่เป็นธรรม สุขภาพ ความปลอดภัย หรือความเป็นอยู่ที่ดีของบุคคลหรือไม่
ผลกระทบที่รุนแรงหรือยากต่อการแก้ไข	ผลลัพธ์ของระบบอาจนำไปสู่ผลกระทบที่ย้อนกลับได้ยาก หรือเกี่ยวข้องกับการตัดสินใจที่ส่งผลต่อชีวิต ความปลอดภัย หรือสิทธิสำคัญของบุคคลหรือไม่

แนวทางการประเมิน

ในการประเมินส่วนนี้ ควรเริ่มจากการพิจารณาว่าเหตุใดจึงจำเป็นต้องใช้ AI และมีทางเลือกอื่นที่เหมาะสมกว่าหรือมีความเสี่ยงต่ำกว่าหรือไม่ การเลือกใช้ AI ควรสอดคล้องกับวัตถุประสงค์ของโครงการ และไม่ควรขยายขอบเขตการใช้งานเกินกว่าความจำเป็น ควรกำหนดข้อจำกัดของระบบให้ชัดเจน เช่น ใช้กับกลุ่มผู้ใช้งานใด ใช้ในบริบทใด ใช้ข้อมูลประเภทใด และผลลัพธ์ของระบบมีบทบาทต่อการตัดสินใจมากน้อยเพียงใด โดยเฉพาะในกรณีที่ระบบ AI อาจส่งผลต่อสิทธิ โอกาส การเข้าถึงบริการ หรือการได้รับการปฏิบัติของบุคคล

นอกจากนี้ ควรพิจารณาว่าระบบ AI อาจถูกนำไปใช้ในลักษณะที่ก่อให้เกิดอันตรายหรือกระทบต่อสิทธิมนุษยชนหรือไม่ เช่น การเลือกปฏิบัติ การละเมิดความเป็นส่วนตัว การเฝ้าระวังในวงกว้าง การให้คะแนนหรือจัดลำดับบุคคล การลดทอนบทบาทของมนุษย์ในการตัดสินใจ หรือการสร้างผลกระทบที่ผู้ได้รับผลกระทบไม่สามารถโต้แย้งหรือแก้ไขได้

โดยสรุป ข้อมูลในส่วนนี้ควรช่วยให้องค์กรเห็นว่า การใช้ AI ในโครงการนี้มีความจำเป็นและได้สัดส่วนหรือไม่ มีความเสี่ยงต่อการก่อให้เกิดอันตรายในลักษณะใด และควรมีมาตรการใดเพื่อป้องกันหรือลดผลกระทบที่อาจเกิดขึ้น

2.2.4 การมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย (Stakeholder Engagement)

ส่วนนี้มีวัตถุประสงค์เพื่อพิจารณาว่าองค์กรได้เปิดโอกาสให้ผู้มีส่วนได้ส่วนเสียที่อาจได้รับผลกระทบจากระบบ AI เข้ามามีส่วนร่วมในกระบวนการออกแบบ พัฒนา นำไปใช้งาน และติดตามผลหรือไม่ รวมถึงพิจารณาว่าการมีส่วนร่วมนั้นเกิดขึ้นเพื่อวัตถุประสงค์ใด และข้อมูลหรือข้อเสนอแนะที่ได้รับถูกนำไปใช้ประกอบการตัดสินใจอย่างไร ข้อมูลในส่วนนี้ช่วยให้องค์กรไม่พิจารณาระบบ AI จากมุมมองของผู้พัฒนาเจ้าของระบบ หรือผู้ใช้งานภายในองค์กรเพียงฝ่ายเดียว แต่คำนึงถึงมุมมองของผู้ที่อาจได้รับผลกระทบโดยตรงหรือโดยอ้อม โดยเฉพาะกลุ่มเปราะบาง กลุ่มชายขอบ หรือกลุ่มที่อาจได้รับผลกระทบมากกว่ากลุ่มอื่น โดยมีประเด็นคำถามดังนี้

ตารางที่ 2-4 ประเด็นคำถามสำคัญสำหรับการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสีย (Stakeholder Engagement)

กลุ่มประเด็น	รายการที่ควรพิจารณา
การระบุและมีส่วนร่วมกับผู้มีส่วนได้ส่วนเสีย	ผู้มีส่วนได้ส่วนเสียที่อาจได้รับผลกระทบจากระบบ AI คือใคร และมีการรับฟังหรือปรึกษาหารือในช่วงใดของวงจรชีวิตระบบ AI
วัตถุประสงค์ของการมีส่วนร่วม	การมีส่วนร่วมมีเป้าหมายเพื่ออะไร เช่น เพิ่มความถูกต้องของระบบ ส่งเสริมความเป็นธรรม ระบุและลดความเสี่ยง หรือสร้างความไว้วางใจต่อระบบ AI
การนำข้อคิดเห็นไปใช้	มีกลไกใดในการรวบรวม พิจารณา และนำเสนอแนะจากผู้มีส่วนได้ส่วนเสีย ไปใช้ในการออกแบบ พัฒนา ปรับปรุง หรือกำกับดูแลระบบ AI รวมถึงแจ้งผลการนำข้อคิดเห็นไปใช้แก่ผู้มีส่วนได้ส่วนเสียอย่างไร

แนวทางการประเมิน

ในการประเมินส่วนนี้ ควรเริ่มจากการระบุว่าผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้องกับระบบ AI มีใครบ้าง โดยไม่จำกัดเฉพาะผู้ใช้งานระบบโดยตรง แต่ควรรวมถึงผู้ที่อาจได้รับผลกระทบจากผลลัพธ์ ข้อเสนอแนะ หรือการตัดสินใจที่ใช้ระบบ AI เป็นส่วนหนึ่งของกระบวนการ ควรพิจารณาว่าผู้มีส่วนได้ส่วนเสียเหล่านี้ได้รับการมีส่วนร่วมในช่วงใดของวงจรชีวิตระบบ AI เช่น ระยะออกแบบ ระยะพัฒนา ระยะทดสอบ ระยะนำไปใช้งาน หรือระยะติดตามและประเมินผล หากยังไม่มีส่วนร่วม ควรระบุแผนการดำเนินการในอนาคตให้ชัดเจน

นอกจากนี้ ควรพิจารณาว่าการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสียมีเป้าหมายเพื่ออะไร และข้อคิดเห็นที่ได้รับถูกนำไปใช้จริงหรือไม่ เช่น ใช้ปรับปรุงการออกแบบระบบ เพิ่มมาตรการลดความเสี่ยง ทบทวนผลกระทบต่อกลุ่มเปราะบาง หรือกำหนดกลไกสื่อสารและรับเรื่องร้องเรียน

โดยสรุป ข้อมูลในส่วนนี้ควรช่วยให้องค์กรเห็นว่า ผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้องกับระบบ AI ได้รับการรับฟังอย่างเหมาะสมหรือไม่ การมีส่วนร่วมนั้นช่วยลดความเสี่ยงหรือเพิ่มความเป็นธรรมของระบบอย่างไร และข้อเสนอแนะที่ได้รับถูกนำไปใช้ประกอบการตัดสินใจอย่างเป็นรูปธรรมหรือไม่

ผลลัพธ์ทั้งหมดจากการกำหนดขอบเขตการประเมิน (Scoping Questions) จะถูกนำไปใช้เป็นข้อมูลตั้งต้นสำหรับการประเมินตามหลักการด้านจริยธรรมและการวิเคราะห์ผลกระทบทางจริยธรรมในขั้นตอนถัดไป ดังนั้น การตอบคำถามในขั้นตอนนี้จึงควรมีความครบถ้วน ชัดเจน และสะท้อนบริบทของการใช้งานจริงให้มากที่สุด เพื่อให้องค์กรสามารถระบุประเด็นสำคัญ ความเสี่ยง ช่องว่าง และมาตรการกำกับดูแลที่เหมาะสมได้อย่างมีประสิทธิภาพ

บทที่ 3

ความปลอดภัย และ ความมั่นคงปลอดภัย



มิติด้านความปลอดภัยและความมั่นคงปลอดภัย (Safety and Security) มีวัตถุประสงค์เพื่อประเมินว่าระบบ AI มีมาตรการที่เหมาะสมในการป้องกัน ตรวจสอบ ติดตาม และตอบสนองต่อความเสี่ยงที่อาจเกิดขึ้นจากการใช้งานระบบหรือไม่ โดยครอบคลุมทั้งความเสี่ยงจากการใช้งานผิดวัตถุประสงค์ การโจมตีหรือแทรกแซงระบบ ความผิดพลาดของระบบ การเข้าถึงข้อมูลโดยไม่ได้รับอนุญาต รวมถึงเหตุการณ์ด้านความปลอดภัยที่อาจส่งผลกระทบต่อผู้ใช้งาน ผู้ที่เกี่ยวข้อง หรือข้อมูลที่ระบบประมวลผล

การประเมินในมิตินี้จึงควรมองให้ไกลกว่าการตรวจสอบว่าระบบ AI ทำงานได้ตามที่ออกแบบไว้หรือไม่ แต่ควรพิจารณาด้วยว่าองค์กรมีความพร้อมเพียงใดในการดูแลความปลอดภัยของระบบตลอดการใช้งานจริง ทั้งในด้านการป้องกันความเสี่ยง การติดตามความผิดปกติ การทดสอบระบบอย่างต่อเนื่อง รวมถึงการแจ้งเตือนและรับมือเมื่อเกิดเหตุการณ์ที่อาจกระทบต่อความปลอดภัยของระบบ ข้อมูล หรือผู้ใช้งาน

ตัวอย่างความเสี่ยงที่ควรคำนึงถึงในมิตินี้ ได้แก่ การป้อนข้อมูลที่เป็นอันตรายต่อระบบ การเปลี่ยนแปลงหรือปนเปื้อนข้อมูลที่ใช้ในการพัฒนาและปรับปรุงระบบ การแทรกแซงการทำงานของระบบ การทำให้ระบบสร้างผลลัพธ์ที่คลาดเคลื่อน การเข้าถึงข้อมูลโดยไม่ได้รับอนุญาต การรั่วไหลของข้อมูล หรือการใช้ระบบในลักษณะที่อาจก่อให้เกิดอันตรายต่อบุคคลหรือองค์กร

จากความเสี่ยงดังกล่าว การประเมินด้าน Safety and Security จึงควรพิจารณาทั้งมาตรการป้องกัน กลไกการรายงานและตอบสนองต่อเหตุการณ์ การทดสอบระบบ และผลกระทบที่อาจเกิดขึ้น โดยสามารถสรุปเป็นประเด็นสำคัญได้ดังตารางต่อไปนี้

ตารางที่ 3-1 รายการคำถามประเมินด้านความปลอดภัยและความมั่นคงปลอดภัย
(Safety and Security)

รายการคำถาม	แนวทางประกอบการประเมิน
มีกลไกใดในการประเมินและติดตามความเสี่ยงจากการใช้งานที่ไม่เหมาะสม การใช้งานโดยมิชอบ หรือการโจมตีระบบ AI หรือไม่	ระบุกลไกที่ใช้ประเมิน ติดตาม และป้องกันความเสี่ยงจากการใช้งานระบบในทางที่ไม่เหมาะสม การใช้งานโดยมิชอบ การโจมตี การแทรกแซง หรือการเจาะระบบ เช่น การควบคุมสิทธิ์การเข้าถึง การติดตามพฤติกรรมผิดปกติ การตรวจสอบการใช้งาน หรือการทดสอบช่องโหว่
มีกลไกให้ผู้ใช้งานรายงานข้อผิดพลาดหรือผลกระทบที่เกิดจากระบบ AI หรือไม่	ระบุช่องทางที่ผู้ใช้งานสามารถแจ้งข้อผิดพลาด ความผิดปกติ หรือปัญหาที่พบจากการใช้งานระบบ เช่น แบบฟอร์มออนไลน์ อีเมล Helpdesk ระบบ Ticket หรือช่องทางติดต่อเจ้าหน้าที่ รวมถึงกระบวนการรับเรื่องและติดตามผล
ระบบ AI ปฏิบัติตามมาตรฐานหรือกรอบด้านความมั่นคงปลอดภัยไซเบอร์ใด	ระบุมาตรฐาน กรอบการดำเนินงาน หรือแนวปฏิบัติด้านความมั่นคงปลอดภัยไซเบอร์ที่เกี่ยวข้องกับระบบ เช่น มาตรฐานภายในองค์กร มาตรฐานอุตสาหกรรม หรือแนวทางด้าน Cybersecurity ที่ใช้ในการคุ้มครองระบบ ข้อมูล และผู้ใช้งาน
มีกลไกใดในการตรวจสอบและตอบสนองต่อเหตุการณ์ด้าน	ระบุกระบวนการตรวจสอบ แจ้งเตือน ตอบสนอง และจัดการเหตุการณ์ที่เกี่ยวข้องกับความปลอดภัยหรือความมั่นคง

รายการคำถาม	แนวทางประกอบการประเมิน
ความปลอดภัยและความมั่นคงปลอดภัย	ปลอดภัยของระบบ AI เช่น การเฝ้าระวังระบบ การบันทึกเหตุการณ์ การกำหนดผู้รับผิดชอบเมื่อเกิดเหตุ และขั้นตอนการแก้ไขหรือจำกัดผลกระทบ
มีการแจ้งผู้ใช้งานเกี่ยวกับเหตุการณ์ด้านความปลอดภัยและความมั่นคงปลอดภัยที่อาจเกี่ยวข้องกับผู้ใช้งานหรือข้อมูลของผู้ใช้งานอย่างไร	ระบุวิธีการสื่อสารกับผู้ใช้งานเมื่อเกิดเหตุการณ์ที่อาจกระทบต่อระบบหรือข้อมูลของผู้ใช้งาน เช่น การแจ้งเตือนผ่านอีเมล ระบบประกาศภายใน หน้าเว็บไซต์ หรือช่องทางติดต่อโดยตรง รวมถึงเงื่อนไขว่าเหตุการณ์ประเภทใดต้องแจ้งผู้ใช้งาน
ระบบ AI ได้ผ่านการทดสอบด้านความปลอดภัยอย่างเข้มงวดและสม่ำเสมอหรือไม่	ระบุว่าระบบเคยผ่านการทดสอบด้านความปลอดภัยหรือไม่ ทดสอบในประเด็นใดบ้าง เช่น ความปลอดภัยของระบบ ความถูกต้องของผลลัพธ์ ความทนทานต่อการใช้งานผิดปกติ หรือความเสี่ยงจากการเปลี่ยนแปลงของโมเดลและข้อมูล
จะมีการทดสอบและทวนสอบระบบ AI บ่อยเพียงใด	ระบุความถี่ในการทดสอบ ทวนสอบ หรือประเมินระบบซ้ำ เช่น ก่อนนำไปใช้งานจริง หลังการปรับปรุงระบบ เป็นรายไตรมาส รายปี หรือเมื่อมีการเปลี่ยนแปลงของข้อมูล โมเดล สภาพแวดล้อม หรือบริบทการใช้งาน
มีมาตรการเพิ่มเติมเพื่อรักษาความปลอดภัยและความมั่นคงปลอดภัยของระบบ AI และข้อมูลหรือไม่	ระบุมาตรการเพิ่มเติมที่ใช้คุ้มครองระบบ AI และข้อมูล เช่น การจำกัดสิทธิ์การเข้าถึง การเข้ารหัสข้อมูล การสำรองข้อมูล การติดตามความผิดปกติ การตรวจสอบการใช้งาน หรือมาตรการลดความเสี่ยงอื่นที่เหมาะสมกับบริบทของระบบ
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความปลอดภัยและความมั่นคงปลอดภัยโดยย่อ	สรุปทั้งประโยชน์และความเสี่ยงด้านความปลอดภัยและความมั่นคงปลอดภัย เช่น ระบบช่วยลดข้อผิดพลาด เพิ่มการตรวจจับความผิดปกติ หรือเพิ่มความปลอดภัยของกระบวนการได้อย่างไร ขณะเดียวกันอาจมีความเสี่ยงจากการโจมตี การรั่วไหลของข้อมูล การใช้งานผิดพลาดวัตถุประสงค์ หรือความผิดพลาดของระบบในลักษณะใด

หมายเหตุ: แนวทางในตารางเป็นข้อเสนอประกอบการประเมิน ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับบริบทของระบบ AI โดยระบุเฉพาะข้อมูลที่เกี่ยวข้องกับการใช้งานจริงของระบบนั้น ๆ

แนวทางการพิจารณา

การประเมินในมิตินี้ควรครอบคลุมตลอดวงจรชีวิตของระบบ AI ตั้งแต่การออกแบบและพัฒนา การทดสอบ การนำไปใช้งาน จนถึงการติดตามและปรับปรุงระบบ โดยควรติดตามและทดสอบระบบอย่างต่อเนื่อง เพื่อให้สามารถตรวจพบความผิดปกติ การเสื่อมลงของประสิทธิภาพ หรือความเสี่ยงที่เกิดขึ้นภายหลังการใช้งาน

นอกจากการตรวจสอบว่ามีมาตรการป้องกันหรือไม่ องค์กรควรพิจารณาด้วยว่ามาตรการดังกล่าวถูกนำไปใช้จริงและเหมาะสมกับระดับความเสี่ยงของระบบหรือไม่ โดยควรกำหนดผู้รับผิดชอบ กลไก

ตรวจจับและแจ้งเตือน ขั้นตอนการจำกัดผลกระทบ การแก้ไขและฟื้นฟูระบบ ตลอดจนการสื่อสารกับ
ผู้ใช้งานหรือผู้ที่เกี่ยวข้องให้ชัดเจน

การประเมินควรอ้างอิงข้อมูลหรือหลักฐานที่มีอยู่ เช่น ผลการทดสอบ บันทึกเหตุการณ์ รายงาน
ข้อผิดพลาด และผลการทวนสอบระบบ หากข้อมูลหรือมาตรการบางส่วนยังไม่พร้อม ควรระบุเป็น
ช่องว่าง พร้อมกำหนดผู้รับผิดชอบและแนวทางในการปรับปรุงหรือติดตามเพิ่มเติม

โดยสรุป ข้อมูลในมิตินี้ควรช่วยให้องค์กรเห็นว่า ระบบ AI มีมาตรการเพียงพอในการป้องกัน
ตรวจจับ ตอบสนอง และลดผลกระทบจากความเสี่ยงด้านความปลอดภัยและความมั่นคงปลอดภัยหรือไม่
รวมถึงมีจุดแข็ง ช่องว่าง หรือประเด็นใดที่ควรนำไปวิเคราะห์ต่อในขั้นตอนการวิเคราะห์ผลกระทบทาง
จริยธรรม (Impact Mapping)

บทที่ 4

ความเป็นธรรม การไม่เลือกปฏิบัติ

และ ความหลากหลาย



มิติด้านความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย (Fairness, Non-discrimination and Diversity) มีวัตถุประสงค์เพื่อประเมินว่าระบบ AI ได้รับการออกแบบ พัฒนา และทดสอบโดยคำนึงถึงความเสี่ยงด้านอคติ ความไม่เป็นธรรม และการเลือกปฏิบัติต่อบุคคลหรือกลุ่มบุคคลที่แตกต่างกันหรือไม่ โดยพิจารณาทั้งข้อมูลที่ใช้พัฒนาระบบ โมเดล ผลลัพธ์ และการนำระบบไปใช้งาน

การประเมินในมิตินี้จึงไม่ควรพิจารณาเพียงว่าระบบ AI ให้ผลลัพธ์ที่ถูกต้องเท่านั้น แต่ควรพิจารณาด้วยว่าข้อมูลครอบคลุมกลุ่มเป้าหมายอย่างเพียงพอหรือไม่ และผลลัพธ์ของระบบอาจสร้างความได้เปรียบ เสียเปรียบ หรือผลกระทบที่แตกต่างกันต่อผู้ใช้งานและผู้ที่เกี่ยวข้องได้รับผลกระทบแต่ละกลุ่มหรือไม่ นอกจากนี้ ระบบควรได้รับการออกแบบให้ผู้ใช้งานที่หลากหลายสามารถเข้าถึงและใช้งานได้ โดยไม่สร้างอุปสรรคเพิ่มเติมจากความพิการ ภาษา อายุ ทักษะดิจิทัล หรือข้อจำกัดด้านอุปกรณ์

ตัวอย่างความเสี่ยงที่ควรคำนึงถึงในมิตินี้ ได้แก่ ข้อมูลฝึกสอนที่ไม่ครอบคลุมกลุ่มผู้ใช้งานบางกลุ่ม ข้อมูลที่มีความคลาดเคลื่อนหรือสะท้อนอคติที่มีอยู่เดิม และผลลัพธ์ของระบบที่อาจก่อให้เกิดการเลือกปฏิบัติต่อผู้ใช้งานบางกลุ่ม รวมถึงการออกแบบระบบที่ไม่รองรับผู้พิการ ผู้สูงอายุ ผู้ใช้ภาษาที่แตกต่างกัน หรือผู้ที่มีข้อจำกัดด้านทักษะดิจิทัลและอุปกรณ์

จากความเสี่ยงดังกล่าว การประเมินด้าน Fairness, Non-discrimination and Diversity จึงควรพิจารณาทั้งคุณภาพและความเป็นตัวแทนของข้อมูล การทดสอบอคติของระบบ ความสามารถในการเข้าถึงและใช้งานของผู้ใช้ที่หลากหลาย ตลอดจนผลกระทบเชิงบวกและเชิงลบที่อาจเกิดขึ้น โดยสามารถสรุปเป็นรายการประเมินได้ดังตารางต่อไปนี้

ตารางที่ 4-1 รายการคำถามประเมินด้านความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย (Fairness, Non-discrimination and Diversity)

รายการคำถาม	แนวทางประกอบการประเมิน
ข้อมูลที่ใช้ฝึกสอนระบบ AI ได้รับการทดสอบด้านอคติและความเป็นตัวแทนของกลุ่มเป้าหมายอย่างไร	ระบุวิธีการตรวจสอบว่าข้อมูลที่ใช้ฝึกสอนระบบมีอคติหรือไม่ และสะท้อนความหลากหลายของกลุ่มผู้ใช้งานหรือบริบทการใช้งานจริงเพียงใด เช่น การตรวจสอบเชิงสถิติ การประเมินความครอบคลุมของกลุ่มข้อมูล การสุ่มตัวอย่าง การตรวจสอบคุณภาพข้อมูล หรือการใช้เครื่องมือตรวจจับอคติ
ระบบ AI ได้รับการทดสอบด้านอคติอย่างไร	ระบุวิธีการทดสอบว่าระบบ โมเดล หรือผลลัพธ์ของระบบอาจก่อให้เกิดอคติหรือผลลัพธ์ที่เลือกปฏิบัติต่อผู้ใช้งานบางกลุ่มหรือไม่ เช่น การทำ fairness audit การทดสอบสถานการณ์จำลอง การวิเคราะห์ผลลัพธ์แยกตามกลุ่มผู้ใช้งาน หรือมาตรการลดอคติที่นำมาใช้
ระบบ AI สามารถใช้งานได้และเข้าถึงได้สำหรับทุกกลุ่มหรือไม่	ระบุว่าระบบได้รับการออกแบบให้ผู้ใช้งานที่หลากหลายสามารถเข้าถึงและใช้งานได้หรือไม่ เช่น การรองรับผู้พิการ การใช้ภาษาที่เข้าใจง่าย การรองรับผู้ที่มีทักษะดิจิทัลแตกต่างกัน การลดข้อจำกัดด้านอุปกรณ์ การออกแบบที่เป็นมิตรต่อผู้สูงอายุ หรือการใช้สีและองค์ประกอบที่ไม่เป็นอุปสรรคต่อผู้ใช้
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความเป็น	สรุปทั้งประโยชน์และความเสี่ยงในมิตินี้ เช่น ระบบช่วยเพิ่มการเข้าถึงบริการ ลดความแตกต่างของการให้บริการ หรือสนับสนุน

รายการคำถาม	แนวทางประกอบการประเมิน
ธรรม การไม่เลือกปฏิบัติ และความหลากหลายโดยย่อ	การตัดสินใจที่สม่ำเสมอได้อย่างไร ขณะเดียวกันอาจมีความเสี่ยงจากข้อมูลที่ไม่ครอบคลุม อคติของผลลัพธ์ ข้อจำกัดในการเข้าถึง หรือผลกระทบต่อกลุ่มเปราะบางในลักษณะใด

หมายเหตุ: แนวทางในตารางเป็นข้อเสนอประกอบการประเมิน ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับบริบทของระบบ AI โดยระบุเฉพาะข้อมูลที่เกี่ยวข้องกับการใช้งานจริงของระบบนั้น ๆ

แนวทางการพิจารณา

การประเมินในมิตินี้ควรพิจารณาทั้งข้อมูลที่ใช้พัฒนาระบบ กระบวนการออกแบบและทดสอบ ตลอดจนผลลัพธ์ที่เกิดขึ้นเมื่อระบบถูกนำไปใช้งานจริง เนื่องจากอคติอาจไม่ได้เกิดจากข้อมูลเพียงอย่างเดียว แต่อาจเกี่ยวข้องกับวิธีการกำหนดปัญหา การออกแบบระบบ การตีความผลลัพธ์ หรือบริบทที่ระบบถูกนำไปใช้ด้วย นอกจากการตรวจสอบผลลัพธ์โดยรวมแล้ว ควรพิจารณาผลลัพธ์แยกตามกลุ่มผู้ใช้งาน หรือผู้ที่อาจได้รับผลกระทบ เพื่อให้เห็นว่าระบบสร้างผลกระทบที่แตกต่างกันระหว่างกลุ่มหรือไม่ ขณะเดียวกัน ควรตรวจสอบว่าการออกแบบและช่องทางการใช้งานไม่สร้างอุปสรรคต่อผู้ใช้งานบางกลุ่ม เช่น การใช้สีเป็นช่องทางเดียวในการสื่อความหมาย การไม่มีข้อความทดแทนสำหรับภาพ หรือการออกแบบที่ไม่รองรับเทคโนโลยีช่วยเหลือ

การประเมินควรอ้างอิงข้อมูลหรือหลักฐานที่มีอยู่ เช่น ผลการตรวจสอบข้อมูล ผลการทดสอบแยกตามกลุ่ม รายงานข้อร้องเรียน หรือผลการประเมินการเข้าถึงระบบ หากยังไม่มีข้อมูลหรือยังไม่เคยดำเนินการในบางประเด็น ควรระบุเป็นช่องว่าง พร้อมกำหนดแนวทางและผู้รับผิดชอบในการดำเนินการเพิ่มเติม

เมื่อประเมินมิตินี้แล้ว องค์กรควรสามารถทราบความพร้อมของมาตรการด้านความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย ระบุจุดแข็ง ช่องว่าง และประเด็นที่ควรปรับปรุง รวมถึงสรุปผลกระทบเชิงบวกและเชิงลบที่ควรนำไปวิเคราะห์ต่อในขั้นตอนการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping)

บทที่ 5

ความยั่งยืน



มิติด้านความยั่งยืน (Sustainability) มีวัตถุประสงค์เพื่อประเมินผลกระทบต่อสิ่งแวดล้อมที่เกี่ยวข้องกับระบบ AI โดยครอบคลุมการใช้พลังงานและทรัพยากร การปล่อยก๊าซเรือนกระจก ตลอดจนผลกระทบจากฮาร์ดแวร์และโครงสร้างพื้นฐานที่สนับสนุนการทำงานของระบบ การพิจารณาในมิตินี้ควรครอบคลุมทั้งผลกระทบที่เกิดจากการพัฒนา การใช้งาน และการสิ้นสุดการใช้งานระบบ AI

ในการประเมินโครงการ AI องค์กรควรเริ่มจากการพิจารณาว่าระบบใช้ทรัพยากรในการประมวลผลอย่างไร เช่น องค์กรพัฒนาและฝึกสอนโมเดลเอง หรือใช้โมเดลและบริการคลาวด์จากผู้ให้บริการภายนอก ระบบมีขนาดและความถี่ในการใช้งานเพียงใด รวมถึงมีข้อมูลเกี่ยวกับการใช้พลังงาน การปล่อยก๊าซเรือนกระจก หรือผลกระทบจากศูนย์ข้อมูลและฮาร์ดแวร์ให้ตรวจสอบหรือไม่ เนื่องจากการพัฒนาและใช้งาน AI ต้องอาศัยการประมวลผลและพลังงานจากศูนย์ข้อมูล

ตัวอย่างเช่น โครงการแชทบอตภายในองค์กรที่ใช้โมเดลผ่านบริการคลาวด์ อาจพิจารณาจากจำนวนและความถี่ในการเรียกใช้งาน ข้อมูลด้านสิ่งแวดล้อมที่ผู้ให้บริการเปิดเผย และมาตรการที่องค์กรสามารถดำเนินการได้ เช่น การเลือกใช้โมเดลที่เหมาะสมกับลักษณะงาน การลดการประมวลผลที่ไม่จำเป็น หรือการเลือกใช้โครงสร้างพื้นฐานที่มีประสิทธิภาพด้านพลังงาน ทั้งนี้ ระดับการประเมินควรปรับให้เหมาะสมกับขนาด ลักษณะ และบริบทของระบบ

จากแนวทางดังกล่าว การประเมินด้านความยั่งยืนจึงครอบคลุมการประเมินและติดตามผลกระทบต่อสิ่งแวดล้อม การเปิดเผยข้อมูลต่อสาธารณะ การวัดการใช้พลังงาน ตลอดจนมาตรการลดการใช้พลังงานและคาร์บอนฟุตพริ้นท์ โดยมีรายการประเมินดังตารางต่อไปนี้

ตารางที่ 5-1 รายการคำถามประเมินด้านความยั่งยืน (Sustainability)

รายการคำถาม	แนวทางประกอบการประเมิน
มีการประเมินผลกระทบต่อสิ่งแวดล้อมของระบบ AI หรือไม่	หากมีการประเมิน ให้ระบุประเภทของการประเมิน เช่น การประเมินผลกระทบต่อสิ่งแวดล้อมอย่างเป็นทางการ การวิเคราะห์ตลอดวงจรชีวิต การประเมินคาร์บอนฟุตพริ้นท์ หรือการทบทวนผลกระทบจากการใช้พลังงาน ทรัพยากร และการปล่อยก๊าซเรือนกระจก พร้อมระบุขอบเขต วิธีการ และผลการประเมินที่สำคัญ หากยังไม่มี การประเมิน ให้ระบุเหตุผลและแผนการดำเนินงานในอนาคต
มีการติดตามผลกระทบต่อสิ่งแวดล้อมอย่างสม่ำเสมอหรือไม่	หากมีการติดตาม ให้ระบุวิธีการ ความถี่ และตัวชี้วัดที่ใช้ เช่น การใช้พลังงาน การปล่อยก๊าซเรือนกระจก หรือการใช้ทรัพยากร รวมถึงวิธีการรายงานและการนำผลที่ได้ไปใช้ลดผลกระทบ หากยังไม่มี การติดตาม ให้ระบุเหตุผลและแผนการดำเนินงานในอนาคต
มีการเปิดเผยผลกระทบต่อสิ่งแวดล้อมต่อสาธารณะหรือไม่	ระบุว่ามีการเปิดเผยข้อมูลผ่านรายงานประจำปี ตัวชี้วัดด้านสิ่งแวดล้อม สังคม และธรรมาภิบาล (ESG) การสื่อสารต่อนักลงทุน แลกเปลี่ยนกับสาธารณะ หรือช่องทางอื่นหรือไม่ รวมถึงระบุรูปแบบ ความถี่ ช่องทางการเผยแพร่ และข้อมูลดังกล่าวสามารถเข้าถึงได้โดยสาธารณะหรือไม่
มีการวัดการใช้พลังงานของระบบ AI หรือไม่	หากมีการวัด ให้ระบุวิธีการ ตัวชี้วัด และข้อมูลที่เก็บรวบรวม เช่น ความเข้มข้นของคาร์บอนจากการประมวลผล ผลกระทบตลอดวงจรชีวิตของฮาร์ดแวร์ และความยั่งยืนของศูนย์ข้อมูล รวมถึงขั้นตอนที่ดำเนินการเพื่อลด

รายการคำถาม	แนวทางประกอบการประเมิน
	การใช้พลังงานหรือการปล่อยคาร์บอน หากยังไม่มีการวัด ให้ระบุเหตุผลและแผนการดำเนินงานในอนาคต
มีมาตรการเพื่อลดการใช้พลังงานและคาร์บอนฟุตพริ้นท์ของระบบหรือไม่	ระบุมาตรการที่นำมาใช้ เช่น การเพิ่มประสิทธิภาพของอัลกอริทึมหรือโมเดล การเลือกใช้ฮาร์ดแวร์ที่ประหยัดพลังงาน หรือการใช้พลังงานหมุนเวียน หากยังไม่มีมาตรการ ให้ระบุเหตุผลและแผนการลดการใช้พลังงานหรือการปล่อยคาร์บอนในอนาคต
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความยั่งยืนโดยย่อ	สรุปผลกระทบเชิงบวกและเชิงลบด้านความยั่งยืนที่พบจากการประเมินในข้อก่อนหน้า เพื่อใช้เป็นข้อมูลประกอบการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ในขั้นตอนถัดไป

หมายเหตุ: แนวทางในตารางเป็นข้อเสนอประกอบการประเมิน ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับบริบทของระบบ AI โดยระบุเฉพาะข้อมูลที่เกี่ยวข้องกับการใช้งานจริงของระบบนั้น ๆ

แนวทางการพิจารณา

การประเมินในมิตินี้ไม่จำเป็นต้องมีระดับรายละเอียดเท่ากันในทุกโครงการ แต่ควรพิจารณาให้สอดคล้องกับขนาด ลักษณะการใช้งาน และทรัพยากรที่ระบบใช้ หากองค์กรยังไม่มี การประเมิน การวัด หรือการติดตามอย่างเป็นระบบ ควรระบุเหตุผล ข้อจำกัดของข้อมูล และแผนการดำเนินงานในอนาคตตามความเหมาะสม

นอกจากนี้ ควรแยกให้เห็นว่าผลกระทบส่วนใดอยู่ภายใต้การควบคุมขององค์กรโดยตรง และส่วนใดเกี่ยวข้องกับผู้ให้บริการภายนอก เช่น ผู้ให้บริการโมเดล คลาวด์ หรือศูนย์ข้อมูล หากองค์กรไม่สามารถเข้าถึงข้อมูลบางส่วนได้ ควรระบุเป็นช่องว่างของข้อมูลที่ต้องติดตาม เนื่องจากการประเมินผลกระทบของ AI จำเป็นต้องพิจารณาให้กว้างกว่าการใช้พลังงานระหว่างการดำเนินงานเพียงอย่างเดียว

โดยสรุป ข้อมูลในมิตินี้ควรช่วยให้องค์กรเห็นว่า ระบบ AI มีผลกระทบต่อสิ่งแวดล้อมในลักษณะใด มีข้อมูลและกลไกในการติดตามผลเพียงพอหรือไม่ และมีมาตรการใดในการลดการใช้พลังงาน ทรัพยากร และคาร์บอนฟุตพริ้นท์ของระบบ

บทที่ 6

ความเป็นส่วนตัว และ การคุ้มครอง ข้อมูล



มิติด้านความเป็นส่วนตัวและการคุ้มครองข้อมูล (Privacy and Data Protection) มีวัตถุประสงค์เพื่อประเมินว่าระบบ AI มีการเก็บรวบรวม ใช้ เชื่อมโยง เปิดเผย และจัดเก็บข้อมูลส่วนบุคคลอย่างเหมาะสมหรือไม่ รวมถึงมีมาตรการเพียงพอในการป้องกันการเข้าถึง การสูญหาย การเปลี่ยนแปลง หรือการเปิดเผยข้อมูลโดยไม่ได้รับอนุญาต

การประเมินในมิตินี้ควรเริ่มจากการทำความเข้าใจว่าระบบ AI ใช้ข้อมูลส่วนบุคคลหรือไม่ ใช้ข้อมูลประเภทใด และใช้เพื่อวัตถุประสงค์ใด โดยข้อมูลส่วนบุคคลอาจรวมถึงข้อมูลที่ระบุตัวบุคคลได้โดยตรงหรือโดยอ้อม เช่น ชื่อ ข้อมูลติดต่อ ข้อมูลประชากร ตัวระบุออนไลน์ ข้อมูลพฤติกรรม ภาพ หรือเสียง รวมถึงข้อมูลที่มีความอ่อนไหว ข้อมูลของเด็ก หรือข้อมูลที่ได้รับจากบุคคลภายนอก

นอกจากการคุ้มครองข้อมูลให้ปลอดภัยแล้ว องค์กรควรพิจารณาด้วยว่าระบบใช้ข้อมูลเท่าที่จำเป็นต่อวัตถุประสงค์หรือไม่ มีการเชื่อมโยงข้อมูลจากหลายแหล่งจนสามารถอนุมานข้อมูลใหม่เกี่ยวกับบุคคลหรือไม่ บุคคลสามารถใช้สิทธิเกี่ยวกับข้อมูลของตนได้เพียงพอ และมีการนำหลักการคุ้มครองความเป็นส่วนตัวมาใช้ตั้งแต่ขั้นตอนการออกแบบระบบหรือไม่

จากประเด็นดังกล่าว การประเมินด้าน Privacy and Data Protection จึงครอบคลุมตั้งแต่ประเภทและวัตถุประสงค์ของข้อมูล มาตรการรักษาความปลอดภัย ความจำเป็นในการใช้ข้อมูล การเชื่อมโยงข้อมูล ฐานในการประมวลผล สิทธิของบุคคล การเข้าถึงโดยบุคคลภายนอก ตลอดจนที่มาและความโปร่งใสของชุดข้อมูล โดยมีรายการคำถามประเมินดังตารางต่อไปนี้

ตารางที่ 6-1 รายการคำถามประเมินด้านความเป็นส่วนตัวและการคุ้มครองข้อมูล

(Privacy and Data Protection)

รายการคำถาม	แนวทางประกอบการประเมิน
ระบบ AI ได้รับการฝึกสอนด้วยข้อมูลส่วนบุคคลหรือมีการประมวลผลข้อมูลส่วนบุคคลหรือไม่	ระบุว่าระบบใช้ข้อมูลที่สามารถระบุตัวบุคคลได้โดยตรงหรือโดยอ้อมในขั้นตอนการฝึกสอน ทดสอบ หรือใช้งานจริงหรือไม่ หากระบบใช้เฉพาะข้อมูลที่ไม่ใช่ข้อมูลส่วนบุคคล ข้อมูลที่ไม่สามารถระบุตัวบุคคลได้ หรือข้อมูลสังเคราะห์ ควรระบุให้ชัดเจน
ระบบ AI ได้รับการฝึกสอนหรือประมวลผลข้อมูลส่วนบุคคลประเภทใดบ้าง	ระบุประเภทของข้อมูลส่วนบุคคลที่ระบบใช้ เช่น ชื่อหรือข้อมูลติดต่อ ข้อมูลประชากร ตัวระบุออนไลน์ ข้อมูลพฤติกรรม ภาพ เสียง หรือข้อมูลอื่นที่เกี่ยวข้องกับบุคคล รวมถึงระบุว่าข้อมูลส่วนบุคคลที่มีความอ่อนไหว ข้อมูลของเด็ก หรือข้อมูลจากบุคคลภายนอกหรือไม่
ข้อมูลได้รับการจัดเก็บและประมวลผลโดยมีมาตรการรักษาความปลอดภัยที่เหมาะสมกับระดับความอ่อนไหวหรือไม่	ระบุมาตรการทางเทคนิคและมาตรการขององค์กรที่ใช้ป้องกันการเข้าถึง การสูญหาย การเปลี่ยนแปลง หรือการเปิดเผยข้อมูลโดยไม่ได้รับอนุญาต เช่น การควบคุมสิทธิ์ การเข้ารหัส การจัดเก็บและส่งข้อมูลอย่างปลอดภัย การบันทึกและติดตามการใช้งาน การทดสอบด้านความปลอดภัย และการอบรมบุคลากร โดยควรอธิบายว่ามาตรการแตกต่างกันตามประเภทหรือความอ่อนไหวของข้อมูลอย่างไร

รายการคำถาม	แนวทางประกอบการประเมิน
ระบบ AI ประมวลผลเฉพาะข้อมูล ที่จำเป็นต่อวัตถุประสงค์ที่กำหนด ไว้หรือไม่	ระบุว่าการพิจารณาความเกี่ยวข้องและความจำเป็นของข้อมูลแต่ ละประเภทก่อนนำมาใช้หรือไม่ รวมถึงตรวจสอบว่าระบบไม่ได้เก็บ หรือประมวลผลข้อมูลมากเกินไปจนเกินความจำเป็น ไม่เกี่ยวข้อง หรือไม่ เหตุผลรองรับอย่างชัดเจน
ระบบเชื่อมโยงข้อมูลจาก ฐานข้อมูลหรือแหล่งข้อมูลต่าง ๆ จนสามารถอนุมานข้อมูลใหม่ เกี่ยวกับลักษณะหรือพฤติกรรม ของบุคคลหรือไม่	ระบุว่าระบบมีการรวม เชื่อมโยง หรือเปรียบเทียบข้อมูลจากหลาย แหล่งหรือไม่ และการเชื่อมโยงดังกล่าวทำให้สามารถอนุมาน คาดการณ์ หรือสร้างข้อมูลใหม่เกี่ยวกับบุคคลได้อย่างไร รวมถึง พิจารณาว่าอาจนำไปสู่การจัดทำโปรไฟล์ การวิเคราะห์พฤติกรรม หรือการสร้างระเบียบข้อมูลบุคคลที่มีรายละเอียดเพิ่มขึ้นหรือไม่
บุคคลต้องให้ความยินยอมโดยชัด แจ้งก่อนที่ระบบ AI จะประมวลผล ข้อมูลหรือไม่	ระบุว่าการประมวลผลข้อมูลต้องอาศัยความยินยอมหรือใช้ฐาน ทางกฎหมายอื่น หากใช้ความยินยอม ควรพิจารณาว่าความ ยินยอมนั้นได้รับก่อนเริ่มประมวลผล มีความเฉพาะเจาะจง ได้รับ โดยเสรี และบุคคลสามารถปฏิเสธหรือถอนความยินยอมได้โดยไม่ เสียสิทธิหรือไม่
มีการประเมินผลกระทบด้านความ เป็นส่วนตัวของระบบ AI หรือไม่	ระบุว่าการประเมินผลกระทบด้านความเป็นส่วนตัว เช่น การ ประเมินผลกระทบด้านการคุ้มครองข้อมูลส่วนบุคคล (Data Protection Impact Assessment: DPIA) ก่อนนำระบบไปใช้หรือ เมื่อมีการเปลี่ยนแปลงที่สำคัญหรือไม่ รวมถึงระบุความเสี่ยงต่อ สิทธิและเสรีภาพของบุคคลและมาตรการที่ใช้ลดความเสี่ยง
มีการนำหลักการคุ้มครองความ เป็นส่วนตัวตั้งแต่การออกแบบมา ใช้กับระบบหรือไม่	ระบุว่าการนำหลักการ Privacy by Design มาใช้ตลอดวงจร ชีวิตของระบบ ตั้งแต่การออกแบบ พัฒนา นำไปใช้งาน และ ดำเนินงานจริงหรือไม่ เช่น การลดการใช้ข้อมูล การจำกัด วัตถุประสงค์ การควบคุมสิทธิ์ มาตรการรักษาความปลอดภัย ความโปร่งใส และการกำหนดค่าพื้นฐานที่คุ้มครองความเป็น ส่วนตัว
ผู้ใช้งานสามารถเข้าถึง แก้ไข หรือลบข้อมูลของตน และเลือกให้ ระบบหยุดประมวลผลข้อมูลได้ หรือไม่	ระบุกลไกที่ช่วยให้บุคคลสามารถตรวจสอบข้อมูลที่ระบบจัดเก็บ แก้ไขข้อมูลที่ไม่ถูกต้องหรือไม่ครบถ้วน ขอให้อลบข้อมูล หรือเลือก หยุดและถอนความยินยอมในการประมวลผลได้ รวมถึงขั้นตอน และช่องทางที่ใช้ดำเนินการ
หากบุคคลภายนอกสามารถ เข้าถึงข้อมูล มีมาตรการป้องกัน การนำข้อมูลไปใช้ในทางที่ไม่ เหมาะสมหรือไม่	ระบุว่าบุคคลหรือหน่วยงานภายนอกใดสามารถเข้าถึงหรือ ประมวลผลข้อมูล และมีมาตรการใดป้องกันการใช้งานโดยมิชอบ การเข้าถึงโดยไม่ได้รับอนุญาต หรือการเปิดเผยข้อมูลโดยไม่ตั้งใจ เช่น ข้อตกลงตามสัญญา การเข้ารหัสหรือใช้นามแฝง การ ควบคุมสิทธิ์ การติดตามการใช้งาน และนโยบายกำหนดขอบเขต การใช้ข้อมูล

รายการคำถาม	แนวทางประกอบการประเมิน
ชุดข้อมูลที่ใช้ฝึกสอนระบบ AI มีความโปร่งใสและสามารถตรวจสอบย้อนกลับได้เพียงใด	ระบุแหล่งที่มา เนื้อหา และลักษณะสำคัญของชุดข้อมูล รวมถึงเอกสารหรือ Metadata ที่เกี่ยวข้อง ตลอดจนกระบวนการเก็บรวบรวม ทำความสะอาด คัดกรอง เปลี่ยนแปลง และนำข้อมูลไปใช้ฝึกสอน เพื่อให้สามารถตรวจสอบที่มาและกระบวนการจัดการข้อมูลย้อนหลังได้
องค์กรทำให้มั่นใจได้อย่างไรว่าชุดข้อมูลที่ใช้ได้มาอย่างมีจริยธรรม	ระบุกระบวนการตรวจสอบว่าข้อมูลไม่มีเนื้อหาที่ผิดกฎหมาย เป็นอันตราย หรือได้มาในลักษณะที่ไม่เหมาะสม รวมถึงการเคารพสิทธิของบุคคลและองค์กร การปฏิบัติตามกฎหมายลิขสิทธิ์และทรัพย์สินทางปัญญา การมีใบอนุญาตหรือการอนุญาตที่เหมาะสม และการตรวจสอบความรับผิดชอบของผู้จัดหาข้อมูล
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความเป็นส่วนตัวและการคุ้มครองข้อมูลโดยย่อ	สรุปผลกระทบเชิงบวกและเชิงลบที่พบจากการประเมินในข้อก่อนหน้า เพื่อใช้เป็นข้อมูลตั้งต้นสำหรับการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ในขั้นตอนถัดไป

หมายเหตุ: รายการคำถามและแนวทางประกอบการประเมินในตารางเรียงเรียงจากคำถามและคำอธิบายประกอบในแบบฟอร์ม UNESCO EIA ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับลักษณะและบริบทการใช้งานจริงของระบบ AI

แนวทางการพิจารณา

การประเมินในมิตินี้ควรเริ่มจากการจัดทำภาพรวมการใช้ข้อมูลของระบบให้ชัดเจนว่า ระบบใช้ข้อมูลอะไร ได้ข้อมูลมาจากไหน ใช้ในขั้นตอนใด ใช้เพื่อวัตถุประสงค์ใด และมีใครสามารถเข้าถึงข้อมูลได้บ้าง โดยควรครอบคลุมทั้งข้อมูลที่ใช้ฝึกสอนและทดสอบระบบ รวมถึงข้อมูลที่ระบบประมวลผลระหว่างการใช้งานจริง

นอกจากการตรวจสอบว่ามีมาตรการรักษาความปลอดภัยหรือไม่ องค์กรควรพิจารณาด้วยว่าการใช้ข้อมูลแต่ละประเภทมีความจำเป็นและเหมาะสมกับวัตถุประสงค์หรือไม่ มีการเชื่อมโยงข้อมูลจนสามารถอนุมานข้อมูลใหม่เกี่ยวกับบุคคลหรือไม่ และบุคคลได้รับข้อมูลหรือสามารถใช้สิทธิเกี่ยวกับข้อมูลของตนได้อย่างเหมาะสมเพียงใด

ในกรณีที่มีบุคคลภายนอก ผู้ให้บริการ หรือแหล่งข้อมูลภายนอกเข้ามาเกี่ยวข้อง ควรพิจารณาทั้งขอบเขตการเข้าถึงข้อมูล มาตรการควบคุมการใช้งาน ที่มาของชุดข้อมูล และความสามารถในการตรวจสอบย้อนกลับ หากข้อมูล กระบวนการ หรือมาตรการบางส่วนยังไม่ชัดเจน ควรระบุเป็นช่องว่าง และกำหนดผู้รับผิดชอบในการตรวจสอบหรือปรับปรุงเพิ่มเติม

ผลจากการประเมินในมิตินี้จะช่วยให้องค์กรเห็นภาพรวมการใช้ข้อมูลส่วนบุคคลของระบบ AI ตั้งแต่ประเภทและที่มาของข้อมูล วัตถุประสงค์ในการใช้ มาตรการคุ้มครองข้อมูล สิทธิของบุคคล ตลอดจนการเข้าถึงข้อมูลโดยบุคคลภายนอก พร้อมทั้งระบุจุดแข็ง ช่องว่าง และประเด็นที่ควรปรับปรุงด้านความเป็นส่วนตัวและการคุ้มครองข้อมูล

ผลกระทบเชิงบวกและเชิงลบที่พบจากการประเมินจะถูกนำไปใช้เป็นข้อมูลตั้งต้นในการวิเคราะห์ผู้
อาจได้รับผลกระทบ ลักษณะของผลกระทบ ระดับความรุนแรง และโอกาสเกิดในขั้นตอนการวิเคราะห์
ผลกระทบทางจริยธรรม (Impact Mapping) ต่อไป

บทที่ 7

การกำกับดูแลและ การตัดสินใจโดยมนุษย์



มิติด้านการกำกับดูแลและการตัดสินใจโดยมนุษย์ (Human Oversight and Determination) มีวัตถุประสงค์เพื่อประเมินว่าระบบ AI เข้ามามีบทบาทในกระบวนการทำงานและการตัดสินใจอย่างไร ระบบทำหน้าที่เพียงให้ข้อมูลหรือคำแนะนำ หรือสามารถตัดสินใจและดำเนินการได้เอง รวมถึงมนุษย์ยังสามารถตรวจสอบ แทรกแซง หยุด หรือเปลี่ยนแปลงผลลัพธ์ของระบบได้อย่างเหมาะสมหรือไม่

การมีมนุษย์อยู่ในกระบวนการเพียงอย่างเดียวอาจยังไม่เพียงพอ ผู้ที่ทำหน้าที่กำกับดูแลควรมีข้อมูล ความรู้ อำนาจ และเวลาเพียงพอในการพิจารณาผลลัพธ์ของระบบ ตลอดจนทราบว่ากรณีใดต้องส่งต่อให้ผู้มีอำนาจทบทวน ใครสามารถเปลี่ยนแปลงการตัดสินใจ และต้องดำเนินการอย่างไรเมื่อพบข้อผิดพลาดหรือความเสี่ยง

ตัวอย่างเช่น หากระบบ AI ให้คำแนะนำด้านการรักษาพยาบาลหรือการอนุมัติสินเชื่อ ผู้รับผิดชอบไม่ควรยอมรับผลลัพธ์ของระบบโดยอัตโนมัติ แต่ควรพิจารณาข้อมูลอื่นร่วมด้วย และสามารถโต้แย้งหรือเปลี่ยนแปลงผลลัพธ์ได้เมื่อพบว่าคำแนะนำของระบบไม่สอดคล้องกับข้อเท็จจริงหรือบริบทของกรณีนั้น

จากประเด็นดังกล่าว การประเมินด้าน Human Oversight and Determination จึงควรพิจารณาทั้งบทบาทของระบบ ระดับที่ระบบสามารถตัดสินใจหรือดำเนินการได้เอง กลไกการแทรกแซง และส่งต่อให้มนุษย์ทบทวน อำนาจและความพร้อมของผู้กำกับดูแล ตลอดจนมาตรการป้องกันการพึ่งพาผลลัพธ์ของ AI มากเกินไป โดยมีรายการคำถามประเมินดังตารางต่อไปนี้

ตารางที่ 7-1 รายการคำถามประเมินด้านการกำกับดูแลและการตัดสินใจโดยมนุษย์

(Human Oversight and Determination)

รายการคำถาม	แนวทางประกอบการประเมิน
ระบบ AI มีบทบาทหลักอย่างไรในองค์กร	ระบุบทบาทหลักของระบบตามการใช้งานจริง ได้แก่ การแทนที่ระบบดิจิทัลเดิม การทำงานแทนงานที่มนุษย์เคยดำเนินการ การเพิ่มความสามารถหรือบริการใหม่ หรือการจัดทำข้อมูล การวิเคราะห์ หรือข้อมูลเชิงลึกเพื่อสนับสนุนการตัดสินใจของมนุษย์
ระบบ AI สามารถตัดสินใจหรือดำเนินการได้เองมากน้อยเพียงใด	ระบุว่าระบบสามารถตัดสินใจหรือดำเนินการได้เองโดยไม่ต้องได้รับความเห็นชอบจากมนุษย์ หรือทำหน้าที่เสนอคำแนะนำและข้อมูล โดยมนุษย์ยังเป็นผู้ตัดสินใจขั้นสุดท้าย หากระบบมีรูปแบบการทำงานอื่น ควรอธิบายขอบเขตการทำงานและจุดที่มนุษย์ต้องเข้ามากำกับดูแล
มีกลไกใดที่ทำให้มนุษย์สามารถแทรกแซง หยุด หรือเปลี่ยนแปลงการตัดสินใจของระบบ AI ได้	ระบุกลไกที่ใช้ทบทวน แก้ไข ยกเลิก หรือหยุดการตัดสินใจของระบบ เช่น การตรวจสอบโดยมนุษย์ กระบวนการอุทธรณ์ ปุ่มพักหรือหยุดระบบ หรือการแจ้งเตือนให้มนุษย์เข้ามาพิจารณา รวมถึงวิธีติดตามว่ากลไกดังกล่าวยังทำงานได้อย่างถูกต้องและถูกนำมาใช้อย่างสม่ำเสมอ
มีกลไกใดในการส่งต่อกรณีให้มนุษย์ทบทวน โดยเฉพาะกรณี	ระบุวิธีการค้นหา แจ้งเตือน หรือจัดลำดับกรณีที่ต้องส่งต่อให้มนุษย์พิจารณา เช่น การกำหนดเกณฑ์ระดับความเสี่ยง การแจ้งเตือนเมื่อผลลัพธ์มีความไม่แน่นอน หรือการกำหนดให้กรณีที่มีผลกระทบสูงต้อง

รายการคำถาม	แนวทางประกอบการประเมิน
ที่มีความเสี่ยงสูงหรือมีความอ่อนไหว	ได้รับอนุมัติจากมนุษย์ รวมถึงกระบวนการทบทวน แก้ไข และติดตามผล
ใครมีอำนาจในการแทรกแซงหรือเปลี่ยนแปลงการตัดสินใจของระบบ AI	ระบุชื่อ บทบาท หรือตำแหน่งของบุคคลหรือทีมที่มีอำนาจ พร้อมอธิบายขอบเขตความรับผิดชอบและวิธีติดตามการใช้อำนาจดังกล่าว รวมถึงพิจารณากรณีที่อาจเกิดจากวิจาณญาณส่วนบุคคล แรงกดดันภายในองค์กร หรือมุมมองที่จำกัด และมาตรการที่ใช้ตรวจสอบหรือลดอคติ
มีกลไกใดในการยกเลิกสิทธิ์เข้าถึงระบบของบุคคล รวมถึงอำนาจในการแทรกแซงการตัดสินใจของระบบ AI	ระบุกระบวนการบริหารและยกเลิกสิทธิ์เข้าถึงระบบ ผู้รับผิดชอบ และระยะเวลาที่ใช้ เช่น เมื่อบุคคลเปลี่ยนบทบาท พ้นจากองค์กร ไม่ปฏิบัติตามขั้นตอน หรือใช้ระบบอย่างไม่เหมาะสม รวมถึงการทบทวนสิทธิ์เป็นระยะเพื่อให้เฉพาะผู้ที่ได้รับอนุญาตสามารถเปลี่ยนแปลงผลลัพธ์ของระบบได้
ทีมที่ทำหน้าที่กำกับดูแลมีความรู้และทักษะเพียงพอในการประเมิน แก้ไข หรือเปลี่ยนแปลงการตัดสินใจของระบบ AI หรือไม่	ระบุว่าทีมมีความรู้ การฝึกอบรม ประสบการณ์ และอำนาจที่จำเป็นในการทำความเข้าใจการทำงาน ข้อจำกัด และความเสี่ยงของระบบ สามารถตีความผลลัพธ์ ตรวจสอบข้อผิดพลาดหรืออคติ และแทรกแซงได้อย่างเหมาะสมหรือไม่ รวมถึงมีกระบวนการสนับสนุนเมื่อต้องพิจารณากรณีที่ซับซ้อนหรือมีความเสี่ยงสูงอย่างไร
มีมาตรการใดในการป้องกันการพึ่งพาคำแนะนำของระบบ AI มากเกินไป	ระบุมาตรการที่ช่วยให้ผู้ตัดสินใจใช้วิจาณญาณและพิจารณาข้อมูลจากแหล่งอื่นร่วมด้วย ไม่ใช่ผลลัพธ์ของ AI เป็นเหตุผลเพียงอย่างเดียว เช่น การกำหนดให้มนุษย์ทบทวนผลลัพธ์ การอบรมเกี่ยวกับข้อจำกัดของระบบ การแจ้งเตือนเมื่อระบบมีความมั่นใจต่ำ การตรวจสอบการตัดสินใจย้อนหลัง หรือการเปรียบเทียบคำแนะนำของ AI กับข้อมูลและความเห็นจากผู้เชี่ยวชาญ
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อการกำกับดูแลและการตัดสินใจโดยมนุษย์โดยย่อ	สรุปผลกระทบเชิงบวกและเชิงลบที่พบจากการประเมินการกำหนดบทบาทและผู้รับผิดชอบ การตรวจสอบและติดตามย้อนกลับ ความรับผิดชอบตามกฎหมาย การตรวจประเมินระบบ ตลอดจนกระบวนการจัดการข้อผิดพลาดและข้อร้องเรียน เพื่อใช้เป็นข้อมูลตั้งต้นสำหรับการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ในขั้นตอนถัดไป

หมายเหตุ: รายการคำถามและแนวทางประกอบการประเมินในตารางเรียงจากคำถามและคำอธิบายประกอบในแบบฟอร์ม UNESCO EIA ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับลักษณะและบริบทการใช้งานจริงของระบบ AI

แนวทางการพิจารณา

การประเมินในมิตินี้ควรเริ่มจากการทำความเข้าใจกระบวนการทำงานของระบบว่า AI เข้ามาเกี่ยวข้องในขั้นตอนใด สร้างผลลัพธ์อะไร ใครเป็นผู้ใช้ผลลัพธ์ และผลลัพธ์ดังกล่าวถูกนำไปใช้ประกอบการตัดสินใจหรือดำเนินการอย่างไร เพื่อให้เห็นชัดว่ามนุษย์ยังมีบทบาทและความรับผิดชอบอยู่ตรงจุดใด

ควรพิจารณาว่าการกำกับดูแลสามารถนำไปใช้ได้จริง ไม่ใช่มีเพียงชื่อผู้รับผิดชอบหรือขั้นตอนที่กำหนดไว้ในเอกสาร ผู้กำกับดูแลควรเข้าถึงข้อมูลที่จำเป็น มีเวลาและความเชี่ยวชาญเพียงพอ สามารถตั้งคำถามกับผลลัพธ์ของระบบ และมีอำนาจในการหยุด แก้ไข หรือส่งต่อกรณีให้ผู้มีอำนาจพิจารณาได้

สำหรับกรณีที่มีความเสี่ยงสูงหรือมีผลกระทบสำคัญ ควรกำหนดเงื่อนไขในการส่งต่อให้มนุษย์ทบทวนอย่างชัดเจน เช่น เมื่อผลลัพธ์มีความไม่แน่นอน ข้อมูลไม่ครบถ้วน หรือการตัดสินใจอาจส่งผลกระทบต่อสิทธิ โอกาส สุขภาพ ความปลอดภัย หรือการเข้าถึงบริการของบุคคล

นอกจากนี้ ควรติดตามประสิทธิภาพของการกำกับดูแลอย่างต่อเนื่อง เช่น ตรวจสอบบันทึกการแทรกแซง การเปลี่ยนแปลงผลลัพธ์ การส่งต่อกรณี และการใช้สิทธิ์เข้าถึงระบบ เพื่อดูว่ากลไกที่กำหนดไว้ถูกนำมาใช้จริง มีความสม่ำเสมอ และไม่สร้างอคติหรือความเสี่ยงใหม่จากการตัดสินใจของมนุษย์

สรุปผลการประเมินและการนำไปใช้ต่อ

ผลจากการประเมินในมิตินี้จะช่วยให้องค์กรเห็นว่าระบบ AI มีบทบาทต่อการทำงานและการตัดสินใจมากน้อยเพียงใด มนุษย์ยังสามารถตรวจสอบ แทรกแซง และรับผิดชอบต่อผลการตัดสินใจได้จริงหรือไม่ รวมถึงผู้กำกับดูแลมีอำนาจ ความรู้ และกลไกสนับสนุนที่เพียงพอหรือไม่

ผลกระทบเชิงบวกและเชิงลบที่พบจะถูกนำไปใช้เป็นข้อมูลตั้งต้นในการวิเคราะห์ผู้ที่อาจได้รับผลกระทบ ลักษณะของผลกระทบ ระดับความรุนแรง และโอกาสเกิดในขั้นตอนการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ต่อไป

บทที่ 8

ความรับผิดชอบ และ ภาระรับผิดชอบ



มิติด้านความรับผิดชอบและการรับผิดชอบ (Accountability and Responsibility) มีวัตถุประสงค์เพื่อประเมินว่าองค์กรได้กำหนดบทบาทหน้าที่ และผู้รับผิดชอบต่อระบบ AI ไว้อย่างชัดเจนหรือไม่ รวมถึงสามารถระบุได้ว่าใครต้องชี้แจง รับผิดชอบต่อผลที่เกิดขึ้น และดำเนินการแก้ไขเมื่อระบบเกิดข้อผิดพลาดหรือก่อให้เกิดผลกระทบ

การระบุชื่อผู้รับผิดชอบเพียงอย่างเดียวอาจยังไม่เพียงพอ องค์กรควรมีกกลไกที่ทำให้การดำเนินงานของระบบสามารถตรวจสอบและติดตามย้อนกลับได้ ตั้งแต่ที่มาของข้อมูล การพัฒนาและเปลี่ยนแปลงโมเดล การตัดสินใจที่ใช้ AI สนับสนุน จนถึงผลลัพธ์ที่เกิดขึ้น รวมถึงควรมีหลักฐานที่ช่วยให้ทีมภายใน ผู้ตรวจสอบ หรือหน่วยงานที่เกี่ยวข้องสามารถทบทวนและตรวจสอบการดำเนินงานได้

ตัวอย่างความเสี่ยงที่ควรคำนึงถึงในมิตินี้ ได้แก่ การไม่มีผู้รับผิดชอบที่ชัดเจนในช่วงของวงจรชีวิตระบบ การแบ่งความรับผิดชอบระหว่างองค์กรและผู้ให้บริการภายนอกที่คลุมเครือ การไม่มีบันทึกที่เพียงพอสำหรับตรวจสอบย้อนหลัง การตรวจสอบที่อาจมีผลประโยชน์ทับซ้อน ตลอดจนการไม่มีการบันทึกการรับเรื่องร้องเรียน แก้ไขข้อผิดพลาด หรือเยียวยาผู้ที่ได้รับผลกระทบอย่างเหมาะสม

จากประเด็นดังกล่าว การประเมินด้าน Accountability and Responsibility จึงควรพิจารณาทั้งการกำหนดบทบาทและความรับผิดชอบ การตรวจสอบและติดตามย้อนกลับ ความรับผิดชอบตามกฎหมาย การตรวจสอบประเมินระบบ ตลอดจนกระบวนการจัดการข้อผิดพลาดและข้อร้องเรียน โดยมีรายการคำถามประเมินดังตารางต่อไปนี้

ตารางที่ 8-1 รายการคำถามประเมินด้านความรับผิดชอบและการรับผิดชอบ

(Accountability and Responsibility)

รายการคำถาม	แนวทางประกอบการประเมิน
มีกลไกใดเพื่อให้สามารถระบุผู้รับผิดชอบ ตรวจสอบการดำเนินงาน และชี้แจงผลที่เกิดจากระบบ AI ได้	ระบุกลไกที่ทำให้สามารถระบุได้ว่าใครรับผิดชอบต่อการออกแบบ การนำไปใช้ การดำเนินงาน และผลลัพธ์ของระบบ รวมถึงวิธีบันทึก ติดตาม และตรวจสอบการดำเนินงาน เช่น สายการรายงาน การตรวจสอบภายในหรือภายนอก กระบวนการทบทวนหรืออุทธรณ์ และขั้นตอนติดตามประสิทธิภาพและการปฏิบัติตามนโยบายหรือข้อกำหนด นอกจากนี้ ควรอธิบายว่าสามารถตรวจสอบย้อนกลับจากข้อมูลนำเข้า การฝึกสอนโมเดล และกระบวนการตัดสินใจไปยังผลลัพธ์ของระบบได้อย่างไร
ระบบ AI เคยได้รับการตรวจสอบประเมิน หรือมีแผนดำเนินการตรวจสอบประเมินในอนาคตหรือไม่	ระบุว่าระบบเคยได้รับการตรวจสอบประเมินหรือมีแผนดำเนินการหรือไม่ ใครเป็นผู้กำกับดูแลหรือดำเนินการ เช่น ทีมภายใน ผู้ตรวจสอบภายนอก หรือบุคคลที่สามที่เป็นอิสระ รวมถึงวิธีตรวจสอบผลประโยชน์ทับซ้อน ขอบเขต ช่วงเวลา ความถี่ และประเภทของการตรวจสอบประเมิน เช่น ด้านความปลอดภัย คุณภาพ การปฏิบัติตามข้อกำหนด หรือจริยธรรม ตลอดจนระบุว่า มี Audit Trail ที่บันทึกการตัดสินใจและจุดตัดสินใจสำคัญของระบบหรือไม่
สามารถระบุบุคคลหรือนิติบุคคลที่ต้องรับผิดชอบ	ระบุบุคคลหรือนิติบุคคลที่มีความรับผิดชอบตามกฎหมายในแต่ละช่วง ตั้งแต่การออกแบบ พัฒนา ทดสอบ นำไปใช้ ดำเนินงาน บำรุงรักษา จนถึงยุติการใช้งาน พร้อมอธิบายบทบาทของทีมภายใน หน่วยงานภาครัฐ ผู้

รายการคำถาม	แนวทางประกอบการประเมิน
กฎหมายในทุกช่วงของวงจรชีวิตระบบ AI ได้หรือไม่	ให้บริการ และองค์กรภายนอกที่เกี่ยวข้อง หากเป็นระบบที่จัดทำจากภายนอก ควรระบุความรับผิดชอบของผู้ให้บริการและข้อตกลงในสัญญา เช่น การคุ้มครองข้อมูล ความปลอดภัย มาตรฐานด้านจริยธรรม และความถูกต้องของผลลัพธ์ รวมถึงวิธีป้องกันไม่ให้ความรับผิดชอบคลุมเครือ หรือไม่มีผู้รับผิดชอบ
มีกระบวนการใดเพื่อให้ข้อผิดพลาดหรืออันตรายที่เกิดจากระบบ AI ได้รับการแก้ไขอย่างรวดเร็ว	ระบุกระบวนการรับเรื่อง ตรวจสอบ และแก้ไขข้อผิดพลาด อันตราย หรือผลกระทบที่ไม่ตั้งใจ รวมถึงวิธีกำหนดผู้รับผิดชอบ ช่องทางการร้องเรียนหรืออุทธรณ์ ผู้พิจารณาเรื่องร้องเรียน ระยะเวลาดำเนินการ และวิธีแจ้งผลแก่ผู้ที่ได้รับผลกระทบ ควรระบุด้วยว่ามีการบันทึกเรื่องร้องเรียน การทบทวน และการแก้ไขเพื่อตรวจสอบย้อนหลังหรือไม่ รวมถึงมีการนำบทเรียนที่พบกลับไปปรับปรุงระบบและกระบวนการกำกับดูแลอย่างไร
มีกระบวนการตรวจสอบข้อกล่าวอ้างหรือข้อกังวลเกี่ยวกับระบบ AI ที่มาจากประชาชน นักวิจัย หรือสื่อมวลชนหรือไม่	ระบุว่ามีกระบวนการอย่างเป็นทางการสำหรับรับและตรวจสอบข้อกล่าวอ้าง ข้อกังวล หรือข้อร้องเรียนจากภายนอกหรือไม่ รวมถึงช่องทางการรับเรื่อง ผู้รับผิดชอบ ขั้นตอนการตรวจสอบ และวิธีสื่อสารผลการตรวจสอบ หากยังไม่มีกระบวนการที่เป็นระบบ ควรระบุสถานะปัจจุบัน และแนวทางดำเนินการต่อไป
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความรับผิดชอบและการรับผิดชอบโดยย่อ	สรุปผลกระทบเชิงบวกและเชิงลบตามประเด็นที่ระบุในแบบฟอร์ม เพื่อใช้เป็นข้อมูลตั้งต้นสำหรับการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ในขั้นตอนถัดไป

หมายเหตุ: รายการคำถามและแนวทางประกอบการประเมินในตารางเรียงเรียงจากคำถามและคำอธิบายประกอบในแบบฟอร์ม UNESCO EIA ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับลักษณะและบริบทการใช้งานจริงของระบบ AI

แนวทางการพิจารณา

การประเมินในมิตินี้ควรเริ่มจากการระบุบุคคลและหน่วยงานที่เกี่ยวข้องกับระบบ AI ตลอดวงจรชีวิต พร้อมกำหนดให้ชัดเจนว่าแต่ละฝ่ายมีหน้าที่ อำนาจในการตัดสินใจ และความรับผิดชอบต่อผลลัพธ์ในส่วนใด โดยเฉพาะกรณีที่มีผู้ให้บริการหรือองค์กรภายนอกเข้ามาเกี่ยวข้อง

ควรพิจารณาว่าการทำงานของระบบสามารถตรวจสอบย้อนหลังได้จริงหรือไม่ เช่น สามารถติดตามที่มาของข้อมูล รุ่นของโมเดล ขั้นตอนการประมวลผล การตัดสินใจที่ใช้ AI สนับสนุน และการแทรกแซงโดยมนุษย์ได้หรือไม่ รวมถึงมีข้อมูลเพียงพอให้ผู้รับผิดชอบ ผู้ตรวจสอบ หรือหน่วยงานที่เกี่ยวข้องทบทวนได้อย่างเหมาะสม

การตรวจประเมินควรมีขอบเขตและความถี่ที่ชัดเจน รวมถึงพิจารณาความเป็นอิสระและผลประโยชน์ทับซ้อนของผู้ตรวจสอบ ตลอดจนตรวจสอบว่าข้อค้นพบจากการประเมินถูกนำไปใช้แก้ไขหรือปรับปรุงระบบจริงหรือไม่

นอกจากนี้ องค์กรควรมีกระบวนการจัดการเมื่อเกิดข้อผิดพลาดหรือผลกระทบที่เข้าถึงได้และดำเนินการได้จริง โดยกำหนดช่องทางรับเรื่อง ผู้รับผิดชอบ ระยะเวลาดำเนินการ กระบวนการแก้ไขหรือเยียวยา และวิธีแจ้งผลแก่ผู้ที่ได้รับผลกระทบอย่างชัดเจน

สรุปผลการประเมินและการนำไปใช้ต่อ

ผลจากการประเมินในมิตินี้จะช่วยให้องค์กรเห็นว่า มีการกำหนดผู้รับผิดชอบต่อระบบ AI อย่างชัดเจนตลอดวงจรชีวิตหรือไม่ บุคคลหรือหน่วยงานที่เกี่ยวข้องสามารถชี้แจงและรับผิดชอบต่อผลที่เกิดขึ้นได้เพียงใด รวมถึงการทำงานและการตัดสินใจของระบบสามารถตรวจสอบย้อนหลังได้หรือไม่

ประเด็นและผลกระทบที่พบจากการประเมินจะถูกนำไปใช้เป็นข้อมูลตั้งต้นในการวิเคราะห์ผู้ที่อาจได้รับผลกระทบ ลักษณะของผลกระทบ ระดับความรุนแรง และโอกาสเกิดในขั้นตอนการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ต่อไป

บทที่ 9

ความโปร่งใส ความสามารถในการ อธิบายและความตระหนักรู้



การประเมินด้านความโปร่งใส ความสามารถในการอธิบาย และความตระหนักรู้ มีวัตถุประสงค์เพื่อพิจารณาว่าองค์กรเปิดเผยข้อมูลเกี่ยวกับการนำระบบ AI มาใช้อย่างเหมาะสมหรือไม่ ผู้ใช้งานและผู้ที่เกี่ยวข้องได้รับผลกระทบหรือไม่ว่า AI เข้ามามีบทบาทในกระบวนการใด และสามารถเข้าถึงคำอธิบายเกี่ยวกับการทำงานและผลลัพธ์ของระบบได้เพียงใด

การแจ้งเพียงว่ามีการใช้ AI อาจยังไม่เพียงพอ ข้อมูลที่สื่อสารควรชัดเจน ทันเวลา และเข้าใจได้ สำหรับผู้ที่ไม่มีความรู้ด้านเทคนิค โดยควรอธิบายให้เห็นว่า AI มีบทบาทอย่างไร ปัจจุบันมีส่วนต่อผลลัพธ์และระบบมีข้อจำกัดหรือความไม่แน่นอนในเรื่องใด นอกจากนี้ ผู้ใช้งานควรได้รับคำแนะนำหรือการฝึกอบรมที่ช่วยให้สามารถใช้ระบบและพิจารณาผลลัพธ์ได้อย่างเหมาะสม

ตัวอย่างความเสี่ยงที่ควรคำนึงถึง ได้แก่ ผู้ใช้งานไม่ทราบว่าระบบที่กำลังใช้งานมี AI เป็นส่วนหนึ่ง บุคคลไม่ทราบว่า AI ถูกนำมาใช้ประกอบการตัดสินใจที่ส่งผลกระทบต่อตนเอง ไม่สามารถขอคำอธิบายเกี่ยวกับการทำงานหรือผลลัพธ์ของระบบได้ คำอธิบายมีความซับซ้อนหรือใช้ศัพท์ทางเทคนิคมากเกินไป รวมถึงผู้ใช้งานไม่ได้รับข้อมูลเกี่ยวกับข้อจำกัดและแนวทางการใช้ระบบอย่างเหมาะสม

จากประเด็นดังกล่าว การประเมินด้าน Transparency, Explainability, and Awareness จึงควรพิจารณาทั้งการเปิดเผยการนำ AI มาใช้ การแจ้งให้ผู้ใช้งานและผู้ที่เกี่ยวข้องได้รับผลกระทบ การเข้าถึงคำอธิบาย ความเข้าใจได้ของผลลัพธ์ ตลอดจนคำแนะนำและการฝึกอบรมที่เกี่ยวข้อง โดยมีรายการคำถามประเมินดังตารางต่อไปนี้

ตารางที่ 9-1 รายการคำถามประเมินด้านความโปร่งใส ความสามารถในการอธิบาย และความตระหนักรู้ (Transparency, Explainability, and Awareness)

รายการคำถาม	แนวทางประกอบการประเมิน
การตัดสินใจนำระบบ AI มาใช้ ได้รับการบันทึกอย่างเป็นทางการและสื่อสารต่อสาธารณะทางออนไลน์หรือไม่	ระบุว่ามีการจัดทำเอกสารเกี่ยวกับการตัดสินใจนำระบบ AI มาใช้หรือไม่ เอกสารดังกล่าวครอบคลุมข้อมูลเรื่องใด และมีการเผยแพร่ผ่านช่องทางออนไลน์ใด รวมถึงประชาชน ผู้ใช้งาน หรือผู้มีส่วนได้ส่วนเสียสามารถเข้าถึงข้อมูลดังกล่าวได้หรือไม่
ผู้ใช้งานได้รับแจ้งหรือไม่ว่าระบบที่กำลังใช้งานมี AI เป็นส่วนหนึ่ง หรือ AI มีส่วนต่อผลลัพธ์ที่ส่งผลกระทบต่อตนเอง	ระบุวิธีการแจ้งให้บุคคลทราบถึงบทบาทของ AI ทั้งกรณีที่ใช้งานระบบโดยตรงและกรณีที่ AI ถูกนำมาใช้ประกอบการตัดสินใจโดยบุคคลไม่ได้ใช้งานระบบด้วยตนเอง เช่น ข้อความแจ้งเตือน ป้ายกำกับ สัญลักษณ์ หรือข้อความบนหน้าจอ รวมถึงรูปแบบ ช่วงเวลา และความสามารถในการเข้าถึงข้อมูลดังกล่าว
ประชาชนหรือผู้ใช้งานสามารถขอและเข้าถึงคำอธิบายเกี่ยวกับการทำงานและกระบวนการตัดสินใจของระบบ AI ได้หรือไม่	ระบุช่องทางและกระบวนการที่ประชาชนหรือผู้ใช้งานสามารถขอคำอธิบายเกี่ยวกับความสามารถ วัตถุประสงค์ การใช้งาน และผลกระทบของระบบ เช่น เว็บไซต์ด้านความโปร่งใส ทะเบียนระบบ AI ช่องทางสอบถาม เอกสารเผยแพร่ หรือช่องทางสนับสนุนผู้ใช้งาน รวมถึงประเภทของข้อมูลที่เปิดเผยและช่วงเวลาที่สามารถขอข้อมูลได้
ผลลัพธ์ของระบบ AI สามารถอธิบายด้วยภาษาที่ผู้ไม่มี	ระบุวิธีการอธิบายผลลัพธ์ของระบบโดยใช้ภาษาที่ชัดเจนและหลีกเลี่ยงศัพท์ทางเทคนิคที่ไม่จำเป็น โดยควรอธิบายปัจจัยสำคัญที่มีส่วนต่อผลลัพธ์ บทบาทของระบบ AI ตลอดจนข้อจำกัดหรือความไม่แน่นอนที่เกี่ยวข้อง เช่น คำอธิบายแบบข้อความ ภาพประกอบ ข้อความแจ้งผู้ใช้งาน

รายการคำถาม	แนวทางประกอบการประเมิน
ความรู้ด้านเทคนิคเข้าใจได้หรือไม่	Dashboard หรือหนังสือแจ้งผลการตัดสินใจ รวมถึงการปรับรูปแบบคำอธิบายให้เหมาะสมกับผู้รับข้อมูลแต่ละกลุ่ม
ผู้ใช้งานได้รับคำแนะนำหรือการฝึกอบรมเพื่อให้เข้าใจและใช้ระบบ AI อย่างเหมาะสมหรือไม่	ระบุคำแนะนำ การฝึกอบรม หรือสื่อการเรียนรู้ที่จัดให้แก่ผู้ใช้งาน ผู้ปฏิบัติงาน หรือผู้ตัดสินใจ เช่น คู่มือการใช้งาน การฝึกอบรม หลักสูตรออนไลน์ นโยบายภายใน เอกสารอธิบายสำหรับประชาชน กิจกรรมสร้างความตระหนักรู้ หรือกิจกรรมร่วมกับชุมชน หากยังไม่มี การดำเนินการ ให้ระบุเหตุผลและแผนการพัฒนาในอนาคต
สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความโปร่งใส ความสามารถในการอธิบาย และความตระหนักรู้	สรุปผลกระทบเชิงบวกและเชิงลบตามประเด็นที่ระบุในแบบฟอร์ม เพื่อใช้เป็นข้อมูลตั้งต้นสำหรับการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ในขั้นตอนถัดไป

หมายเหตุ: รายการคำถามและแนวทางประกอบการประเมินในตารางเรียงเรียงจากคำถามและคำอธิบายประกอบในแบบฟอร์ม UNESCO EIA ผู้ประเมินควรเลือกใช้และปรับให้เหมาะสมกับลักษณะและบริบทการใช้งานจริงของระบบ AI

แนวทางการพิจารณา

การประเมินในส่วนนี้ควรเริ่มจากการระบุว่าใครจำเป็นต้องรับทราบข้อมูลเกี่ยวกับระบบ AI ต้องได้รับข้อมูลเรื่องใด ในช่วงเวลาใด และผ่านช่องทางใด โดยควรครอบคลุมทั้งผู้ใช้งานโดยตรง ผู้ที่ได้รับผลกระทบจากผลลัพธ์ของระบบ และประชาชนหรือผู้มีส่วนได้ส่วนเสียที่ต้องการตรวจสอบข้อมูลเกี่ยวกับระบบ

ควรพิจารณาคุณภาพของข้อมูลที่สื่อสาร ไม่ใช่เพียงตรวจสอบว่ามีเอกสารหรือข้อความแจ้งเตือนหรือไม่ ผู้รับข้อมูลควรเข้าใจได้ว่า AI เข้ามามีบทบาทอย่างไร ผลลัพธ์ถูกนำไปใช้อย่างไร และมีข้อจำกัดหรือความไม่แน่นอนใดที่ควรทราบ โดยรูปแบบและระดับรายละเอียดของคำอธิบายควรเหมาะสมกับผู้รับข้อมูลแต่ละกลุ่ม

นอกจากนี้ ควรตรวจสอบว่าผู้ใช้งานสามารถขอข้อมูลหรือคำอธิบายเพิ่มเติมได้จริง มีผู้รับผิดชอบในการตอบคำถาม และมีคำแนะนำหรือการฝึกอบรมที่ช่วยให้ผู้ใช้งานเข้าใจบทบาทและข้อจำกัดของระบบ หากข้อมูล ช่องทาง หรือสื่อสนับสนุนบางส่วนยังไม่พร้อม ควรระบุเป็นช่องว่างและกำหนดแนวทางในการพัฒนาหรือปรับปรุงเพิ่มเติม

สรุปผลการประเมินและการนำไปใช้ต่อ

ผลจากการประเมินในส่วนนี้จะช่วยให้องค์กรเห็นว่า การนำระบบ AI มาใช้ได้รับการเปิดเผยอย่างเหมาะสมหรือไม่ ผู้ใช้งานและผู้ที่เกี่ยวข้องได้รับผลกระทบรับรู้และเข้าใจบทบาทของ AI ได้เพียงใด รวมถึงสามารถเข้าถึงคำอธิบาย คำแนะนำ และข้อมูลที่จำเป็นต่อการใช้งานหรือการตรวจสอบระบบได้หรือไม่

ประเด็นและผลกระทบที่พบจากการประเมินจะถูกนำไปใช้เป็นข้อมูลตั้งต้นในการวิเคราะห์ผู้ที่อาจได้รับผลกระทบ ลักษณะของผลกระทบ ระดับความรุนแรง และโอกาสเกิดในขั้นตอนการวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) ต่อไป

บทที่ 10

การวิเคราะห์ผลกระทบทางจริยธรรม



ภายหลังจากประเมินมาตรการและกลไกกำกับดูแลของระบบ AI ครบทุกมิติแล้ว องค์กรควรนำข้อมูลที่ได้มาพิจารณาต่อว่า ระบบอาจก่อให้เกิดผลกระทบเชิงบวกหรือเชิงลบอย่างไร เกิดขึ้นกับใคร และมีความสำคัญเพียงใด เพื่อให้สามารถกำหนดมาตรการกำกับดูแลและตัดสินใจเกี่ยวกับการพัฒนาและใช้งานระบบได้อย่างเหมาะสม

การวิเคราะห์ผลกระทบทางจริยธรรม (Impact Mapping) เป็นแนวทางที่ช่วยให้ทีมประเมินเปรียบเทียบผลกระทบที่สำคัญของระบบ AI ในแต่ละมิติ โดยไม่ได้มีวัตถุประสงค์เพื่อคำนวณคะแนนความเสี่ยงแบบตายตัว หรือใช้ทดแทนการประเมินด้านกฎหมาย ด้านเทคนิค และด้านการดำเนินงาน แต่ช่วยให้ทีมพิจารณาผลกระทบอย่างเป็นระบบ พร้อมบันทึกเหตุผลและข้อมูลที่ใช้ประกอบการตัดสินใจ

10.1 ควรดำเนินการ Impact Mapping เมื่อใด

Impact Mapping ควรดำเนินการภายหลังจากการกำหนดขอบเขตการประเมิน (Scoping Questions) และการประเมินตามหลักการด้านจริยธรรม (Principles-Based Questions) เสร็จสิ้นแล้ว เนื่องจากข้อมูลจากสองขั้นตอนดังกล่าวจะช่วยให้ทีมเข้าใจบริบทของระบบ มาตรการที่มีอยู่ และผลกระทบเบื้องต้นในแต่ละมิติได้ชัดเจนยิ่งขึ้น

การวิเคราะห์นี้ไม่ควรดำเนินการเพียงครั้งเดียว แต่ควรทบทวนเมื่อระบบมีการเปลี่ยนแปลงที่สำคัญ ถูกนำไปใช้ในบริบทใหม่ ขยายการใช้งานไปยังผู้ใช้งานหรือประชากรกลุ่มใหม่ หรือเชื่อมต่อกับระบบอื่น เพื่อให้ผลกระทบที่เกิดขึ้นใหม่และการเปลี่ยนแปลงของบริบทได้รับการพิจารณาอย่างต่อเนื่อง

10.2 ผู้ที่ควรมีส่วนร่วมในการวิเคราะห์

การจัดทำ Impact Mapping ควรเป็นความรับผิดชอบร่วมกันของสมาชิกทีมและผู้ที่เกี่ยวข้องกับระบบ AI โดยอาศัยความรู้และมุมมองจากหลายด้าน เช่น ด้านเทคนิค การดำเนินงาน กฎหมาย จริยธรรม การกำกับดูแล และบริบทของกลุ่มผู้ใช้งาน

การมีส่วนร่วมจากบุคลากรที่มีบทบาทและความเชี่ยวชาญแตกต่างกันจะช่วยลดโอกาสที่ผลกระทบบางประเด็นจะถูกมองข้าม และทำให้การประเมินสะท้อนทั้งมิติทางเทคนิค การดำเนินงาน จริยธรรม และสังคมได้อย่างรอบด้านมากขึ้น

10.3 การระบุผลกระทบเชิงบวกและเชิงลบ

ทีมควรนำผลกระทบเชิงบวกและเชิงลบที่สรุปไว้ในข้อสุดท้ายของการประเมินแต่ละมิติมาทบทวนให้ครบถ้วนได้แก่

1. ความปลอดภัยและความมั่นคงปลอดภัย
2. ความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย
3. ความยั่งยืน
4. ความเป็นส่วนตัวและการคุ้มครองข้อมูล
5. การกำกับดูแลและการตัดสินใจโดยมนุษย์
6. ความรับผิดชอบและการรับผิดชอบ
7. ความโปร่งใส ความสามารถในการอธิบาย และความตระหนักรู้

ผลกระทบควรเขียนเป็นสิ่งที่อาจเกิดขึ้นอย่างเป็นรูปธรรม โดยอธิบายให้เห็นว่าเกิดขึ้นอย่างไรและเกิดกับใคร หลีกเลี่ยงข้อความที่กว้างเกินไป เช่น “เพิ่มประสิทธิภาพ” โดยไม่ระบุว่าเพิ่มประสิทธิภาพในเรื่องใด หรือส่งผลกระทบต่อใคร เพื่อให้การวิเคราะห์มีความชัดเจนและไม่ซับซ้อนเกินไปควรเลือกเฉพาะผลกระทบที่สำคัญ

ตัวอย่างการเขียนผลกระทบ

ข้อความที่ยังกว้าง: ระบบช่วยเพิ่มประสิทธิภาพการให้บริการ

ข้อความที่เห็นผลกระทบชัดเจน: ระบบช่วยลดระยะเวลาในการคัดแยกคำร้อง ทำให้เจ้าหน้าที่สามารถพิจารณาคำร้องที่ซับซ้อนได้เร็วขึ้น และประชาชนได้รับการส่งต่อเรื่องภายในระยะเวลาที่สั้นลง

10.3.1 ตัวอย่างผลกระทบเชิงบวกและเชิงลบในแต่ละมิติ

ตัวอย่างต่อไปนี้มีวัตถุประสงค์เพื่อช่วยให้ทีมเห็นลักษณะของผลกระทบที่อาจนำมาพิจารณา โดยไม่ใช่รายการคำตอบสำเร็จรูป ทีมควรระบุผลกระทบให้สอดคล้องกับวัตถุประสงค์ ลักษณะการทำงาน ผู้ใช้งาน และบริบทของระบบ AI ที่ประเมิน

มิติ	ตัวอย่างผลกระทบเชิงบวก	ตัวอย่างผลกระทบเชิงลบ
ความปลอดภัยและความมั่นคงปลอดภัย	ระบบช่วยตรวจพบความผิดปกติหรือเหตุการณ์ที่อาจก่อให้เกิดอันตรายได้เร็วขึ้น	ระบบให้ผลลัพธ์ที่คลาดเคลื่อนหรือถูกแทรกแซง จนนำไปสู่การตัดสินใจที่ไม่ปลอดภัย
ความเป็นธรรม การไม่เลือกปฏิบัติ และความหลากหลาย	ระบบช่วยให้การให้บริการมีความสม่ำเสมอและเพิ่มโอกาสในการเข้าถึงบริการของกลุ่มที่เคยมีข้อจำกัด	ระบบให้ผลลัพธ์ที่ไม่เป็นธรรมต่อบางกลุ่ม ตอกย้ำภาพจำ หรือสร้างอุปสรรคต่อผู้พิการและผู้ที่มีความพิการด้านภาษาและทักษะดิจิทัล
ความยั่งยืน	ระบบช่วยลดการใช้ทรัพยากร หรือสนับสนุนการจัดการพลังงานและสิ่งแวดล้อมได้ดีขึ้น	การพัฒนาและใช้งานระบบใช้พลังงานและทรัพยากรสูง หรือก่อให้เกิดผลกระทบต่อสิ่งแวดล้อม
ความเป็นส่วนตัวและการคุ้มครองข้อมูล	ระบบใช้ข้อมูลที่ไม่สามารถระบุตัวบุคคลได้อย่างเหมาะสม ทำให้สามารถสร้างประโยชน์จากข้อมูลโดยลดความเสี่ยงต่อบุคคล	ข้อมูลส่วนบุคคลถูกใช้เกินวัตถุประสงค์ รั่วไหล หรือได้รับการคุ้มครองไม่เพียงพอ
การกำกับดูแลและการตัดสินใจโดยมนุษย์	มนุษย์สามารถตรวจสอบ แทรกแซง และแก้ไขผลลัพธ์ของระบบได้ก่อนเกิดความเสียหาย	ผู้ตัดสินใจพึ่งพาคำแนะนำของ AI มากเกินไป หรือไม่สามารถแทรกแซงระบบได้เมื่อเกิดกรณีที่มีความเสี่ยง

มิติ	ตัวอย่างผลกระทบเชิงบวก	ตัวอย่างผลกระทบเชิงลบ
ความรับผิดชอบและ การรับผิดชอบ	สามารถระบุผู้รับผิดชอบ ตรวจสอบ ย้อนกลับ และแก้ไขข้อผิดพลาดได้ อย่างชัดเจน	ไม่สามารถระบุผู้รับผิดชอบหรือสาเหตุ ของข้อผิดพลาดได้ ทำให้การ ร้องเรียน การแก้ไข หรือการเยียวยา ทำได้ยาก
ความโปร่งใส ความสามารถในการ อธิบาย และความ ตระหนักรู้	ผู้ใช้งานทราบว่า AI มีบทบาท อย่างไร และได้รับคำอธิบายที่ช่วยให้ เข้าใจผลลัพธ์และตัดสินใจได้อย่าง เหมาะสม	ผู้ใช้งานไม่ทราบว่า AI มีส่วนต่อ ผลลัพธ์ที่ส่งผลต่อตนเอง หรือได้รับ คำอธิบายที่ซับซ้อน ไม่ครบถ้วน และไม่สามารถนำไปใช้ได้จริง

10.4 การวิเคราะห์ผลกระทบเชิงบวก

นำผลกระทบเชิงบวกที่สรุปจากการประเมินใน Part 2 มาวิเคราะห์รายละเอียดตามองค์ประกอบต่อไปนี้

องค์ประกอบ	แนวทางการพิจารณา
ผลกระทบเชิงบวกที่อาจเกิดขึ้น	อธิบายประโยชน์ที่อาจเกิดขึ้นให้ชัดเจนว่าเกิดกับใครและเกิดขึ้น อย่างไร
ระดับของผลกระทบ (Scale)	พิจารณาขนาดและความสำคัญของประโยชน์หากผลกระทบนั้น เกิดขึ้น
ผู้ที่ได้รับผลกระทบ (Impacted Parties)	ระบุผู้ได้รับผลกระทบโดยตรง โดยอ้อม หรือกลุ่มที่ไม่ได้ คาดการณ์ไว้
ระยะเวลา (Timescale)	พิจารณาว่าผลกระทบอาจเกิดในระยะสั้น ระยะกลาง ระยะยาว หรือข้ามรุ่น
โอกาสเกิด (Likelihood)	ประเมินว่าผลกระทบมีแนวโน้มเกิดขึ้นมากน้อยเพียงใด
หมายเหตุของทีม (Team Notes)	บันทึกเหตุผล หลักฐาน สมมติฐาน และข้อมูลที่ใช้ประกอบการ ประเมิน

10.5 การวิเคราะห์ผลกระทบเชิงลบ

นำผลกระทบเชิงลบที่สรุปจากการประเมินใน Part 2 มาวิเคราะห์ตามองค์ประกอบเดียวกับผลกระทบเชิงบวก และพิจารณาเพิ่มเติมถึงความสามารถในการเยียวยา มาตรการที่มีอยู่ และมาตรการเพิ่มเติมที่จำเป็น

องค์ประกอบ	แนวทางการพิจารณา
ผลกระทบเชิงลบที่อาจเกิดขึ้น	อธิบายอันตราย ความเสียหาย หรือผลที่ไม่พึงประสงค์ให้ชัดเจน
ระดับของผลกระทบ (Scale)	พิจารณาความรุนแรงของผลกระทบหากเกิดขึ้น
ผู้ที่ได้รับผลกระทบ (Impacted Parties)	ระบุผู้ได้รับผลกระทบโดยตรง โดยอ้อม หรือกลุ่มที่ไม่ได้คาดการณ์ไว้
ระยะเวลา (Timescale)	ระบุว่าผลกระทบอาจเกิดขึ้นหรือคงอยู่ในช่วงเวลาใด
ความสามารถในการแก้ไขหรือเยียวยา (Remediability)	พิจารณาว่าผลกระทบสามารถย้อนกลับ แก้ไข หรือลดความเสียหายได้เพียงใด
โอกาสเกิด (Likelihood)	ประเมินว่าผลกระทบมีแนวโน้มเกิดขึ้นมากน้อยเพียงใด
มาตรการที่มีอยู่	พิจารณาว่ามาตรการและกลไกจากการประเมินใน Part 2 สามารถลดผลกระทบได้เพียงใด
มาตรการเพิ่มเติม	ระบุมาตรการที่ต้องเพิ่มเพื่อป้องกัน ลด แก้ไข หรือเยียวยาผลกระทบ
หมายเหตุของทีม (Team Notes)	บันทึกเหตุผล หลักฐาน สมมติฐาน และข้อจำกัดของข้อมูล

10.6 เกณฑ์ประกอบการวิเคราะห์ผลกระทบ

การวิเคราะห์ผลกระทบเชิงบวกและเชิงลบในหัวข้อก่อนหน้า ให้พิจารณาตามเกณฑ์ต่อไปนี้ โดยเลือกคำตอบที่สะท้อนบริบทของระบบมากที่สุด เกณฑ์เหล่านี้ไม่ได้นำไปคำนวณเป็นคะแนนรวม แต่ใช้เพื่อเปรียบเทียบผลกระทบและบันทึกเหตุผลประกอบการตัดสินใจอย่างเป็นระบบ

10.6.1 ระดับของผลกระทบ (Scale)

ระดับของผลกระทบใช้พิจารณาว่า หากผลกระทบเกิดขึ้นแล้ว จะก่อให้เกิดประโยชน์หรือความเสียหายมากเพียงใด

ผลกระทบเชิงบวก

ระดับ	แนวทางการพิจารณา
สูงมาก (Very High)	ก่อให้เกิดการเปลี่ยนแปลงเชิงบวกในระดับมากหรือระดับระบบ และประโยชน์มีแนวโน้มคงอยู่ในระยะยาวหรือถาวร

ระดับ	แนวทางการพิจารณา
สูง (High)	ก่อให้เกิดประโยชน์อย่างมีนัยสำคัญและต่อเนื่องแก่ผู้ที่ได้รับผลกระทบ
ปานกลาง (Medium)	ก่อให้เกิดประโยชน์ที่เห็นได้ชัด แต่อาจจำกัดอยู่ในบางช่วงเวลา บางกระบวนการ หรือบางกลุ่ม
ปานกลางหรือเล็กน้อย (Moderate/Minor)	ก่อให้เกิดประโยชน์ในระดับจำกัด และไม่ได้เปลี่ยนแปลงผลลัพธ์หลักอย่างมีนัยสำคัญ

ผลกระทบเชิงลบ

ระดับ	แนวทางการพิจารณา
ร้ายแรงอย่างยิ่ง (Catastrophic)	ก่อให้เกิดความเสียหายอย่างร้ายแรงที่สุด เช่น การสูญเสียชีวิต เสรีภาพ หรือความเสียหายที่ไม่สามารถย้อนกลับได้
วิกฤต (Critical)	ก่อให้เกิดความเสียหายอย่างมีนัยสำคัญและยาวนานต่อบุคคล ชุมชน สังคม หรือระบบกฎหมาย
รุนแรง (Serious)	ก่อให้เกิดความเสียหายอย่างมีนัยสำคัญในช่วงระยะเวลาหนึ่ง แม้ผลกระทบอาจไม่ถาวร
ปานกลางหรือเล็กน้อย (Moderate/Minor)	ก่อให้เกิดความเสียหายในวงจำกัด และไม่ทำให้เกิดผลกระทบที่รุนแรงหรือยาวนาน

ในการเลือกระดับ ทิมควรถามว่า หากผลกระทบเกิดขึ้นจริง จะเปลี่ยนแปลงคุณภาพชีวิต สิทธิ โอกาส สุขภาพ ความปลอดภัย สังคม สิ่งแวดล้อม หรือโครงสร้างพื้นฐานมากเพียงใด และผลกระทบจะคงอยู่นานหรือไม่

10.6.2 ผู้ที่ได้รับผลกระทบ (Impacted Parties)

เกณฑ์นี้ใช้ระบุใครหรือสิ่งใดอาจได้รับผลจากระบบ โดยไม่ควรมองเฉพาะผู้ใช้งานโดยตรงเท่านั้น

กลุ่ม	แนวทางการพิจารณา
ผู้ได้รับผลกระทบโดยตรง (Primary)	บุคคลหรือกลุ่มที่เกี่ยวข้องโดยตรงกับการพัฒนา จัดหา ใช้งาน หรือเป็นเป้าหมายของระบบ
ผู้ได้รับผลกระทบโดยอ้อม (Secondary)	บุคคลหรือกลุ่มที่ได้รับผลต่อเนื่องจากผลกระทบที่เกิดกับผู้ได้รับผลกระทบโดยตรง เช่น ครอบครัว ผู้พึ่งพา หรือชุมชนที่เกี่ยวข้อง

กลุ่ม	แนวทางการพิจารณา
ผู้ได้รับผลกระทบที่ไม่ได้คาดการณ์ไว้ (Unexpected/Unintended)	บุคคล กลุ่ม สิ่งแวดล้อม หรือสิ่งอื่นที่อาจได้รับผลกระทบ แม้ไม่ได้อยู่ในขอบเขตที่คาดไว้ตั้งแต่ต้น

ทีมควรถามว่า ใครได้รับผลโดยตรง ใครอาจได้รับผลต่อเนื่อง และยังมีบุคคล กลุ่ม หรือสิ่งแวดล้อมใดที่อาจได้รับผลโดยไม่ได้คาดการณ์ไว้หรือไม่

10.6.3 ระยะเวลาที่ผลกระทบเกิดขึ้น (Timescale)

ระยะเวลาใช้พิจารณาว่าผลกระทบจะเริ่มปรากฏเมื่อใด และอาจคงอยู่เป็นเวลานานเพียงใด

ระยะเวลา	แนวทางการพิจารณา
ระยะสั้น (Short-term)	ผลกระทบเกิดขึ้นในช่วงแรกหรือไม่หลังจากนำระบบไปใช้
ระยะกลาง (Medium-term)	ผลกระทบค่อย ๆ ปรากฏเมื่อระบบถูกใช้งานต่อเนื่องในช่วงระยะเวลาหนึ่ง
ระยะยาว (Long-term)	ผลกระทบสะสมหรือคงอยู่จากการใช้งานระบบอย่างต่อเนื่อง
ข้ามรุ่น (Intergenerational)	ผลกระทบอาจส่งต่อไปยังคนรุ่นถัดไป สังคม หรือสิ่งแวดล้อมในอนาคต

การประเมินนี้ไม่ได้กำหนดแต่ละช่วงเป็นจำนวนเดือนหรือปี ทีมจึงควรกำหนดความหมายให้เหมาะสมกับบริบทของโครงการ และใช้เกณฑ์เดียวกันตลอดการประเมิน นอกจากนี้ ควรคำนึงถึงผลกระทบสะสมจากการใช้งานหลายครั้ง การขยายการใช้งาน หรือการทำงานร่วมกับระบบอื่นด้วย

10.6.4 ความสามารถในการแก้ไขหรือเยียวยา (Remediability)

เกณฑ์นี้ใช้เฉพาะผลกระทบเชิงลบ โดยพิจารณาว่า หากความเสียหายเกิดขึ้นแล้ว สามารถย้อนกลับ แก้ไข ลดผลกระทบ หรือเยียวยาผู้ที่ได้รับผลกระทบได้มากน้อยเพียงใด

ระดับ	แนวทางการพิจารณา
ต่ำมาก (Very Low)	ผลกระทบไม่สามารถย้อนกลับหรือแก้ไขได้
ต่ำ (Low)	แก้ไขได้ยาก และอาจมีความเสียหายหรือผลต่อเนื่องหลงเหลืออยู่
ปานกลาง (Medium)	แก้ไขได้ แต่ต้องใช้เวลา ทรัพยากร หรือสร้างภาระแก่ผู้ที่ได้รับผลกระทบ
สูง (High)	แก้ไขได้โดยไม่เกิดความยุ่งยากหรือความเสียหายต่อเนื่องอย่างมีนัยสำคัญ

การประเมินควรมองจากมุมมองของผู้ที่ได้รับผลกระทบ ไม่ใช่พิจารณาเพียงว่าองค์กรสามารถแก้ไขระบบทางเทคนิคได้หรือไม่ เช่น แม้จะแก้ไขข้อมูลในระบบได้ทันที แต่หากบุคคลสูญเสียโอกาสหรือได้รับความเสียหายไปแล้ว ผลกระทบนั้นอาจยังไม่ถือว่าแก้ไขได้ง่าย

10.6.5 โอกาสเกิดของผลกระทบ (Likelihood)

โอกาสเกิดใช้พิจารณาว่า ภายใต้รูปแบบและบริบทการใช้งานจริง ผลกระทบแต่ละประเด็นมีแนวโน้มเกิดขึ้นมากน้อยเพียงใด

ระดับ	แนวทางการพิจารณา
สูงมาก (Very High)	มีความเป็นไปได้สูงมากและมีแนวโน้มว่าจะเกิดขึ้น
สูง (High)	มีความเป็นไปได้สูงว่าผลกระทบจะเกิดขึ้น
ปานกลาง (Medium)	ผลกระทบมีโอกาสเกิดขึ้นได้
ต่ำ (Low)	ผลกระทบมีโอกาสเกิดขึ้นน้อย

การประเมินไม่ควรอาศัยความรู้สึกเพียงอย่างเดียว แต่ควรพิจารณาข้อมูล เช่น ลักษณะการใช้งาน เหตุการณ์ที่เคยเกิดขึ้น ผลการทดสอบ แรงจูงใจของผู้ใช้งานและผู้พัฒนา นโยบายและกฎหมายที่เกี่ยวข้อง ตลอดจนระดับความตระหนักรู้และความรู้ความสามารถด้าน AI ของผู้ใช้งาน

10.7 การพิจารณามาตรการที่มีอยู่และมาตรการเพิ่มเติม

สำหรับผลกระทบเชิงลบแต่ละประเด็น ทีมควรนำมาตรการและกลไกที่ประเมินไว้ใน Part 2 มาพิจารณาว่าสามารถลดโอกาสเกิดหรือความรุนแรงของผลกระทบได้เพียงใด โดยไม่ควรพิจารณาเพียงว่ามีมาตรการอยู่หรือไม่ แต่ควรดูว่ามาตรการถูกนำไปใช้จริง มีผู้รับผิดชอบ และมีหลักฐานแสดงประสิทธิผลหรือไม่

หากมาตรการที่มีอยู่ยังไม่เพียงพอ ควรระบุมาตรการเพิ่มเติมที่จำเป็น เช่น การปรับปรุงระบบ การเพิ่มการตรวจสอบโดยมนุษย์ การจำกัดขอบเขตการใช้งาน การจัดทำช่องทางร้องเรียน หรือการเยียวยาผู้ที่ได้รับผลกระทบ

10.8 การบันทึกเหตุผลและข้อจำกัดของข้อมูล (Team Notes)

ทีมควรบันทึกเหตุผลที่ใช้เลือกระดับของผลกระทบ ผู้ที่ได้รับผลกระทบ ระยะเวลา โอกาสเกิด และความสามารถในการเยียวยา โดยอ้างอิงข้อมูลหรือหลักฐานที่เกี่ยวข้อง เช่น ผลการทดสอบ ข้อมูลจากผู้มีส่วนได้ส่วนเสีย เหตุการณ์ที่เคยเกิดขึ้น การประเมินก่อนหน้า หรือระบบที่มีลักษณะใกล้เคียง

หากผลกระทบยังมีความไม่แน่นอน ไม่ควรตัดออกจากการวิเคราะห์ แต่ควรบันทึกสมมติฐาน ประเด็นที่ยังไม่ทราบ และช่องว่างของข้อมูลไว้ เพื่อใช้ติดตามและทบทวนเมื่อมีข้อมูลเพิ่มเติม

10.9 สรุปผลการวิเคราะห์และการนำไปใช้ต่อ

ผลจากการจัดทำ Impact Mapping จะช่วยให้องค์กรเห็นภาพรวมของผลกระทบเชิงบวกและเชิงลบที่สำคัญของระบบ AI รวมถึงผู้ที่อาจได้รับผลกระทบ ระดับและช่วงเวลาของผลกระทบ โอกาสเกิด และความสามารถในการแก้ไขหรือเยียวยา

ข้อมูลดังกล่าวจะถูกนำไปใช้พิจารณาว่า มาตรการที่มีอยู่เพียงพอหรือไม่ ประเด็นใดควรได้รับการจัดลำดับความสำคัญ และต้องกำหนดมาตรการลดผลกระทบ การเยียวยา หรือการกำกับดูแลเพิ่มเติมในเรื่องใด โดยควรทบทวนผลการวิเคราะห์เมื่อระบบหรือบริบทการใช้งานเปลี่ยนแปลงไป

บทส่งท้าย

การประเมินผลกระทบทางจริยธรรมไม่ได้มีวัตถุประสงค์เพื่อตัดสินว่าระบบ AI “ผ่าน” หรือ “ไม่ผ่าน” แต่เป็นกระบวนการที่ช่วยให้องค์กรทำความเข้าใจระบบและบริบทการใช้งาน ระบุผู้ที่อาจได้รับผลกระทบ ตรวจสอบความพร้อมของมาตรการกำกับดูแล และวิเคราะห์ผลกระทบเชิงบวกและเชิงลบอย่างเป็นระบบ เพื่อประกอบการตัดสินใจเกี่ยวกับการพัฒนาและนำระบบไปใช้งานอย่างเหมาะสม

ผลจากการประเมินควรถูกนำไปแปลงเป็นการดำเนินงานที่ชัดเจน เช่น การปรับปรุงระบบ การเพิ่มมาตรการป้องกันหรือบรรเทาผลกระทบ การจำกัดขอบเขตการใช้งาน การกำหนดผู้รับผิดชอบ ตลอดจนการติดตามผลภายหลังการนำระบบไปใช้ หากพบว่าผลกระทบบางประเด็นยังไม่สามารถจัดการได้อย่างเพียงพอ องค์กรอาจพิจารณาปรับเปลี่ยน ชะลอ หรือทบทวนแนวทางการใช้งานระบบก่อนดำเนินการต่อ

การประเมินผลกระทบทางจริยธรรมจึงไม่ใช่เอกสารที่จัดทำเพียงครั้งเดียวแล้วสิ้นสุด แต่เป็นข้อมูลสำคัญสำหรับการตัดสินใจและการกำกับดูแลที่ควรได้รับการทบทวนตลอดวงจรชีวิตของระบบ AI โดยเฉพาะเมื่อระบบ ข้อมูล กลุ่มผู้ใช้งาน หรือบริบทการใช้งานมีการเปลี่ยนแปลง เพื่อให้การใช้ AI ยังคงสอดคล้องกับวัตถุประสงค์ขององค์กร เคารพสิทธิของผู้ที่เกี่ยวข้อง และสามารถสร้างประโยชน์ได้โดยลดผลกระทบที่ไม่พึงประสงค์อย่างเหมาะสม

บทที่ 11

การนำไปใช้ผ่านกรณีตัวอย่าง



หมายเหตุ: กรณีใช้งาน (Use Case) ในบทนี้เป็นกรณีสมมุติที่จัดทำขึ้นเพื่อสาธิตการนำ EIA Playbook ไปใช้ ไม่ได้อ้างอิงระบบหรือโครงการจริง

กรณีใช้งานสมมุติ

ระบบ AI ประเมินและแนะนำ

แนวทางส่งเสริมสุขภาพส่วนบุคคล

ระบบนี้ช่วยให้ประชาชนวางแผนดูแลสุขภาพของตนเอง โดยผู้ใช้กำหนดเป้าหมาย ให้ข้อมูลสุขภาพ และเลือกเชื่อมต่ออุปกรณ์สวมใส่ จากนั้น AI จะประเมินแนวโน้มและจัดทำคำแนะนำเฉพาะบุคคล ผู้ใช้เป็นผู้ตัดสินใจนำคำแนะนำไปปฏิบัติและระบบจะแนะนำให้ปรึกษาผู้เชี่ยวชาญเมื่อพบกรณีที่อาจมีความเสี่ยง



เส้นทางการใช้งาน

01	02	03	04	05
กำหนดเป้าหมาย	ให้ข้อมูล	AI วิเคราะห์	รับคำแนะนำ	ติดตามหรือส่งต่อ
ระบุเป้าหมายสุขภาพ	กรอกข้อมูลและเชื่อมต่ออุปกรณ์	ประเมินแนวโน้มและข้อจำกัด	ออกกำลังกายและโภชนาการ	ติดตามผลหรือปรึกษาผู้เชี่ยวชาญ

ข้อมูลของกรณีใช้งาน

ข้อมูลชุดนี้อธิบายระบบ ณ เวลาที่เริ่มการประเมินเพื่อใช้ประกอบการพิจารณาในแต่ละมิติ

01 ระบบนี้ทำอะไรและอยู่ในระยะใด ระบบประเมินแนวโน้มสุขภาพและสร้างคำแนะนำด้านการออกกำลังกายและโภชนาการให้เหมาะกับเป้าหมายพฤติกรรม และข้อจำกัดของผู้ใช้แต่ละราย เพื่อสนับสนุนการดูแลสุขภาพเชิงป้องกัน ไม่ใช่เพื่อวินิจฉัยโรคหรือกำหนดการรักษา	02 ใครใช้และใครอาจได้รับผลกระทบ ผู้ใช้งานหลักคือประชาชนที่ต้องการดูแลสุขภาพเชิงป้องกัน ส่วนบุคลากรด้านสุขภาพที่ได้รับอนุญาตอาจให้คำปรึกษาหรือเข้าถึงข้อมูลเฉพาะเมื่อผู้ใช้ร้องขอความช่วยเหลือหรือเข้าเงื่อนไขที่ต้องส่งต่อ
--	--

ระบบอยู่ระหว่างการพัฒนาต้นแบบและยังไม่ได้ทดลองใช้กับผู้ใช้งาน โดยโครงการมีความตั้งใจจะทดลองใช้ระบบกับผู้ใช้กลุ่มจำกัดในขั้นตอนถัดไป แต่ยังไม่ได้จัดทำแผนการทดลองใช้ การประเมินครั้งนี้จึงดำเนินการในช่วงก่อนนำ AI ไปใช้งาน (Before Deployment) เพื่อระบุผลกระทบที่อาจเกิดขึ้น และนำผลการประเมินไปปรับปรุงการออกแบบระบบ กำหนดมาตรการกำกับดูแล และจัดทำแผนการทดลองใช้ก่อนเริ่มใช้งานจริง	กลุ่มที่ต้องได้รับการพิจารณาเป็นพิเศษ ได้แก่ ผู้มีโรคประจำตัว ผู้สูงอายุ ผู้มีข้อจำกัดทางร่างกาย ผู้ที่มีข้อมูลสุขภาพไม่ครบถ้วนหรือคลาดเคลื่อน และผู้ที่ไม่สามารถเข้าถึงอุปกรณ์สวมใส่หรือบริการดิจิทัลได้อย่างเท่าเทียม นอกจากนี้ ครอบครัวหรือผู้ดูแลอาจได้รับผลกระทบทางอ้อม หากผู้ใช้นำคำแนะนำที่ไม่เหมาะสมไปปฏิบัติ
03 ระบบทำงานและสร้างคำแนะนำอย่างไร ผู้ใช้กำหนดเป้าหมายและให้ข้อมูลสุขภาพ พฤติกรรมในชีวิตประจำวัน และข้อจำกัดของตน รวมถึงอาจเชื่อมต่อข้อมูลจากอุปกรณ์สวมใส่ ระบบจะวิเคราะห์ข้อมูลเพื่อประเมินแนวโน้มและสร้างคำแนะนำเฉพาะบุคคล โดยผู้ใช้เป็นผู้ตัดสินใจนำไปปฏิบัติ อย่างไรก็ตาม ยังไม่ได้ระบุว่าระบบจะแสดงผล แหล่งที่มา หรือระดับความเชื่อมั่นของคำแนะนำอย่างไร รวมถึงวิธีดำเนินการเมื่อข้อมูลไม่ครบถ้วน ขัดแย้งกัน หรืออยู่นอกขอบเขตที่ระบบรองรับ	
04 ใช้ข้อมูลและระบบภายนอกใดบ้าง ระบบใช้ข้อมูลสุขภาพ เช่น ค่าดัชนีมวลกายและเปอร์เซ็นต์ไขมัน ข้อมูลพฤติกรรม เป้าหมายของผู้ใช้ และข้อมูลจากอุปกรณ์สวมใส่ เฉพาะข้อมูลที่ผู้ใช้ให้ความยินยอม ระบบอาจเชื่อมต่อกับบริการคลาวด์ ฐานข้อมูลด้านการออกกำลังกายหรือโภชนาการ และบริการ AI ภายนอก ระบบกำหนดสิทธิ์การเข้าถึงข้อมูลตามบทบาทของผู้เกี่ยวข้อง โดยผู้ใช้สามารถเข้าถึงและจัดการข้อมูลของตนเอง ทีมพัฒนาระบบเข้าถึงเฉพาะข้อมูลที่จำเป็นต่อการตรวจสอบและปรับปรุงระบบ และบุคลากรด้านสุขภาพเข้าถึงข้อมูลเฉพาะกรณีที่ผู้ใช้ร้องขอคำปรึกษาหรือเข้าเงื่อนไขการส่งต่อ อย่างไรก็ตาม ยังไม่ได้ระบุผู้ให้บริการ สถานที่ประมวลผล รายละเอียดด้านการเข้ารหัสและการบันทึกการเข้าถึง ระยะเวลาการเก็บรักษาและการ	05 ใครรับผิดชอบและมนุษย์ตัดสินใจตรงไหน เจ้าของโครงการรับผิดชอบการพัฒนาและตัดสินใจเกี่ยวกับการนำระบบไปทดลองใช้ ทีมเทคนิคดูแลการพัฒนา การเชื่อมต่อข้อมูล และการติดตามการทำงานของระบบ บุคลากรด้านสุขภาพให้คำปรึกษาเมื่อได้รับการร้องขอหรือเมื่อมีการส่งต่อ แต่ยังไม่ได้รับมอบหมายให้ตรวจทานคำแนะนำทุกกรณีก่อนแสดงต่อผู้ใช้ ผู้ใช้เป็นผู้ตัดสินใจว่าจะนำคำแนะนำไปปฏิบัติหรือไม่ และสามารถหยุดใช้ระบบได้ด้วยตนเอง ปัจจุบันโครงการยังไม่ได้กำหนดผู้ตรวจสอบคำแนะนำในกรณีเสี่ยง ผู้มีอำนาจจะจับระบบ และกระบวนการให้ผู้ใช้ขอคำอธิบาย คัดค้าน หรือขอให้ทบทวนคำแนะนำ

ลบข้อมูล ตลอดจนเงื่อนไขการแบ่งปันข้อมูลกับผู้ให้บริการภายนอก	
06 ขอบเขตและกรณีที่มีความเสี่ยง ระบบไม่ใช่เพื่อวินิจฉัยโรค กำหนดการรักษา แนะนำการใช้ยา หรือทดแทนบุคลากรทางการแพทย์ ระบบรับข้อมูลเกี่ยวกับข้อจำกัดทางสุขภาพของผู้ใช้เพื่อประกอบการสร้างคำแนะนำ ความเสี่ยงอาจเกิดจากข้อมูลที่ไม่ครบถ้วนหรือคลาดเคลื่อน คำแนะนำที่หนักเกินข้อจำกัดทางร่างกาย คำแนะนำที่ไม่เหมาะกับผู้มีโรคประจำตัว หรือการที่ผู้ใช้เข้าใจว่าคำแนะนำของระบบเป็นคำวินิจฉัยทางการแพทย์ ปัจจุบันระบบยังไม่มีเงื่อนไขหยุดหรือระงับการสร้างคำแนะนำโดยอัตโนมัติเมื่อพบข้อมูลผิดปกติ หรือกรณีที่อยู่นอกขอบเขต รวมถึงยังไม่ได้กำหนดเกณฑ์ส่งต่อและวิธีแจ้งเตือนที่ชัดเจน	07 การทดลองใช้ การติดตาม และข้อมูลที่ต้องยืนยัน โครงการมีความตั้งใจจะทดลองใช้ระบบกับผู้ใช้กลุ่มจำกัดในขั้นตอนถัดไป แต่ยังไม่ได้จัดทำแผนการทดลองใช้ ได้แก่ กลุ่มเป้าหมาย เกณฑ์คัดเลือก จำนวนผู้เข้าร่วม ระยะเวลาทดลอง วิธีเก็บข้อคิดเห็น และเกณฑ์วัดผลการทดลอง ระบบมีเมนูให้ผู้ใช้รายงานคำแนะนำที่ไม่เหมาะสมหรือข้อผิดพลาดที่พบจากการใช้งาน อย่างไรก็ตาม ยังไม่ได้กำหนดผู้รับผิดชอบในการรับเรื่อง ระยะเวลาตอบสนอง ขั้นตอนการติดตามผล และเกณฑ์การยกระดับเหตุการณ์ ระบบยังไม่มีผลการทดสอบด้านความปลอดภัยและความถูกต้องของคำแนะนำ และยังไม่ได้กำหนดรอบการทดสอบและทวนสอบ กระบวนการจัดการเหตุการณ์ รวมถึงมาตรการกำกับผู้ให้บริการภายนอก ประเด็นเหล่านี้ต้องดำเนินการให้ชัดเจนก่อนเริ่มทดลองใช้

เมื่อเข้าใจข้อมูลตั้งต้นของ Use Case แล้ว ขั้นตอนต่อไปคือการนำระบบมาประเมินตามกระบวนการ EIA เพื่อพิจารณาว่าระบบมีความเหมาะสม มีมาตรการกำกับดูแลเพียงพอ และอาจก่อให้เกิดผลกระทบอย่างไร

การประเมินแบ่งเป็น 3 ส่วน เริ่มจากการกำหนดขอบเขตและบริบทของระบบใน Part 1 จากนั้นประเมินมาตรการตามหลักการด้านจริยธรรมใน Part 2 และนำผลกระทบที่พบมาวิเคราะห์รายละเอียดใน Part 3

ผลที่ได้จะช่วยให้โครงการเห็นทั้งโอกาส ความเสี่ยง และประเด็นที่ต้องปรับปรุง ก่อนกำหนดมาตรการและตัดสินใจเกี่ยวกับการทดลองใช้ระบบในขั้นตอนต่อไป

Part 1 การกำหนดขอบเขตการประเมิน (Scoping Questions)

เมื่อเข้าใจการทำงานของระบบแล้ว ข้อมูลจากกรณีใช้งานสามารถนำมาตอบคำถามสำหรับการกำหนด

ขอบเขตการประเมินได้ ดังตัวอย่างต่อไปนี้

01 AI System Description

คำถามที่ใช้พิจารณา

ระบบ AI นี้ใช้เพื่ออะไร ใครเป็นผู้ใช้งาน
ใครอาจได้รับผลกระทบ และระบบพึ่งพาอะไรบ้าง

ผลการประเมินจาก Use Case

ระบบสนับสนุนการดูแลสุขภาพเชิงป้องกัน โดยวิเคราะห์ข้อมูลสุขภาพ พฤติกรรม เป้าหมาย และข้อจำกัดของผู้ใช้ เพื่อสร้างคำแนะนำด้านการออกกำลังกายและโภชนาการเฉพาะบุคคล ผลลัพธ์เป็นคำแนะนำ ไม่ใช่การวินิจฉัยโรคหรือกำหนดการรักษา และผู้ใช้เป็นผู้ตัดสินใจว่าจะนำคำแนะนำไปปฏิบัติหรือไม่

ระบบยังอยู่ระหว่างการพัฒนาต้นแบบและยังไม่ได้ทดลองใช้กับผู้ใช้งาน การประเมินครั้งนี้จึงดำเนินการในช่วงก่อนนำ AI ไปใช้งาน เพื่อใช้ผลการประเมินปรับปรุงระบบและจัดทำแผนการทดลองใช้

02 Project Governance

คำถามที่ใช้พิจารณา

ใครเป็นผู้รับผิดชอบและมีอำนาจตัดสินใจ
และมีกลไกใดในการทบทวนความเสี่ยงของระบบ

ผลการประเมินจาก Use Case

เจ้าของโครงการรับผิดชอบการพัฒนาและตัดสินใจเกี่ยวกับการนำระบบไปทดลองใช้ ทีมเทคนิคดูแลการพัฒนา การเชื่อมต่อข้อมูล และการติดตามการทำงาน ส่วนบุคลากรด้านสุขภาพให้คำปรึกษาเมื่อได้รับการร้องขอหรือเมื่อมีการส่งต่อ โดยผู้ใช้ยังเป็นผู้ตัดสินใจว่าจะนำคำแนะนำไปปฏิบัติหรือไม่

อย่างไรก็ตาม โครงการยังไม่ได้กำหนดผู้ตรวจสอบคำแนะนำในกรณีเสี่ยง ผู้มีอำนาจระงับระบบ และกระบวนการให้ผู้ใช้ขอคำอธิบาย คัดค้าน หรือขอให้ทบทวนคำแนะนำ

03 Proportionality and Do No Harm

คำถามที่ใช้พิจารณา

การใช้ AI มีความจำเป็นและเหมาะสมหรือไม่
และควรกำหนดขอบเขตอย่างไรเพื่อป้องกันอันตราย

ผลการประเมินจาก Use Case

คำแนะนำสุขภาพทั่วไปสามารถจัดทำได้โดยไม่ใช้ AI แต่ AI ช่วยวิเคราะห์ข้อมูลหลายแหล่งและสร้างคำแนะนำเฉพาะบุคคลได้ จึงมีความเหมาะสมแบบมีเงื่อนไข

ระบบต้องจำกัดบทบาทไว้เฉพาะการส่งเสริมสุขภาพ ไม่ใช่เพื่อวินิจฉัยหรือกำหนดการรักษา และให้ผู้ใช้เป็นผู้ตัดสินใจ ทั้งนี้ คำแนะนำที่คลาดเคลื่อนอาจกระทบต่อสุขภาพและความปลอดภัย โดยเฉพาะผู้ใช้กลุ่มเสี่ยง จึงต้องทดสอบระบบและกำหนดเกณฑ์หยุดหรือส่งต่อให้ชัดเจนก่อนเริ่มทดลองใช้

04 Stakeholder Engagement

คำถามที่ใช้พิจารณา

ผู้มีส่วนได้ส่วนเสียกลุ่มใดควรเข้ามามีส่วนร่วม
และข้อคิดเห็นจะถูกนำไปใช้อย่างไร

ผลการประเมินจาก Use Case

ควรเปิดโอกาสให้ผู้ใช้ทั่วไป ผู้มีโรคประจำตัว ผู้สูงอายุ บุคลากรด้านสุขภาพ ทีมเทคนิค และผู้เชี่ยวชาญด้านข้อมูลส่วนบุคคลมีส่วนร่วมในการออกแบบการทดลองและให้ข้อคิดเห็น เพื่อนำไปปรับปรุงคำแนะนำ ข้อความเตือน และเกณฑ์ส่งต่อผู้เชี่ยวชาญ

ผลลัพธ์จากการกำหนดขอบเขตการประเมิน

ระบบมีการใช้ข้อมูลสุขภาพและสร้างคำแนะนำที่อาจส่งผลกระทบต่อสุขภาพและความปลอดภัยของผู้ใช้โดยตรง จึงควรดำเนินการ EIA ก่อนเริ่มทดลองใช้ โดยประเด็นสำคัญที่ต้องนำไปประเมินต่อ ได้แก่ ความปลอดภัยของคำแนะนำ การคุ้มครองข้อมูลสุขภาพ ความเป็นธรรมต่อผู้ใช้ที่มีข้อจำกัด และการกำกับดูแลโดยมนุษย์

Part 2 การประเมินตามหลักการด้านจริยธรรม (Principles-Based Assessment)

ในกรณีตัวอย่างนี้ จะสาธิตการประเมินเฉพาะมิติด้านความปลอดภัยและความมั่นคงปลอดภัย (Safety and Security)

ประเมินด้านความปลอดภัยและความมั่นคงปลอดภัย

Safety and Security

คำถามตามแบบประเมิน	ผลการประเมินจาก Use Case
01 มีกลไกใดในการประเมินและติดตามความเสี่ยงจากการใช้งานที่ไม่เหมาะสม การใช้งานโดยมิชอบ หรือการโจมตีระบบ AI หรือไม่	ระบบกำหนดสิทธิ์เข้าถึงข้อมูลตามบทบาท และจำกัดขอบเขตไม่ให้วินิจฉัยโรคหรือกำหนดการรักษา เพื่อลดความเสี่ยงจากการเข้าถึงข้อมูลโดยไม่ได้รับอนุญาตและการนำระบบไปใช้เกินวัตถุประสงค์ อย่างไรก็ตาม ยังไม่ได้กำหนดกลไกติดตามพฤติกรรมการใช้งานที่ผิดปกติ การตรวจสอบการใช้งาน และการทดสอบช่องโหว่ของระบบ
02 มีกลไกให้ผู้ใช้งานรายงานข้อผิดพลาดหรือผลกระทบที่เกิดจากระบบ AI หรือไม่	ผู้ใช้งานสามารถรายงานคำแนะนำที่ไม่เหมาะสมหรือข้อผิดพลาดผ่านเมนูภายในระบบได้ แต่ยังไม่ได้กำหนดผู้รับผิดชอบในการรับเรื่อง ระยะเวลาตอบสนอง และกระบวนการติดตามผลหลังได้รับรายงาน
03 ระบบ AI ปฏิบัติตามมาตรฐานหรือกรอบด้านความมั่นคงปลอดภัยใด	ยังไม่ได้ระบุมาตรฐาน กรอบการดำเนินงาน หรือนโยบายด้านความมั่นคงปลอดภัยไซเบอร์ที่ใช้กับระบบ รวมถึงข้อกำหนดสำหรับบริการคลาวด์และบริการ AI ภายนอก
04 มีกลไกใดในการตรวจจับและตอบสนองต่อเหตุการณ์ด้านความปลอดภัยและความมั่นคงปลอดภัย	ยังไม่ได้กำหนดกระบวนการตรวจจับ แจ้งเตือน ตอบสนอง และบันทึกเหตุการณ์ ตลอดจนผู้รับผิดชอบอำนาจในการระงับระบบ วิธีจำกัดผลกระทบ และขั้นตอนการกู้คืน

คำถามตามแบบประเมิน	ผลการประเมินจาก Use Case
05 มีการแจ้งผู้ใช้งานเกี่ยวกับเหตุการณ์ด้านความปลอดภัยและความมั่นคงปลอดภัยที่อาจเกี่ยวข้องกับผู้ใช้งานหรือข้อมูลของผู้ใช้งานอย่างไร	ยังไม่ได้กำหนดว่าเหตุการณ์ประเภทใดต้องแจ้งผู้ใช้งาน ช่องทางและระยะเวลาในการแจ้ง รวมถึงข้อมูลที่ผู้ใช้งานควรได้รับเมื่อระบบหรือข้อมูลสุขภาพของตนได้รับผลกระทบ
06 ระบบ AI ได้ผ่านการทดสอบด้านความปลอดภัยอย่างเข้มงวดและสม่ำเสมอหรือไม่	ระบบยังไม่ผ่านการทดสอบด้านความปลอดภัยอย่างเป็นทางการ โดยโครงการกำหนดให้ต้องทดสอบระบบก่อนเริ่มทดลองใช้ แต่ยังไม่ได้จัดทำแผนและเกณฑ์การทดสอบ ทั้งด้านความปลอดภัย ความถูกต้องของคำแนะนำ ความทนทานเมื่อได้รับข้อมูลไม่ครบถ้วนหรือผิดปกติ และความเสี่ยงจากการเชื่อมต่อกับระบบภายนอก
07 จะมีการทดสอบและทวนสอบระบบ AI บ่อยเพียงใด	ยังไม่ได้กำหนดรอบการทดสอบและทวนสอบ รวมถึงเงื่อนไขที่ต้องประเมินระบบใหม่เมื่อมีการเปลี่ยนโมเดลแหล่งข้อมูล ผู้ให้บริการ หรือขยายการใช้งานไปยังผู้ใช้งานกลุ่มใหม่
08 มีมาตรการเพิ่มเติมเพื่อรักษาความปลอดภัยและความมั่นคงปลอดภัยของระบบ AI และข้อมูลหรือไม่	ระบบใช้ข้อมูลตามความยินยอมและจำกัดสิทธิ์เข้าถึงตามบทบาท แต่ยังไม่ได้กำหนดมาตรการเข้ารหัส การบันทึกการเข้าถึงและการใช้งาน การสำรองและลบข้อมูล การติดตามความผิดปกติ และการกำกับดูแลผู้ให้บริการภายนอก

สรุปผลกระทบเชิงบวกและเชิงลบของระบบ AI ต่อความปลอดภัยและความมั่นคงปลอดภัยโดยย่อ

POSITIVE ผลกระทบเชิงบวก	NEGATIVE ผลกระทบเชิงลบ
• การจำกัดสิทธิ์ตามบทบาทช่วยลดโอกาสที่บุคคลซึ่งไม่เกี่ยวข้องจะเข้าถึงข้อมูลสุขภาพ (จากข้อ 01 และ 08) • การมีช่องทางให้ผู้ใช้งานรายงานคำแนะนำที่ไม่เหมาะสมหรือข้อผิดพลาด อาจช่วยให้ทีมโครงการได้รับข้อมูลปัญหาจากการใช้งานจริงและนำไปปรับปรุงระบบได้ ทั้งนี้ ประโยชน์ดังกล่าวจะเกิดขึ้นได้อย่างมีประสิทธิภาพเมื่อมีการกำหนดผู้รับผิดชอบ ระยะเวลาตอบสนอง และกระบวนการติดตามผลอย่างชัดเจน (จากข้อ 02)	• คำแนะนำที่คลาดเคลื่อนหรือไม่เหมาะสมอาจทำให้ผู้ใช้งานเจ็บหรือมีอาการรุนแรงขึ้น โดยเฉพาะผู้มีโรคประจำตัว ผู้สูงอายุ และผู้มีข้อจำกัดทางร่างกาย ขณะที่โครงการยังไม่มีผลการทดสอบเพื่อยืนยันความปลอดภัยและความถูกต้องของคำแนะนำ (จากข้อ 06–07) • ข้อมูลสุขภาพอาจรั่วไหล ถูกเข้าถึงโดยไม่ได้รับอนุญาต หรือผลลัพธ์ของระบบอาจถูกแทรกแซงเนื่องจากมาตรการคุ้มครองข้อมูลและข้อกำหนดสำหรับผู้ให้บริการภายนอกยังไม่ชัดเจน (จากข้อ 01, 03 และ 08) • เมื่อเกิดเหตุการณ์ด้านความปลอดภัย ความเสียหายอาจไม่ได้รับการตรวจพบหรือจำกัดอย่างทันที่ และผู้ใช้ที่ได้รับผลกระทบอาจไม่ได้รับแจ้งหรือความช่วยเหลือที่จำเป็น เนื่องจากยังไม่มีกระบวนการตรวจจับ ตอบสนอง และแจ้งเหตุการณ์ที่ชัดเจน (จากข้อ 04–05)

Part 3 การวิเคราะห์ผลกระทบทางจริยธรรม

(Impact Mapping)

นำผลกระทบเชิงบวกและเชิงลบที่สรุปไว้ใน Part 2 การประเมินตามหลักการด้านจริยธรรม มาวิเคราะห์ตามเกณฑ์ โดยพิจารณาระดับผลกระทบ ผู้ได้รับผลกระทบ ระยะเวลา และโอกาสเกิด สำหรับผลกระทบเชิงลบ จะพิจารณาเพิ่มเติมถึงความสามารถในการเยียวยา มาตรการที่มีอยู่สามารถลดผลกระทบได้เพียงใด และจำเป็นต้องกำหนดมาตรการเพิ่มเติมหรือไม่

ผลกระทบที่นำมาวิเคราะห์ในกรณีตัวอย่าง

กรณีตัวอย่างนี้แสดงการวิเคราะห์เชิงลึกอย่างละหนึ่งผลกระทบเพื่อให้เห็นวิธีใช้เกณฑ์ ส่วนผลกระทบอื่นยังต้องบันทึกและติดตามต่อไป

ผลกระทบเชิงบวก	ผลกระทบเชิงลบ
ช่องทางรายงานอาจช่วยให้ทีมโครงการค้นพบปัญหาจากการใช้งานจริงและนำไปปรับปรุงระบบได้	คำแนะนำที่คลาดเคลื่อนอาจทำให้ผู้ใช้บาดเจ็บหรือมีอาการรุนแรงขึ้น

วิเคราะห์ผลกระทบเชิงบวก

+ ช่องทางรายงานอาจช่วยให้ทีมโครงการค้นพบปัญหาจากการใช้งานจริงและนำไปปรับปรุงระบบได้		
องค์ประกอบ	ผลการประเมิน	เหตุผลประกอบ
ระดับผลกระทบ (Scale)	ปานกลาง (Medium)	หากมีการรับเรื่องและติดตามผลอย่างชัดเจน ช่องทางรายงานจะช่วยให้ทีมโครงการค้นพบปัญหาและนำข้อมูลไปปรับปรุงระบบได้ อย่างไรก็ตาม ประโยชน์ยังจำกัดอยู่ในระดับระบบและขึ้นอยู่กับการดำเนินการหลังได้รับรายงาน
ผู้ได้รับผลกระทบ (Impacted Parties)	โดยตรง: ผู้ใช้และทีมโครงการ โดยอ้อม: ผู้ใช้ในอนาคตและบุคลากรสุขภาพ	ผู้รายงานได้รับการติดตามปัญหา ทีมโครงการได้รับข้อมูลเพื่อแก้ไขและผู้ใช้รายอื่นอาจได้รับประโยชน์จากระบบที่ปรับปรุงแล้ว

ระยะเวลา (Time-scale)	ระยะสั้น (Short-term)	ประโยชน์อาจเริ่มเกิดขึ้นหลังได้รับรายงานและทีมโครงการดำเนินการตรวจสอบ อย่างไรก็ตาม ปัจจุบันยังไม่ได้กำหนดระยะเวลาตอบสนองและติดตามผล
โอกาสเกิด (Likelihood)	ปานกลาง (Medium)	ระบบมีช่องทางรายงานแล้ว จึงมีโอกาสได้รับข้อมูลปัญหาจากผู้ใช้ แต่ยังไม่ได้รับข้อมูลจากการใช้งานจริง และยังไม่ได้กำหนดผู้รับผิดชอบ ระยะเวลาตอบสนอง และกระบวนการติดตามผล
หมายเหตุของทีม (Team Notes)	เป็นการประเมินก่อนทดลองใช้	ผลการประเมินเป็นการคาดการณ์จากการออกแบบระบบและสมมติฐานการใช้งาน จึงต้องยืนยันด้วยข้อมูลการรายงานและผลการดำเนินงานระหว่างการใช้งานทดลองใช้

วิเคราะห์ผลกระทบเชิงลบและกำหนดมาตรการ

สำหรับผลกระทบเชิงลบ ต้องพิจารณาองค์ประกอบเดียวกับผลกระทบเชิงบวก และเพิ่มความสามารถในการเยียวยา มาตรการที่มีอยู่ และมาตรการเพิ่มเติม

คำแนะนำที่ตลาดเคลื่อนอาจทำให้ผู้ใช้บาดเจ็บหรือมีอาการรุนแรงขึ้น		
องค์ประกอบ	ผลการประเมิน	เหตุผลประกอบ
ระดับผลกระทบ (Scale)	รุนแรง (Serious)	หากผลกระทบเกิดขึ้น อาจทำให้ผู้ใช้บาดเจ็บหรือมีสุขภาพแย่ลงอย่างมีนัยสำคัญในช่วงระยะเวลาหนึ่ง แม้ผลกระทบอาจไม่เกิดกับผู้ใช้ทุกรายหรือไม่ถาวรในทุกกรณี
ผู้ได้รับผลกระทบ (Impacted Parties)	โดยตรง: ผู้ใช้ โดยเฉพาะผู้มีโรคประจำตัว ผู้สูงอายุ และผู้มีข้อจำกัดทางร่างกาย โดยอ้อม: ครอบครัว ผู้ดูแล และบุคลากรสุขภาพ	ผู้ใช้รับอันตรายโดยตรง ขณะที่ครอบครัวหรือผู้ดูแลอาจต้องช่วยดูแล และบุคลากรสุขภาพอาจต้องรักษาหรือแก้ไขผลที่เกิดขึ้น
ระยะเวลา (Time-scale)	ระยะสั้น	การบาดเจ็บอาจเกิดขึ้นทันทีหรือไม่นานหลังผู้ใช้ปฏิบัติตามคำแนะนำ แม้ผลต่อสุขภาพบางกรณีอาจคงอยู่ต่อเนื่องในระยะต่อมา

โอกาสเกิด (Likelihood)	ปานกลาง (Medium)	ข้อมูลของผู้ใช้อาจไม่ครบถ้วนหรือคลาดเคลื่อน และระบบยังไม่มีผลการทดสอบด้านความปลอดภัยของคำแนะนำ อย่างไรก็ตาม คำแนะนำที่คลาดเคลื่อนไม่ได้นำไปสู่อันตรายทุกกรณี เนื่องจากผู้ใช้อังเป็น ผู้ตัดสินใจว่าจะนำคำแนะนำไปปฏิบัติหรือไม่ จึงประเมินโอกาสเกิดในระดับปานกลาง
ความสามารถในการเยียวยา (Remediability)	ต่ำ (Low)	ความเสียหายบางกรณีอาจแก้ไขได้ยาก ต้องได้รับการรักษา และอาจมีผลต่อสุขภาพต่อเนื่องแม้ผู้ใช้หยุดปฏิบัติตามคำแนะนำแล้ว

พิจารณามาตรการที่มีอยู่และมาตรการเพิ่มเติม

องค์ประกอบ	ผลการประเมิน	เหตุผลประกอบ
มาตรการที่มีอยู่	มีบางส่วน แต่ยังไม่เพียงพอ	ระบบกำหนดขอบเขตว่าไม่ใช้เพื่อวินิจฉัยโรค กำหนดการรักษา หรือแนะนำยา และรับข้อมูลข้อจำกัดทางสุขภาพเพื่อประกอบการสร้างคำแนะนำ รวมถึงมีช่องทางให้ผู้ใช้งานรายงานข้อผิดพลาด อย่างไรก็ตาม มาตรการเหล่านี้ยังไม่สามารถยืนยันได้ว่าจะป้องกันคำแนะนำที่ไม่เหมาะสมหรืออันตรายได้ เนื่องจากยังไม่มีผลการทดสอบและหลักฐานแสดงประสิทธิผล
มาตรการเพิ่มเติม	ต้องดำเนินการก่อนทดลองใช้	กำหนดเงื่อนไขหยุดให้คำแนะนำและเกณฑ์ส่งต่อ แสดงคำเตือนที่ชัดเจน ให้นักวิชาการสุขภาพตรวจสอบกรณีเสี่ยง ทดสอบความปลอดภัยกับข้อมูลที่ครอบคลุมผู้ใช้กลุ่มต่าง ๆ และจัดทำกระบวนการรับเรื่อง ตอบสนอง และติดตามเหตุการณ์
หมายเหตุของทีม (Team Notes)	ยังมีช่องว่างของข้อมูล	ระดับผลกระทบ โอกาสเกิด และความสามารถในการเยียวยาเป็นการประเมินเบื้องต้นก่อนทดลองใช้ จึงต้องยืนยันด้วยผลการทดสอบ ข้อคิดเห็นจากบุคลากรสุขภาพ และข้อมูลเหตุการณ์ที่พบระหว่างการทดลองใช้

ข้อสรุปจากการวิเคราะห์

โอกาสที่ควรต่อยอด กำหนดผู้รับผิดชอบและขั้นตอนตอบสนอง เพื่อให้ช่องทางรายงานนำไปสู่การตรวจพบ แก้ไข และเรียนรู้จากปัญหาได้จริง

สิ่งที่ต้องทำก่อนทดลองใช้ จัดทำเกณฑ์หยุดให้คำแนะนำและส่งต่อ ทดสอบความปลอดภัย กำหนดการกำกับดูแลโดยมนุษย์ และจัดทำกระบวนการจัดการเหตุการณ์ให้ชัดเจน

การนำผลการวิเคราะห์ไปใช้

ข้อสรุป

ผลการวิเคราะห์สนับสนุนให้โครงการพัฒนาระบบต่อไปได้แบบมีเงื่อนไข โดยต้องดำเนินการมาตรการที่ระบุข้างต้น ให้แล้วเสร็จ และตรวจสอบความพร้อมก่อนอนุมัติให้เริ่มทดลองใช้กับผู้ใช้กลุ่มจำกัด

ระหว่างการทดลองใช้ โครงการควรกำหนดวิธีติดตามความปลอดภัยและความถูกต้องของคำแนะนำ เก็บ ข้อคิดเห็นจากผู้ใช้งานและบุคลากรสุขภาพ และบันทึกข้อผิดพลาดหรือเหตุการณ์ที่อาจส่งผลกระทบต่อสุขภาพ ของผู้ใช้

หลังการทดลองใช้ ควรนำผลการทดสอบ ข้อคิดเห็น และข้อมูลเหตุการณ์ที่เกิดขึ้นกลับมาทบทวนและ ปรับปรุง Impact Mapping ก่อนตัดสินใจพัฒนาระบบต่อไป ปรับขอบเขตการใช้งาน หรือขยายการใช้งานไป ยังผู้ใช้กลุ่มอื่น

อภิธานศัพท์

อภิธานศัพท์นี้รวบรวมคำศัพท์สำคัญที่ใช้ในคู่มือการประเมินผลกระทบทางจริยธรรมของปัญญาประดิษฐ์ เพื่อให้ผู้ประเมินและผู้ที่เกี่ยวข้องเข้าใจความหมายและใช้ถ้อยคำในทิศทางเดียวกัน ความหมายที่กำหนดเป็นการอธิบายในบริบทของคู่มือฉบับนี้ โดยไม่ได้มีวัตถุประสงค์เพื่อใช้แทนนิยามตามกฎหมาย มาตรฐาน หรือข้อกำหนดเฉพาะที่เกี่ยวข้อง

1. คำศัพท์เกี่ยวกับระบบ AI และกระบวนการประเมิน

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ใน คู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
AI SYSTEM	ระบบ ปัญญาประดิษฐ์ หรือระบบ AI	ระบบที่ใช้เทคนิคหรือองค์ประกอบของปัญญาประดิษฐ์ในการประมวลผลข้อมูลและสร้างผลลัพธ์ เช่น การคาดการณ์ เนื้อหา คำแนะนำ หรือการตัดสินใจ ซึ่งอาจส่งผลกระทบต่อสภาพแวดล้อมบุคคล หรือกระบวนการที่ระบบถูกนำไปใช้	[5][11]
AI USE CASE	กรณีการใช้งาน AI	การนำระบบ AI ไปใช้เพื่อวัตถุประสงค์เฉพาะภายในกระบวนการ งาน หรือบริการ โดยต้องพิจารณาร่วมกันทั้งวัตถุประสงค์ ผู้ใช้งาน ข้อมูลบทบาทของระบบ และบริบทการนำไปใช้	[4][5]
AI LIFECYCLE	วงจรชีวิตของระบบ AI	ขั้นตอนที่เกี่ยวข้องกับระบบ AI ตั้งแต่การกำหนดวัตถุประสงค์ ออกแบบ จัดหาและรวบรวมข้อมูล พัฒนา ทดสอบ นำไปใช้งาน ติดตาม ปรับปรุง จนถึงการยุติการใช้งาน	[1][5][11]
ETHICAL IM- PACT ASSESS- MENT: EIA	การประเมินผล กระทบทาง จริยธรรม	กระบวนการที่ช่วยให้องค์กรพิจารณาว่าระบบ AI สอดคล้องกับหลักการด้านจริยธรรมหรือไม่ รวมถึงระบุและวิเคราะห์ผลกระทบเชิงบวกและเชิงลบที่อาจเกิดขึ้น เพื่อกำหนดมาตรการและใช้ประกอบการตัดสินใจ	[1][4]
SCOPING QUESTIONS	การกำหนดขอบเขต การประเมิน	ขั้นตอนที่ใช้รวบรวมข้อมูลพื้นฐานเกี่ยวกับวัตถุประสงค์ ลักษณะการทำงาน ผู้ใช้งาน ผู้ที่อาจได้รับผลกระทบ กลไกการกำกับดูแล และ	[4]

		บริบทของระบบ AI ก่อนเข้าสู่การประเมินในแต่ละมิติ	
PRINCIPLES-BASED QUESTIONS	การประเมินตามหลักการด้านจริยธรรม	ชุดคำถามสำหรับพิจารณาว่าองค์กรมีมาตรการและกลไกการกำกับดูแลที่เหมาะสมและเพียงพอต่อการดำเนินงานตามหลักการด้านจริยธรรมในแต่ละมิติหรือไม่	[1][4]
IMPACT MAPPING	การวิเคราะห์ผลกระทบทางจริยธรรม	ขั้นตอนที่นำข้อมูลจากการกำหนดขอบเขตและการประเมินตามหลักการด้านจริยธรรมมาระบุและวิเคราะห์ผลกระทบเชิงบวกและเชิงลบ รวมถึงผู้ที่ได้รับผลกระทบ ระดับของผลกระทบ โอกาสเกิดและแนวทางจัดการผลกระทบ	[4]
OUTPUT AND OUTCOME	ผลผลิตและผลลัพธ์	Output หมายถึงผลที่ระบบสร้างขึ้นโดยตรง เช่น กระบวนการ คำแนะนำ หรือข้อความ ส่วน Outcome หมายถึงการเปลี่ยนแปลงหรือผลที่เกิดขึ้นเมื่อผลผลิตดังกล่าวถูกนำไปใช้กับบุคคล กระบวนการ หรือการตัดสินใจ	[4][5]
POSITIVE AND NEGATIVE IMPACTS	ผลกระทบเชิงบวกและเชิงลบ	ผลกระทบเชิงบวกหมายถึงประโยชน์หรือการเปลี่ยนแปลงที่พึงประสงค์จากการใช้ระบบ AI ส่วนผลกระทบเชิงลบหมายถึงอันตราย ความเสียหาย หรือผลที่ไม่พึงประสงค์ที่อาจเกิดขึ้นต่อบุคคล กลุ่มบุคคล สังคม หรือสิ่งแวดล้อม	[1][4]
MITIGATION MEASURES	มาตรการลดผลกระทบ	มาตรการที่กำหนดขึ้นเพื่อป้องกัน ลด ควบคุม แก้ไข หรือเยียวยาผลกระทบเชิงลบที่อาจเกิดขึ้นจากการพัฒนาและใช้งานระบบ AI	[1][4][5]
SAFEGUARD	มาตรการป้องกัน	มาตรการ กลไก ขั้นตอน หรือการควบคุมด้านเทคนิคและการบริหารจัดการที่ช่วยป้องกันหรือลดความเสี่ยงและผลกระทบที่อาจเกิดขึ้นจากระบบ AI	[4][5]

2. คำศัพท์เกี่ยวกับขอบเขตและการกำกับดูแลโครงการ

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ใน คู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
PROJECT GOVERNANCE	กลไกการกำกับดูแล โครงการ AI	โครงสร้าง บทบาท อำนาจตัดสินใจ ความ รับผิดชอบ และกระบวนการที่องค์กรกำหนดขึ้น เพื่อกำกับดูแลการออกแบบ พัฒนา จัดหา ใช้ งาน และติดตามระบบ AI	[5][12]
STAKEHOLDER	ผู้มีส่วนได้ส่วนเสีย	บุคคล กลุ่มบุคคล หรือองค์กรที่มีส่วนเกี่ยวข้อง สามารถส่งผลกระทบต่อระบบ หรืออาจได้รับผลกระทบ จากการออกแบบ พัฒนา ใช้งาน และผลลัพธ์ ของระบบ AI	[1][4][11]
STAKEHOLDER ENGAGEMENT	การมีส่วนร่วมของผู้ มีส่วนได้ส่วนเสีย	กระบวนการระบุ รับฟัง ปรึกษาหารือ และนำ ความคิดเห็นของผู้มีส่วนได้ส่วนเสียมาใช้ ประกอบการออกแบบ การตัดสินใจ การ ประเมินผลกระทบ และการกำกับดูแลระบบ AI	[1][4]
INTENDED USER	ผู้ใช้งานที่ระบบมุ่ง หมายให้ใช้	บุคคลหรือกลุ่มบุคคลที่ระบบได้รับการออกแบบ และพัฒนาขึ้นเพื่อให้ใช้งานโดยตรงตาม วัตถุประสงค์และเงื่อนไขที่กำหนด	[4][11]
IMPACTED PARTIES	ผู้ที่ได้รับผลกระทบ	บุคคล กลุ่มบุคคล ชุมชน สังคม หรือ สิ่งแวดล้อมที่อาจได้รับผลกระทบจากระบบ AI ทั้งโดยตรง โดยอ้อม หรือในลักษณะที่ไม่ได้ คาดการณ์ไว้ แม้ไม่ได้เป็นผู้ใช้งานระบบด้วย ตนเอง	[4]
PROPORTIONALITY SCREENING	การคัดกรองความ สมเหตุสมผลและ ความได้สัดส่วน	การพิจารณาว่าการนำ AI มาใช้มีความจำเป็น เหมาะสม และได้สัดส่วนกับวัตถุประสงค์หรือไม่ รวมถึงมีทางเลือกอื่นที่สามารถบรรลุ วัตถุประสงค์โดยก่อให้เกิดความเสี่ยงหรือ ผลกระทบน้อยกว่าหรือไม่	[1][4]
DO NO HARM	หลักการไม่ก่อให้เกิด อันตราย	หลักการที่กำหนดให้องค์กรระบุ ป้องกัน และลด อันตรายที่อาจเกิดขึ้นจากระบบ AI โดยเฉพาะ ผลกระทบต่อสิทธิมนุษยชน ศักดิ์ศรี ความ ปลอดภัย เสรีภาพ สังคม และสิ่งแวดล้อม	[1][4]

3. คำศัพท์เกี่ยวกับความเป็นธรรมและความสามารถในการเข้าถึง

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ใน คู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
ALGORITHM- MIC BIAS	อคติของระบบ AI	แนวโน้มเชิงระบบที่ทำให้ระบบ AI สร้างผลลัพธ์ที่ไม่เป็นธรรม หรือส่งผลแตกต่างกันต่อบุคคลหรือกลุ่มบุคคล โดยอาจเกิดจากข้อมูล การออกแบบโมเดล วิธีทดสอบ การตีความผลลัพธ์ หรือบริบทการนำระบบไปใช้	[1][5]
FAIRNESS AUDIT	การตรวจสอบประเมิน ความเป็นธรรม	การตรวจสอบข้อมูล โมเดล และผลลัพธ์ของระบบ อย่างเป็นระบบ เพื่อประเมินว่าระบบสร้างผลลัพธ์ที่แตกต่างหรือก่อให้เกิดความเสียเปรียบต่อผู้ใช้งานหรือผู้ที่ได้รับผลกระทบบางกลุ่มหรือไม่	[5]
ACCESSIBIL- ITY	ความสามารถใน การเข้าถึงและใช้ งาน	การออกแบบระบบและข้อมูลให้ผู้ใช้งานที่มีความหลากหลายสามารถเข้าถึง เข้าใจ และใช้งานได้ รวมถึงผู้พิการ ผู้สูงอายุ ผู้มีข้อจำกัดด้านภาษา ทักษะดิจิทัล อุปกรณ์ หรือวิธีการรับรู้ข้อมูล	[1][9]

4. คำศัพท์เกี่ยวกับความยั่งยืน

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ใน คู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
CARBON FOOT- PRINT	คาร์บอนฟุตพริ้นท์	ปริมาณการปล่อยก๊าซเรือนกระจกที่เกี่ยวข้องกับการพัฒนาและใช้งานระบบ AI โดยอาจครอบคลุม การฝึกสอนและเรียกใช้โมเดล การใช้พลังงานของ ศูนย์ข้อมูล ตลอดจนการผลิต ใช้งาน และกำจัด ฮาร์ดแวร์	[2][3]
LIFECYCLE EN- VIRONMENTAL IMPACT	ผลกระทบต่อ สิ่งแวดล้อมตลอด วงจรชีวิต	ผลกระทบที่เกิดขึ้นตั้งแต่การผลิตฮาร์ดแวร์ การจัดหาและประมวลผลข้อมูล การพัฒนาและใช้งาน ระบบ ไปจนถึงการยุติการใช้งานและการจัดการ อุปกรณ์	[1][2]
CARBON INTEN- SITY OF COM- PUTE	ความเข้มข้นของ คาร์บอนจากการ ประมวลผล	ปริมาณการปล่อยก๊าซเรือนกระจกที่สัมพันธ์กับ พลังงานไฟฟ้าที่ใช้ในการประมวลผล ซึ่งอาจแตกต่างกันตามแหล่งพลังงาน ช่วงเวลา และ สถานที่ตั้งของศูนย์ข้อมูล	[2][3]

5. คำศัพท์เกี่ยวกับความเป็นส่วนตัวและการคุ้มครองข้อมูล

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ ในคู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
PERSONAL DATA	ข้อมูลส่วนบุคคล	ข้อมูลเกี่ยวกับบุคคลที่ทำให้สามารถระบุตัวบุคคลนั้นได้ ไม่ว่าทางตรงหรือทางอ้อม โดยอาจรวมถึงชื่อ ข้อมูลติดต่อ ตัวระบุออนไลน์ ข้อมูลพฤติกรรม ภาพ เสียง หรือข้อมูลอื่นที่สามารถเชื่อมโยงถึงบุคคลได้	[10]
DATA ANONY-MIZATION	การทำให้ข้อมูลไม่สามารถระบุตัวบุคคลได้	กระบวนการลดหรือตัดข้อมูลระบุตัวตนและความเชื่อมโยงต่าง ๆ จนไม่สามารถระบุกลับไปยังบุคคลได้ด้วยวิธีการที่สมเหตุสมผล หากข้อมูลยังสามารถเชื่อมโยงกลับไปยังบุคคลได้ ข้อมูลนั้นอาจยังคงมีความเสี่ยงต่อการระบุตัวบุคคล	[10][13][14]
PSEUDONY-MIZATION	การใช้นามแฝง	การแทนที่ข้อมูลที่ระบุตัวบุคคลโดยตรงด้วยรหัสหรือนามแฝง โดยเก็บข้อมูลที่เชื่อมโยงกลับไปยังบุคคลแยกต่างหากและมีมาตรการคุ้มครองที่เหมาะสม การใช้นามแฝงช่วยลดความเสี่ยงแต่ไม่ได้ทำให้ข้อมูลหมดสภาพเป็นข้อมูลส่วนบุคคลโดยอัตโนมัติ	[8][13][14]
PRIVACY BY DESIGN	การคุ้มครองความเป็นส่วนตัวตั้งแต่การออกแบบ	การนำหลักการและมาตรการคุ้มครองความเป็นส่วนตัวมาใช้ตั้งแต่เริ่มออกแบบและทดลองวงจรชีวิตของระบบ เช่น การใช้ข้อมูลเท่าที่จำเป็น การจำกัดวัตถุประสงค์ การควบคุมสิทธิ์ และการกำหนดค่าพื้นฐานที่คุ้มครองข้อมูล	[8][10]
DATA PROTECTION IMPACT ASSESSMENT: DPIA	การประเมินผลกระทบด้านการคุ้มครองข้อมูลส่วนบุคคล	กระบวนการประเมินการประมวลผลข้อมูลส่วนบุคคลอย่างเป็นระบบ เพื่อระบุความเสี่ยงต่อสิทธิและเสรีภาพของบุคคล รวมถึงกำหนดมาตรการจัดการความเสี่ยงก่อนเริ่มใช้งานหรือเมื่อระบบมีการเปลี่ยนแปลงที่สำคัญ	[8][10]

6. คำศัพท์เกี่ยวกับการกำกับดูแลและการตัดสินใจโดยมนุษย์

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ใน คู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
HUMAN OVER-SIGHT	การกำกับดูแลโดยมนุษย์	กระบวนการและกลไกที่ทำให้มนุษย์สามารถติดตาม ตรวจสอบ ทำความเข้าใจ และทบทวนการทำงานของหรือผลลัพธ์ของระบบ AI รวมถึงสามารถแทรกแซงได้เมื่อพบข้อผิดพลาดหรือความเสี่ยง	[1][5]
HUMAN OVER-RIDE	การเปลี่ยนแปลงหรือยกเลิกผลลัพธ์โดยมนุษย์	กลไกที่เปิดโอกาสให้บุคคลที่ได้รับมอบอำนาจสามารถปฏิเสธ เปลี่ยนแปลง ย้อนกลับ หยุด หรือยกเลิกการตัดสินใจและการดำเนินการของระบบ AI ได้	[1][5]
HUMAN-IN-THE-LOOP: HITL	มนุษย์มีส่วนร่วมในกระบวนการตัดสินใจ	รูปแบบการทำงานที่กำหนดให้มนุษย์เข้ามาตรวจสอบ ให้ความเห็นชอบ หรือดำเนินการบางขั้นตอนก่อนที่จะผลลัพธ์ของระบบ AI จะถูกนำไปใช้หรือก่อให้เกิดการตัดสินใจ	[5]
AUTOMATION BIAS	อคติจากการพึ่งพาระบบอัตโนมัติ	แนวโน้มที่มนุษย์จะเชื่อถือหรือยอมรับผลลัพธ์ของระบบ AI มากเกินไป โดยไม่ได้ตรวจสอบข้อจำกัด พิจารณาข้อมูลจากแหล่งอื่น หรือใช้วิจารณญาณของตนเองอย่างเพียงพอ	[5][6]

7. คำศัพท์เกี่ยวกับความรับผิดชอบและการรับผิดชอบ

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ใน คู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
RESPONSIBILITY	ความรับผิดชอบ	หน้าที่ บทบาท และขอบเขตการดำเนินงานที่กำหนดให้บุคคลหรือหน่วยงานรับผิดชอบในแต่ละช่วงของวงจรชีวิตระบบ AI	[1][5][12]
ACCOUNTABILITY	การรับผิดชอบ	การที่บุคคลหรือหน่วยงานต้องสามารถชี้แจงเหตุผล แสดงหลักฐาน ยอมรับการตรวจสอบ และรับผิดชอบต่อการทำงานและผลที่เกิดขึ้นจากระบบ AI รวมถึงดำเนินการแก้ไขหรือเยียวยาเมื่อเกิดผลกระทบ	[1][5][12]
TRACEABILITY	การตรวจสอบย้อนกลับ	ความสามารถในการติดตามที่มาของข้อมูล รุ่นของโมเดล ขั้นตอนการประมวลผล การเปลี่ยนแปลงระบบ	[5][12]

		การตัดสินใจ และการแทรกแซงของมนุษย์ตลอดวงจรชีวิตของระบบ AI	
AUDIT TRAIL	บันทึกสำหรับการตรวจสอบย้อนหลัง	บันทึกเหตุการณ์ที่จัดเก็บอย่างเป็นลำดับเวลา เช่น ข้อมูลนำเข้า รุ่นของโมเดล ผลลัพธ์ การเปลี่ยนแปลงระบบ การตัดสินใจ และการแทรกแซงของมนุษย์ เพื่อให้สามารถตรวจสอบและทวนสอบการดำเนินงานได้	[5][12]
AUDITABILITY	ความสามารถในการตรวจสอบประเมิน	คุณลักษณะที่ทำให้กระบวนการ ข้อมูล การตัดสินใจ และผลลัพธ์ของระบบมีหลักฐานเพียงพอสำหรับการทบทวนและตรวจสอบโดยทีมภายใน ผู้ตรวจสอบอิสระ หรือหน่วยงานที่เกี่ยวข้อง	[5][12]

8. คำศัพท์เกี่ยวกับความโปร่งใส ความสามารถในการอธิบาย และความตระหนักรู้

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ ในคู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
TRANSPARENCY	ความโปร่งใส	การเปิดเผยข้อมูลเกี่ยวกับการมีอยู่ วัตถุประสงค์ บทบาท ขอบเขต ผู้รับผิดชอบ ข้อมูลที่ใช้ และข้อจำกัดของระบบ AI อย่างเหมาะสมกับผู้ใช้งาน ผู้ที่ได้รับผลกระทบ และผู้มีส่วนได้ส่วนเสีย	[1][5]
EXPLAINABILITY	ความสามารถในการอธิบาย	ความสามารถในการให้คำอธิบายเกี่ยวกับการทำงานและผลลัพธ์ของระบบ โดยทำให้ผู้รับข้อมูลเข้าใจบทบาทของ AI ปัจจัยสำคัญที่มีส่วนต่อผลลัพธ์ ตลอดจนข้อจำกัดและความไม่แน่นอนที่เกี่ยวข้อง	[1][5]
AI LITERACY	ความรู้ความสามารถด้าน AI	ความรู้ ทักษะ และวิจารณญาณที่ช่วยให้บุคคลสามารถเข้าใจบทบาท หลักการทำงาน ความสามารถ ข้อจำกัด ความเสี่ยง และผลกระทบของ AI รวมถึงสามารถใช้และพิจารณาผลลัพธ์ของระบบได้อย่างเหมาะสม	[1][5]

9. คำศัพท์เกี่ยวกับเกณฑ์วิเคราะห์ผลกระทบ

คำศัพท์ ภาษาอังกฤษ	คำภาษาไทยที่ใช้ ในคู่มือ	ความหมายในบริบทของคู่มือ	อ้างอิง
SCALE	ระดับของผลกระทบ	ระดับความสำคัญหรือขนาดของประโยชน์และความเสียหาย หากผลกระทบนั้นเกิดขึ้นจริง โดยพิจารณา	[4]

	ผลกระทบต่อบุคคล กลุ่มบุคคล สิทธิ สุขภาพ สังคม สิ่งแวดล้อม หรือโครงสร้างพื้นฐาน		
IMPACTED PARTIES	ผู้ที่ได้รับผลกระทบ	ผู้ที่ได้รับผลกระทบจากระบบโดยตรง โดยอัตโนมัติ หรือในลักษณะที่ไม่ได้คาดการณ์ไว้ โดยควรพิจารณาให้กว้างกว่าผู้ใช้งานระบบเพียงอย่างเดียว	[4]
TIMESCALE	ระยะเวลาของผลกระทบ	ช่วงเวลาที่ผลกระทบเริ่มปรากฏและระยะเวลาที่ผลกระทบอาจคงอยู่ เช่น ระยะสั้น ระยะกลาง ระยะยาว หรือข้ามรุ่น	[4]
LIKELIHOOD	โอกาสเกิด	ระดับความเป็นไปได้ที่ผลกระทบจะเกิดขึ้นภายใต้ลักษณะและบริบทการใช้งานจริงของระบบ AI	[4][5]
REMEDIABIL- ITY	ความสามารถในการแก้ไขหรือเยียวยา	ระดับที่ผลกระทบเชิงลบสามารถย้อนกลับ แก้ไข ลดความเสียหาย หรือเยียวยาผู้ที่ได้รับผลกระทบได้ โดยควรพิจารณาจากมุมมองของผู้ได้รับผลกระทบ ไม่ใช่เฉพาะความสามารถในการแก้ไขระบบทางเทคนิค	[4]
TEAM NOTES	หมายเหตุของทีม	พื้นที่สำหรับบันทึกเหตุผล หลักฐาน สมมติฐาน ความไม่แน่นอน ข้อจำกัด และช่องว่างของข้อมูลที่ใช้ประกอบการวิเคราะห์ผลกระทบ	

บรรณานุกรม

- [1] United Nations Educational, Scientific and Cultural Organization (UNESCO), "Recommendation on the Ethics of Artificial Intelligence," adopted Nov. 23, 2021.
- [2] Organisation for Economic Co-operation and Development (OECD), "Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications: The AI Footprint," *OECD Digital Economy Papers*, no. 341, OECD Publishing, Paris, France, Nov. 2022, doi: 10.1787/7babf571-en.
- [3] International Energy Agency (IEA), "Energy and AI," IEA, Paris, France, 2025.
- [4] United Nations Educational, Scientific and Cultural Organization (UNESCO), "Ethical Impact Assessment," *Global AI Ethics and Governance Observatory*. [Online]. Accessed: Jul. 30, 2026.
- [5] E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, National Institute of Standards and Technology, Gaithersburg, MD, USA, Jan. 2023, doi: 10.6028/NIST.AI.100-1.
- [6] C. Autio, R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall, and K. Roberts, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," NIST AI 600-1, National Institute of Standards and Technology, Gaithersburg, MD, USA, Jul. 2024, doi: 10.6028/NIST.AI.600-1.
- [7] A. Vassilev, A. Oprea, A. Fordyce, H. Anderson, X. Davies, and M. Hamin, "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations," NIST AI 100-2e2025, National Institute of Standards and Technology, Gaithersburg, MD, USA, Mar. 2025, doi: 10.6028/NIST.AI.100-2e2025.
- [8] K. Boeckl and N. Lefkovitz, "NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0," National Institute of Standards and Technology, Gaithersburg, MD, USA, Jan. 2020, doi: 10.6028/NIST.CSWP.01162020.
- [9] World Wide Web Consortium (W3C), "Web Content Accessibility Guidelines (WCAG) 2.2," W3C Recommendation, Dec. 12, 2024.
- [10] ราชกิจจานุเบกษา, "พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562," เล่ม 136 ตอนที่ 69 ก, 27 พฤษภาคม 2562.

- [11] International Organization for Standardization and International Electrotechnical Commission, "ISO/IEC 22989:2022 Information Technology—Artificial Intelligence—Artificial Intelligence Concepts and Terminology," ISO/IEC 22989:2022, Geneva, Switzerland, Jul. 2022.
- [12] International Organization for Standardization and International Electrotechnical Commission, "ISO/IEC 42001:2023 Information Technology—Artificial Intelligence—Management System," ISO/IEC 42001:2023, Geneva, Switzerland, Dec. 2023.
- [13] International Organization for Standardization and International Electrotechnical Commission, "ISO/IEC 27559:2022 Information Security, Cybersecurity and Privacy Protection—Privacy Enhancing Data De-identification Framework," ISO/IEC 27559:2022, Geneva, Switzerland, Nov. 2022.
- [14] International Organization for Standardization and International Electrotechnical Commission, "ISO/IEC 20889:2018 Privacy Enhancing Data De-identification Terminology and Classification of Techniques," ISO/IEC 20889:2018, Geneva, Switzerland, Nov. 2018.

คณะทำงาน AI Governance Practice Center (AIGPC)

รจนา ลำเลิศ

ธิดิกร ตระกูลศิริศักดิ์

กัลยรัตน์ เจริญปฐ

ธีระภัทร เจริญปฐ

หทัยชนก พุทธรักษา

หัตถพงษ์ ศิริรัตน์

สุฉายพิมพ์ ศิริวัฒน์

ณัษชา หล่อกรรพย์

บวิญญูทัย ยอดเวียงไชย

ETHICAL IMPACT ASSESSMENT PLAYBOOK

(ร่าง) คู่มือการประเมินผลกระทบ
ทางจริยธรรมปัญญาประดิษฐ์



ศูนย์ปฏิบัติการด้านธรรมาภิบาลปัญญาประดิษฐ์
AI Governance Practice Center : AIGPC

สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์
☎ 02 123 1237 ✉ aigpc@etda.or.th