การกำกับปัญญาประดิษฐ์และมาตรฐานทางเทคนิค: มาตรฐานข้อมูลเพื่อความรับผิดได้ระหว่างวงจรชีวิตของระบบปัญญาประดิษฐ์

(ฉบับย่อตีพิมพ์ใน วารสารกฤษฎีกา เล่ม 4 ตอน 4 พ.ศ. 2567)

อาทิตย์ สุริยะวงศ์กุล

ADAPT Centre, Trinity College Dublin

ด้วยความสามารถที่ใช้ได้จริงมากขึ้นและราคาที่เข้าถึงได้มากขึ้น ทำให้มีการประยุกต์ใช้ปัญญาประดิษฐ์อย่างแพร่หลาย ทั้งในตัวบริการและผลิตภัณฑ์ที่บุคคลทั่วไป เช่น ผู้บริโภค สามารถมองเห็นและสัมผัสได้โดยตรง และทั้งในกระบวนการผลิตและการจัดการองค์กร ที่บุคคลภายนอกองค์กรไม่สามารถมองเห็นหรือสัมผัสได้โดยตรง บริการและผลิตภัณฑ์บางส่วนที่นำปัญญาประดิษฐ์ไปใช้ เป็นบริการและผลิตภัณฑ์ที่อาจเกี่ยวข้องกับความปลอดภัยของบุคคลหรือของสาธารณะ จึงมีความพยายามในการกำกับการใช้ปัญญาประดิษฐ์ในกิจการดังกล่าว เช่นเดียวกับการกำกับกิจการเพื่อความปลอดภัยอื่นๆ ที่มีเครื่องมือหลากหลาย ทั้งกฎระเบียบ ระบบใบอนุญาต มาตรการทางเทคโนโลยี มาตรฐาน และการตรวจสอบ รวมถึงระบบแรงจูงใจทางเศรษฐกิจ การกำกับกิจการที่ใช้ปัญญาประดิษฐ์ก็มีชุดเครื่องมือที่มีลักษณะคล้ายกัน บทความนี้จะพิจารณาถึงการใช้มาตรฐานทางเทคนิคกับการกำกับปัญญาประดิษฐ์

หลักการด้านจริยธรรมและการกำกับปัญญาประดิษฐ์

ตั้งแต่ประมาณ พ.ศ. 2560 ภาคเอกชนและภาครัฐทั่วโลกเริ่มให้ความสนใจพิจารณาประเด็นด้านจริยธรรมที่เกี่ยวข้องกับเทคโนโลยีปัญญาประดิษฐ์ และทยอยออกหลักการหรือแนวปฏิบัติสำหรับองค์กรของตัวเองหรือสำหรับหน่วยงานที่องค์กรของตนให้ทุนสนับสนุนหรือร่วมงานด้วย Fjeld et al. 2020, Zhou et al. 2020 และ Mbiazi et al. 2023 เป็นตัวอย่างของงานที่สำรวจหลักการเหล่านี้และพยายามจัดจำแนก-เปรียบเทียบให้เห็นความคล้ายคลึงและลักษณะที่แตกต่างของหลักการจากแต่ละหน่วยงาน โดย Zhou et al. ได้พยายามจำแนกหลักการออกเป็นสองกลุ่มใหญ่คือกลุ่มแรก “ปัญญาประดิษฐ์ที่มีจริยธรรม” (Ethical AI) ที่เกี่ยวข้องกับการทำให้ระบบปัญญาประดิษฐ์สามารถทำงานได้ตามหลักการจริยธรรม และกลุ่มที่สอง “จริยธรรมปัญญาประดิษฐ์” (AI Ethics) ที่เกี่ยวกับผลกระทบต่อมนุษย์และสังคมจากการมีอยู่ของระบบปัญญาประดิษฐ์ (ซึ่งบางส่วนของทั้งสองกลุ่มทับซ้อนกับจริยธรรมทางธุรกิจและจริยธรรมทางข้อมูล) ในขณะที่ Fjeld et al. และ Mbiazi et al. ได้จำแนกหลักการต่างๆ เข้าหมวดใหญ่ 8 และ 6 หมวด ดังนี้

Fjeld et al. 2020	**Mbiazi et al. 2023**
Privacy	Privacy and data protection

(ความเป็นส่วนตัว)	(ความเป็นส่วนตัวและการคุ้มครองข้อมูล)
Accountability (ความรับผิดได้)	Responsibility and Accountability (ความรับผิดชอบและความรับผิดได้)
Safety and Security (ความปลอดภัยและมั่นคง)	Safety, security and robustness (ความปลอดภัย มั่นคง และแข็งแกร่ง)
Transparency and Explainability (ความโปร่งใสและอธิบายได้)	Transparency and explainability (ความโปร่งใสและอธิบายได้)
Fairness and Non-discrimination (ความเป็นธรรมและไม่แบ่งแยก)	Fairness and equity (ความเป็นธรรมและเท่าเทียม)
Human Control of Technology (มนุษย์เป็นผู้ควบคุมเทคโนโลยี)	
Professional Responsibility (ความรับผิดชอบทางวิชาชีพ)	
Promotion of Human Values (สนับสนุนคุณค่าของมนุษย์)	Social and environmental impact (ผลกระทบต่อสังคมและสิ่งแวดล้อม)

ตารางที่ 1 - เปรียบเทียบการจำแนกหมวดของหลักการจริยธรรมปัญญาประดิษฐ์โดย Fjeld et al. 2020 และ Mbiazi et al. 2023 โดยในแถวเดียวกันคือหมวดจากทั้งสองบทความที่พอจะเทียบเคียงกันได้

สำหรับประเทศไทยมีแนวปฏิบัติด้านจริยธรรมปัญญาประดิษฐ์เท่าที่สำรวจพบเบื้องต้น (อาทิตย์ สุริยะวงศ์กุล 2567) ดังนี้

- **แนวปฏิบัติจริยธรรมปัญญาประดิษฐ์ (Thailand AI Ethics Guideline)** จัดทำโดย สำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ โดยมีคณะผู้จัดทำเป็นนักวิชาการจากคณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล มีบริษัท ไมโครซอฟท์ (ประเทศไทย) จำกัด ร่วมเป็นที่ปรึกษา เผยแพร่ฉบับแรกเมื่อ 21 ตุลาคม 2562 ต่อมาได้ปรับปรุงเพิ่มกรณีศึกษาและวิธีการปฏิบัติ และกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมเสนอคณะรัฐมนตรีเมื่อ 14 ธันวาคม 2563 ให้หน่วยงานราชการใช้เป็นแนวปฏิบัติ และคณะรัฐมนตรีมีมติเห็นชอบเมื่อ 2 กุมภาพันธ์ 2564 (หนังสือสำนักเลขาธิการคณะรัฐมนตรีที่ นร 0505/ว 74 ลงวันที่ 4 กุมภาพันธ์ 2564)
- **แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ของสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (NSTDA AI Ethics Guideline)** จัดทำโดย สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชา (สวทช.) โดย สวทช. ได้ออก “ประกาศสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ เรื่อง แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์” เมื่อ 31 มีนาคม 2565 ให้ใช้แนวปฏิบัตินี้กำกับดูแลงานด้านปัญญาประดิษฐ์ “ที่ดำเนินการโดยบุคลากรของสำนักงาน [สวทช.] ผู้ที่ร่วมวิจัยหรือรับการสนับสนุนการวิจัยจากสำนักงาน ภาคเอกชนที่ใช้พื้นที่ภายในอุทยานวิทยาศาสตร์ประเทศไทย และอาคารอื่น ๆ ของสำนักงาน และผู้รับจ้างช่วงที่เกี่ยวข้อง”

- **แนวปฏิบัติเกี่ยวกับมาตรฐานการใช้ปัญญาประดิษฐ์ (Thailand Artificial Intelligence Guidelines 1.0 – TAIG 1.0)** จัดทำโดย ศูนย์วิจัยกฎหมายและการพัฒนา คณะนิติศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยการสนับสนุนของ บริษัท เอพี (ไทยแลนด์) จำกัด (มหาชน), บริษัท แชนด์เล่อร์ เอ็มเอชเอ็ม จำกัด, บริษัท แอ็ดวานซ์อินฟอร์เมชั่นเทคโนโลยี จำกัด (มหาชน), และ บริษัท เทิร์นคีย์ คอมมูนิเคชั่น เซอร์วิส จำกัด (มหาชน) ตีพิมพ์ ธันวาคม 2565; เปิดตัว 4 มีนาคม 2566 ในงานเปิดตัวโครงการพัฒนาแพลตฟอร์มภาครัฐ เพื่อรองรับการปฏิบัติตามกฎหมายคุ้มครองข้อมูลส่วนบุคคล

นอกจากนี้ยังอาจมีส่วนของบริษัทเอกชน เช่นการประกาศของบริษัท ทรู คอร์ปอเรชั่น จำกัด (มหาชน) เมื่อ 24 กันยายน 2567 ถึงการใช้แผน **GSMA Responsible AI Maturity Roadmap** ของสมาคมจีเอสเอ็ม ในกิจการของตัวเอง โดยเป็นหนึ่งใน 19 ผู้ให้บริการจากทั่วโลกที่นำร่องใช้แผนดำเนินงานดังกล่าว (ทรู คอร์ปอเรชั่น 2567; GSMA Press Office 2024) สำหรับการสำรวจกฎหมายและกฎระเบียบที่เกี่ยวข้องของไทยเพิ่มเติม สามารถดูได้ที่ (Piyatumrong 2024) และ (คณะกรรมการจริยธรรมปัญญาประดิษฐ์ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ 2564:57-59)

เพื่อให้เห็นเป็นตัวอย่าง บทความส่วนนี้ได้เลือกหลักการด้านจริยธรรมและการกำกับปัญญาประดิษฐ์จากหน่วยงาน 4 ประเภทมาเปรียบเทียบโดยคร่าว ได้แก่ หน่วยงานวิจัย (สวทช.) หน่วยงานรัฐบาลที่มีอำนาจหน้าที่ในการกำกับ (กระทรวงดิจิทัลฯ) หน่วยงานระหว่างประเทศ (อาเซียน) และหน่วยงานด้านเทคนิค (มูลนิธิลีนุกซ์ - Linux Foundation) (คณะกรรมการจริยธรรมปัญญาประดิษฐ์ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ 2564; กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม 2562; Association of Southeast Asian Nations 2024; Roy 2021) ดังแสดงในตารางที่ 2

หน่วยงานวิจัย – แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

1. ความเป็นส่วนตัว (Privacy)
2. ความมั่นคงและปลอดภัย (Security and Safety)
3. ความไว้วางใจ (Reliability)
4. ความเป็นธรรม เท่าเทียม และไม่แบ่งแยก (Fairness and non-discrimination)
5. ความโปร่งใสและอธิบายได้ (Transparency and Explainability)
6. ภาระความรับผิด (Accountability)
7. มนุษย์เป็นผู้ควบคุมปัญญาประดิษฐ์ เพื่อความยั่งยืนของมนุษยชาติ (Human Oversight and Human Agency)

หน่วยงานรัฐบาล – หลักการทางจริยธรรมปัญญาประดิษฐ์ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม

1. ความสามารถในการแข่งขันและการพัฒนาอย่างยั่งยืน (Competitiveness and Sustainability Development)
2. ความสอดคล้องกับกฎหมาย จริยธรรม และมาตรฐานสากล (Laws, Ethics, and International Standards)
3. ความโปร่งใสและภาระความรับผิดชอบ (Transparency and Accountability)
4. ความมั่นคงปลอดภัยและความเป็นส่วนตัว (Security and Privacy)

5. ความเท่าเทียม หลากหลาย ครอบคลุม และเป็นธรรม (Fairness)
6. ความน่าเชื่อถือ (Reliability)

หน่วยงานระหว่างประเทศ – ASEAN Guide on AI Governance and Ethics

1. Transparency and Explainability (ความโปร่งใสและอธิธายได้)
2. Fairness and Equity (ความเป็นธรรมและเท่าเทียม)
3. Security and Safety (ความมั่นคงและปลอดภัย)
4. Human-centricity (ความมีมนุษย์เป็นศูนย์กลาง)
5. Privacy and Data Governance (ความเป็นส่วนตัวและการกำกับดูแลข้อมูล)
6. Accountability and Integrity (ความรับผิดได้และซื่อสัตย์)
7. Robustness and Reliability (ความแข็งแกร่งและน่าเชื่อถือ)

หน่วยงานด้านเทคนิค – LF AI & Data's Principles for Trusted AI

1. Reproducibility (ความสามารถในการทำซ้ำ)
2. Robustness (ความแข็งแกร่ง)
3. Equitability (ความเท่าเทียม)
4. Privacy (ความเป็นส่วนตัว)
5. Explainability (ความอธิบายได้)
6. Accountability (ความรับผิดได้)
7. Transparency (ความโปร่งใส)
8. Security (ความมั่นคง)

ตารางที่ 2 หลักการด้านจริยธรรมและการกำกับปัญญาประดิษฐ์จากหน่วยงาน 4 ประเภท ภาษาอังกฤษในวงเล็บท้ายชื่อหลักการใช้ตามที่ปรากฏในเอกสารต้นฉบับ ซึ่งอาจกินความครอบคลุมมากหรือน้อยกว่าหลักการในภาษาไทย ส่วนภาษาไทยในวงเล็บเป็นการแปลโดยผู้แต่ง

จะเห็นได้ว่าในภาพรวม หลักการจากหน่วยงานทั้ง 4 ประเภทในตัวอย่างนั้นมีความคล้ายคลึงกันเป็นอย่างมาก โดยหลักการเรื่องความโปร่งใส (transparency) ความเป็นธรรม (fairness-equitability) ความรับผิดได้ (accountability) ความมั่นคงและปลอดภัย (security-safety) ความแข็งแกร่ง-เชื่อใจได้ (robustness-reliability) และความเป็นส่วนตัว (privacy) เป็นหลักการที่พบได้อย่างสม่ำเสมอจากทุกหน่วยงาน ซึ่งสอดคล้องกับที่ Zhou et al. (2020) พบว่าหลักการด้านจริยธรรมปัญญาประดิษฐ์ที่พบมากที่สุดได้แก่หลักการด้านความโปร่งใส (transparency) ความยุติธรรม-เป็นธรรม (justice and fairness) ความรับผิดชอบ (responsibility) การไม่ก่อให้เกิดอันตราย (non-maleficence) และความเป็นส่วนตัว (privacy)

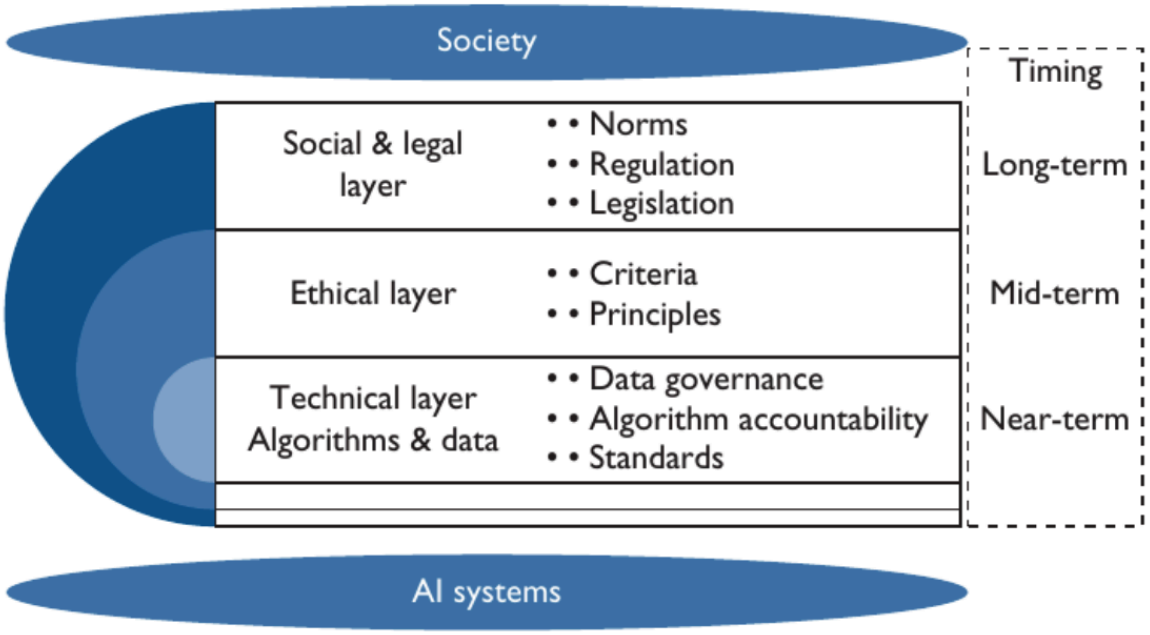
อย่างไรก็ตามหลักการจากแต่ละหน่วยงานก็อาจมีลักษณะพิเศษหรือจุดเน้นตามพันธกิจของหน่วยงาน เช่นกรณีของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม มีข้อที่ว่าด้วยความสามารถในการแข่งขัน ซึ่งเกี่ยวข้องกับพันธกิจของหน่วยงานโดยตรง หรือกรณีของหน่วยงานไทยทั้งสองแห่งที่มีประเด็นว่าด้วยความยั่งยืน (แม้ในกรณีของสวทช.คำแปลภาษาอังกฤษจะไม่มีคำดังกล่าว

ก็ตาม) หรือกรณีของสวทช.และอาเซียน ที่มีจุดร่วมกันในเรื่องการให้ความสำคัญกับมนุษย์ในฐานะผู้ควบคุมระบบหรือเป็นศูนย์กลางของการออกแบบ (ซึ่งสอดคล้องกับหลักการ Human Control of Technology และ Promotion of Human Values ใน Fjeld et al. 2020) ส่วนมูลนิธิลีนุกซ์ซึ่งเป็นหน่วยงานทางเทคนิคก็มีข้อที่ว่าด้วยความสามารถในการทำซ้ำ ซึ่งว่าด้วยระเบียบวิธีวิทยาศาสตร์ที่หากใช้ซอฟต์แวร์ ข้อมูล และสภาพแวดล้อมในลักษณะเดียวกันดังที่ได้อธิบายไว้ในเอกสาร ก็ควรจะได้ผลลัพธ์ลักษณะเดียวกันหรือคล้ายคลึงกัน ไม่ว่าใครจะเป็นผู้กระทำก็ตาม (Gundersen 2023) ซึ่งสิ่งนี้ทำให้สามารถตรวจสอบและคาดเดาพฤติกรรมของระบบได้

การกำกับปัญญาประดิษฐ์และมาตรฐานทางเทคนิค

แกสเซอร์ และ อัลไมดา (Gasser & Almeida 2017) ได้เสนอวิธีทำความเข้าใจถึงความสัมพันธ์ระหว่างเครื่องมือต่างๆ ในการกำกับกิจการ ด้วยโมเดลการกำกับกิจการแบบแบ่งชั้น (layered governance model) ซึ่งตั้งอยู่ระหว่างสังคมและระบบปัญญาประดิษฐ์ ชั้นต่างๆ เหล่านี้ ซึ่งประกอบด้วย ชั้นทางสังคม-กฎหมาย (Social & legal layer) ชั้นทางจริยธรรม (Ethical layer) และชั้นทางเทคนิค (Technical layer) ต่างก็มีปฏิสัมพันธ์กัน ทั้งนี้ชั้นทางเทคนิคเป็นชั้นที่อยู่ใกล้กับระบบปัญญาประดิษฐ์ที่สุด และชั้นทางสังคม-กฎหมายเป็นชั้นที่อยู่ใกล้กับสังคมที่สุด โดยมีชั้นทางจริยธรรมตรงกลางเป็นตัวเชื่อมประสาน

ชั้นทางเทคนิคนั้นอาจมีทั้งเครื่องมือในแง่การดำเนินงานซึ่งยังจำเป็นต้องใช้มนุษย์ในกำกับ เช่น การกำกับดูแลข้อมูลและมาตรฐาน และเครื่องมืออัตโนมัติที่สามารถใช้ซอฟต์แวร์ช่วยตรวจสอบได้ เช่น การตรวจสอบความผิดปกติในอัลกอริทึม โดย แกสเซอร์ และ อัลไมดา มองว่าชั้นทางเทคนิดนี้ เป็นดังรากฐานที่จะสนับสนุนการกำกับในชั้นทางจริยธรรมและชั้นทางสังคม-กฎหมาย



ภาพที่ 1 ตัวแบบการกำกับปัญญาประดิษฐ์แบบแบ่งชั้น

(ที่มา: แกสเซอร์ และ อัลไมดา 2017)

เราอาจมองอีกแบบได้ว่า ในขณะที่การกำกับดูแลทางเทคนิคทำงานโดยตรงกับระบบทางเทคนิคและการกำกับดูแลทางสังคม-กฎหมายทำงานกับสถาบันทางสังคม ชั้นทางจริยธรรมก็ทำหน้าที่เป็นหลักการในภาพกว้าง (เช่น หลักสิทธิมนุษยชน) เป็นดังเข็มทิศสำหรับอีกสองชั้น ในการตรวจสอบว่าการกำกับทั้งทางเทคนิคและทางสังคม-กฎหมายนั้น ยังอยู่ในเส้นทางของหลักการใหญ่หรือไม่

ในแง่ของเวลา แกสเซอร์ และ อัลไมดา ยังเสนออีกด้วยว่า ในขณะที่เครื่องมืออย่างกฎหมาย หลักการทางจริยธรรม และปทัสถานทางสังคม ของสังคมหนึ่งอาจใช้เวลาตั้งแต่ระยะปานกลางไปจนถึงระยะยาว จึงจะสามารถตกลงกันได้ แต่การพัฒนาเครื่องมือในชั้นทางเทคนิค อย่างซอฟต์แวร์ มาตรฐานทางเทคนิค และแนวปฏิบัติภายในองค์กร อาจใช้เวลาน้อยกว่า (ดูด้านขวาของภาพที่ 1) และสามารถเป็นรากฐานสำหรับการพัฒนาหรือบังคับใช้เครื่องมือกำกับในชั้นอื่นต่อไป

อย่างไรก็ตาม นับจากตอนที่ทั้งสองได้เสนอบทความดังกล่าวใน พ.ศ. 2560 เราจะพบว่าในปัจจุบัน สำหรับชั้นทางจริยธรรม เรามีหลักการด้านจริยธรรมปัญญาประดิษฐ์แล้วจำนวนมาก ดังได้แสดงบางส่วนข้างต้น[1] ส่วนในชั้นทางสังคม-กฎหมาย ก็มีกฎหมายปัญญาประดิษฐ์โดยเฉพาะอย่างน้อยหนึ่งฉบับในสหภาพยุโรป แต่ในชั้นทางเทคนิคนั้น แม้จะมีซอฟต์แวร์และมาตรฐานที่เกี่ยวข้องออกมาแล้วจำนวนหนึ่ง[2] แต่ความรับรู้ก็ยังเป็นวงจำกัด และขาดแนวทางการนำมาใช้อย่างชัดเจนนอกแวดวงทางเทคนิค เช่นการเชื่อมโยงมาตรฐานต่างๆ ดังกล่าวเข้ากับการบังคับใช้กฎหมาย

ตัวอย่างการกำกับกิจการและมาตรฐานทางเทคนิค

หากจะยกตัวอย่างจากการกำกับดูแลที่มีอยู่แล้วของไทยให้เห็นความสัมพันธ์ระหว่างหลักการทางจริยธรรม กฎหมาย และมาตรฐานทางเทคนิค ก็เช่นในเรื่องการนำเข้าข้อมูลคอมพิวเตอร์อันเป็นเท็จในประการที่น่าจะเกิดความเสียหายต่อความปลอดภัยสาธารณะ ซึ่งเราอาจมองได้ว่าตรงกับหลักการจริยธรรมว่าด้วยเรื่องความมั่นคงและปลอดภัย ทั้งนี้สำหรับเครื่องมือการกำกับในชั้นทางสังคม-กฎหมาย เพื่อที่จะให้สามารถติดตามผู้กระทำความผิดได้ พ.ร.บ.ว่าด้วยการกระทำความผิดเกี่ยวกับคอมพิวเตอร์ จึงได้ระบุในมาตรา 26 ให้ผู้ให้บริการมีหน้าที่เก็บรักษาข้อมูลจราจรทางคอมพิวเตอร์ (log file) ไว้เป็นจำนวนวันตามที่กำหนด

อย่างไรก็ตามมาตรา 26 ไม่ได้กำหนดรายละเอียดของสิ่งที่ต้องการให้จัดเก็บ หรือวิธีในการจัดเก็บที่จะถูกนับว่าน่าเชื่อถือและนำมาใช้เป็นพยานหลักฐานได้เอาไว้ รายละเอียดดังกล่าวถูกประกาศเพิ่มเติมภายหลังในประกาศของกระทรวง[3] โดยได้ระบุสิ่ง

[1] แม้จะยังถกเถียงได้ว่า เป็นที่ยอมรับร่วมกันของสังคมหนึ่งๆ มากน้อยแค่ไหน

[2] เช่นนับถึงวันที่ 5 ต.ค. 2567 คณะกรรมการ ISO/IEC JTC 1/SC 42 Artificial Intelligence ซึ่งตั้งขึ้นเมื่อปี 2560 ได้เผยแพร่มาตรฐานด้านปัญญาประดิษฐ์ออกมาแล้ว 31 ฉบับ และมีอีก 36 ฉบับที่อยู่ระหว่างการพัฒนา https://www.iso.org/committee/6794475.html

[3] ประกาศกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เรื่องหลักเกณฑ์การเก็บข้อมูลจราจรทางคอมพิวเตอร์ของผู้ให้บริการ พ.ศ. 2564 ซึ่งมาแทนที่ประกาศกระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร เรื่อง หลักเกณฑ์การเก็บรักษาข้อมูลจราจรทางคอมพิวเตอร์ของผู้ให้บริการ พ.ศ. 2550

ที่กฎหมายคาดหวังว่าจะมีในตัวข้อมูลจราจร ซึ่งจะทำให้สามารถใช้ข้อมูลจราจรเพื่อติดตามผู้กระทำผิดได้จริง เช่น ข้อ 9 ที่กำหนดว่าข้อมูลจราจรจะต้องระบุรายละเอียดผู้ใช้บริการเป็นรายบุคคลได้ หรือข้อ 11 ที่กำหนดให้ผู้ให้บริการจะต้องตั้งนาฬิกาของอุปกรณ์บริการให้ตรงกับเวลาอ้างอิงสากล และข้อมูลจราจรต้องมีส่วนประกอบตามมาตรฐานในภาคผนวกท้ายประกาศ ข้อกำหนดเหล่านี้เป็นข้อกำหนดทางเทคนิคทั้งสิ้น นอกจากนี้ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติยังได้ออกเอกสารอีก 2 ชุด คือ มาตรฐานศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ *ระบบเก็บรักษาข้อมูลจราจรทางคอมพิวเตอร์ เล่ม 1 ข้อกำหนด (มศอ. 4003.1-2552)* และ *ระบบเก็บรักษาข้อมูลจราจรทางคอมพิวเตอร์ เล่ม 2 แนวทางในการจัดทำและตรวจสอบ (ศอ. 4003.2 - 2560)* เพื่อเป็นแนวปฏิบัติในการจัดเก็บข้อมูลจราจรและทำให้มั่นใจได้ว่าตัวข้อมูลจราจรนั้นมีความน่าเชื่อถือ ซึ่งสามารถดูตัวอย่างได้ในภาพที่ 2

๔.๑.๔.๒ ข้อกำหนด: ระบบต้องสามารถปรับตั้งนาฬิกาภายใน ให้ตรงกับเวลาอ้างอิงมาตรฐานระดับชาติ ได้โดยอัตโนมัติ และมีการกำหนดความถี่ในการปรับตั้งค่าอัตโนมัติ โดยพิจารณาจากข้อมูลแวดล้อมที่เกี่ยวข้อง อาทิ ความเสถียรของระบบ

คำอธิบาย/วัตถุประสงค์: เพื่อให้ระบบเก็บรักษาข้อมูลจราจรทางคอมพิวเตอร์ มีเวลาที่ตรงกับมาตรฐานสากลและสามารถใช้อ้างอิงในการวิเคราะห์เหตุการณ์ต่างได้ถูกต้อง

หมายเหตุ	รายชื่อหน่วยงานและเครื่องแม่ข่ายที่ให้บริการปรับเทียบเวลาอ้างอิงมาตรฐานระดับชาติ ได้แก่ สถาบันมาตรวิทยาแห่งชาติ ได้แก่ time1.nimt.or.th (203.185.69.60) time2.nimt.or.th (203.185.69.59) time3.nimt.or.th (203.185.69.56) กรมอุทกศาสตร์ กองทัพเรือ ได้แก่ time.navy.mi.th (113.53.247.3) ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ได้แก่ clock.nectec.or.th (203.185.57.115)

ตัวอย่างการดำเนินการ

- **ระบบใช้โพรโทคอล NTP (Network Time Protocol) ในการปรับตั้งค่ากับเครื่องแม่ข่ายที่ให้บริการปรับเทียบเวลาอ้างอิงมาตรฐานระดับชาติ**

ภาพที่ 2 ตัวอย่างข้อกำหนด คำอธิบาย/วัตถุประสงค์ และตัวอย่างการดำเนิการ จาก ศอ. 4003.2 - 2560 ซึ่งสรุปข้อกำหนดจาก มศอ. 4003.1-2552 และจัดเป็นหมวดหมู่ตามมาตรฐาน ISO/IEC 27002 Information security controls

เนื้อหาของหลักเกณฑ์ในปรากฏในประกาศและมาตรฐานข้างต้นเป็นเครื่องมือที่อยู่ในชั้นทางเทคนิค ซึ่งผู้ให้บริการแต่ละรายสามารถสมัครใจใช้ได้เองทันที กล่าวคือมันสามารถมีสภาพเป็นเครื่องมือในการกำกับได้โดยตัวมันเองในระดับหนึ่ง แม้ในประเทศนั้นจะยังไม่มีกฎหมายว่าด้วยการกระทำผิดทางคอมพิวเตอร์เป็นการเฉพาะ (เช่นเมื่อสามารถระบุตัวผู้เผยแพร่ข้อมูลได้โดยอาศัยข้อมูลจากมาตรฐานดังกล่าว ก็อาจใช้กลไกกฎหมายว่าด้วยความปลอดภัยสาธารณะ กฎหมายคุ้มครองผู้บริโภค หรือกฎหมายอื่นๆ เท่าที่มีอยู่ได้) และเมื่อมีกฎหมายเป็นการเฉพาะ ตัวมาตรฐานทางเทคนิคเหล่านั้นก็สามารถเป็นเครื่องมือสนับสนุนในการบังคับใช้กฎระเบียบในชั้นทางสังคม-กฎหมายได้เช่นกัน

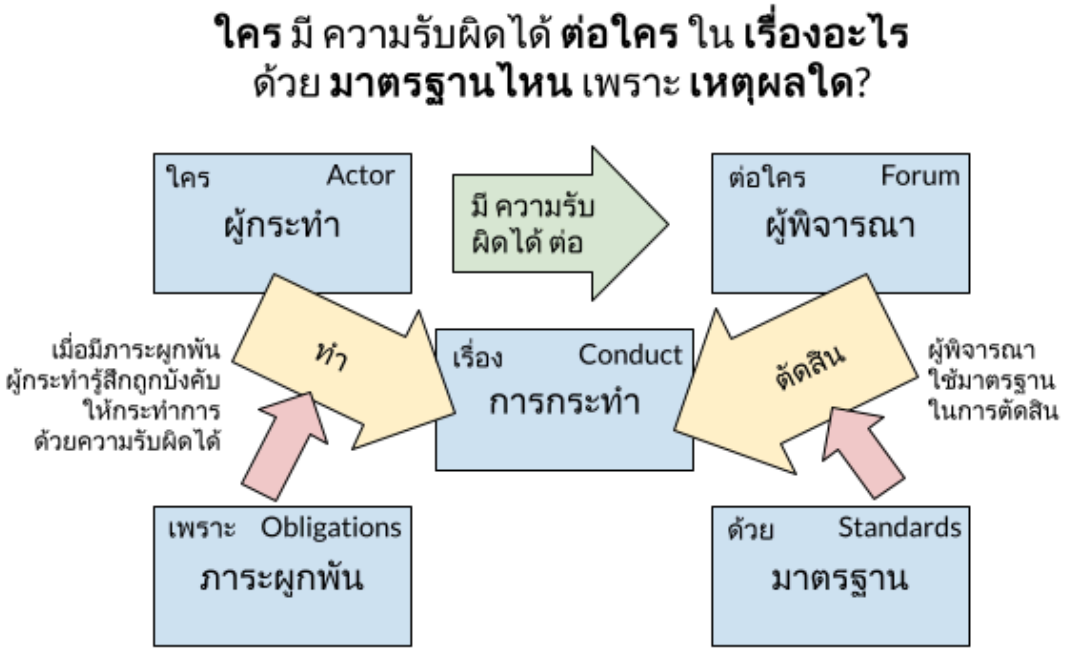
ในสองส่วนถัดไปจะเป็นการพิจารณาหลักการว่าด้วยความรับผิดได้ ซึ่งจำเป็นต้องอาศัยการบันทึกเอกสารบางอย่างเพื่อการบังคับใช้ และการพิจารณาเครื่องมือทางเทคนิค เพื่อวัตถุประสงค์ในการบังคับให้เกิดความรับผิดได้ของระบบปัญญาประดิษฐ์

หลักการว่าด้วยความรับผิดได้ (และความโปร่งใส)

หลักการที่เราให้ความสนใจในบทความส่วนนี้จะว่าด้วยเรื่องความรับผิดได้ (accountability) ซึ่งเรียกได้ว่าเป็น "หลักการข้อสุดท้าย" ในความหมายว่า ไม่ว่าจะมีหลักการกี่ข้อก็ตาม การที่จะทำให้เกิดการทำตามหลักการต่างๆ เหล่านั้นได้จริง จำเป็นจะต้องทำให้เกิดระบบนิเวศที่ผู้เล่นแต่ละรายรับผิดชอบต่อการกระทำ(หรือไม่กระทำ)ของตัวเองอย่างเหมาะสมและได้สัดส่วน ทั้งนี้หลักการว่าด้วยความรับผิดได้นั้นสัมพันธ์อย่างใกล้ชิดกับหลักการว่าด้วยความโปร่งใส (transparency) ซึ่งเราจะอภิปรายไปคู่กัน โดยอิงหน้าที่ของผู้มีส่วนได้ส่วนเสียในการรายงานข้อมูลตามที่กฎหมายปัญญาประดิษฐ์ของสหภาพยุโรป (EU AI Act) ได้เสนอไว้สำหรับระบบปัญญาประดิษฐ์ที่มีความเสี่ยงสูง (high-risk AI systems) (European Parliament and Council of the European Union 2024)

สำหรับการพิจารณาระบบปัญญาประดิษฐ์เพื่อประเมินระดับความเสี่ยงปัญญาประดิษฐ์ ซึ่งจะนำไปสู่ข้อเสนอมาตรการที่เหมาะสมกับระดับความเสี่ยงและผู้มีส่วนได้ส่วนเสีย พีรพัฒ โชคสุวัฒนสกุล, ปิยะบุตร บุญอร่ามเรือง, พัฒนาพร โกวพัฒนกิจ, ชวิน อุ่นภัทร, ฐิติรัตน์ ทิพย์สัมฤทธิ์กุล, และ เยาวลักษณ์ ชาติบัญชาชัย (2565:45-76) ได้เสนอขั้นตอนการพิจารณาที่พัฒนาขึ้นโดยอ้างอิงคำจำกัดความจากกรอบการจำแนกระบบปัญญาประดิษฐ์ของ OECD (OECD Framework for the Classification of AI systems) (Organisation for Economic Cooperation and Development 2022), กรอบการจัดการความเสี่ยงด้านปัญญาประดิษฐ์ (AI Risk Management Framework) ของ National Institute of Standards and Technology 2023), และเอกสารของผู้พัฒนาระบบปัญญาประดิษฐ์

ในที่นี้เราจะใช้กรอบคิดของความรับผิดได้ซึ่ง โบเวนส์ กูดิน และชลีมันส์ (Bovens, Goodin, & Schillemans, 2014) ได้ทบทวนวรรณกรรมด้านความรับผิดได้ทางสาธารณะและสรุปว่าความรับผิดได้นั้นมีโครงสร้างหลักที่ตั้งอยู่บนคำถามที่ว่า "ใครมีความรับผิดได้ต่อใคร ในเรื่องอะไร ด้วยมาตรฐานไหน เพราะเหตุผลใด?" (Who is accountable to whom for what by which standards and why?) คำถามดังกล่าวทำให้เราเห็นความสัมพันธ์ของมาตรฐานในบริบทของความรับผิดได้ ดังแสดงในแผนผังในภาพที่ 3



ภาพที่ 3 "ใครมีความรับผิดได้ต่อใคร ในเรื่องอะไร ด้วยมาตรฐานไหน เพราะเหตุผลใด" (ที่มา: ผู้แต่ง อ้างอิงจากคำถามโดย โบเวนส์ กูดิน และชลีมันส์ 2014)

จากแผนผัง "ผู้พิจารณา" (Forum) อาจเป็นได้ทั้งคู่ค้า ผู้บริโภค ผู้กำกับกิจการ หรือผู้มีส่วนได้ส่วนเสียอื่น แล้วแต่บริบทของการกระทำและภาระผูกพันที่กำลังพิจารณา เช่นในกรณีของกฎหมายปัญญาประดิษฐ์ของสหภาพยุโรป (EU AI Act) หาก "การกระทำ" (Conduct) เป็นการให้บริการหรือวางตลาดสินค้าที่เป็นระบบปัญญาประดิษฐ์ที่มีความเสี่ยงสูง (high-risk AI systems) มาตรา 16 ของ EU AI Act ได้กำหนดหน้าที่หรือภาระผูกพัน (obligations) ให้ "ผู้กระทำ" (Actor) ซึ่งกรณีนี้ตามนิยามของกฎหมายดังกล่าวคือ "ผู้จัดหา" (Provider) จะต้องจัดทำและเก็บรักษาเอกสารต่างๆ ตามที่ระบุไว้ในมาตรา 18 ซึ่งประกอบด้วยเอกสารเช่นเอกสารทางเทคนิค เอกสารว่าด้วยระบบบริหารคุณภาพ เป็นเวลา 10 ปีนับตั้งแต่ระบบดังกล่าวได้วางตลาด เพื่อให้หน่วยงานที่มีอำนาจตรวจสอบได้ และมาตรา 23 กำหนดให้ "ผู้นำเข้า" (Importer) มีหน้าที่ตรวจยืนยันว่าผู้จัดหามีเอกสารทางเทคนิคดังกล่าวจริง ซึ่งรายละเอียดหรือมาตรฐานของรายการข้อมูลที่จะต้องบันทึกไว้ในเอกสารเหล่านี้ได้ถูกกำหนดไว้ในภาคผนวกท้ายกฎหมาย

ตัวอย่างของมาตรฐานรายการข้อมูล เช่น รายการของข้อมูลที่จะต้องมีในเอกสารทางเทคนิคที่กำหนดไว้ในภาคผนวก 9 (Annex IV) ของ EU AI Act โดยระบุถึงข้อมูลเช่น ชื่อของผู้จัดหาระบบปัญญาประดิษฐ์ (AI Provider) จุดประสงค์ตามที่ตั้งใจไว้ (intended purpose) ของระบบเมื่อตอนสร้างระบบ รุ่นของซอฟตแวร์และเฟิร์มแวร์ที่จำเป็นต้องใช้กับระบบปัญญาประดิษฐ์ รายละเอียดว่าระบบปัญญาประดิษฐ์จะทำงานร่วมกับซอฟต์แวร์หรือฮาร์ดแวร์อย่างไร รายละเอียดวิธีการพัฒนาและทดสอบ รายละเอียดของระบบบริหารความเสี่ยง วิธีใช้งานสำหรับผู้ปรับใช้ระบบปัญญาประดิษฐ์ (AI Deployer)[4] ฯลฯ

[4] ซึ่งกรณีนี้ถือเป็นสิ่งที่ผู้จัดหา (Provider) มีความรับผิดได้ต่อผู้ปรับใช้ (Deployer)

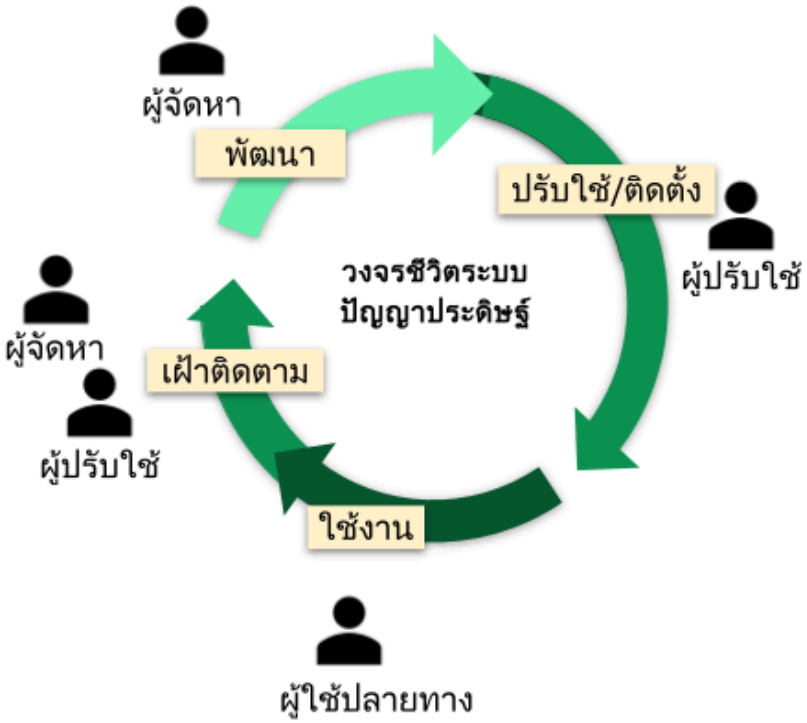
การบันทึกข้อมูลที่จำเป็นในแต่ละช่วงของวงจรชีวิตของระบบปัญญาประดิษฐ์ ตั้งแต่ก่อนเปิดใช้บริการ ระหว่างใช้บริการ และหลังใช้บริการ มีความสำคัญต่อความสามารถในการรับผิดได้ของผู้กระทำต่อผู้พิจารณา

การบันทึกข้อมูลเพื่อสนับสนุนความรับผิดได้ตลอดวงจรชีวิตของระบบปัญญาประดิษฐ์

การแบ่งช่วงวงจรชีวิตของระบบปัญญาประดิษฐ์นั้นทำได้หลายแบบ แต่ส่วนใหญ่มีลักษณะคล้ายคลึงกัน กล่าวคือในภาพใหญ่จะมีส่วนของการพัฒนา การปรับใช้และติดตั้ง การเฝ้าติดตามระหว่างใช้งาน และการนำข้อมูลจากการเฝ้าติดตามวนกลับไปปรับปรุงพัฒนาใหม่ (National Institute of Standards and Technology 2023:10; Organisation for Economic Cooperation and Development 2022; U.S. General Services Administration. n.d.) โดยในทุกช่วงของวงจรชีวิตนี้ล้วนมีโอกาสทำให้ความรับผิดได้ของระบบนั้นดีขึ้น

เอกสารของ OECD.AI แบ่งวงจรชีวิตของระบบปัญญาประดิษฐ์ออกเป็น 4 ช่วงใหญ่คือ 1) การออกแบบ ข้อมูล และตัวแบบ (Design, data and models) 2) การตรวจสอบและการตรวจยืนยัน (Verification and Validation) 3) การปรับใช้-ติดตั้ง (Deployment) และ 4) การดำเนินงานและเฝ้าติดตาม (Operation and monitoring) (Organisation for Economic Cooperation and Development n.d.) ในขณะที่มาตรฐาน ISO/IEC 5338:2023 องค์การระหว่างประเทศว่าด้วยการมาตรฐาน และคณะกรรมาธิการระหว่างประเทศว่าด้วยมาตรฐานสาขาอิเล็กทรอเทคนิกส์ แบ่งกระบวนการที่เกี่ยวกับวงจรชีวิตของระบบปัญญาประดิษฐ์ไว้อย่างละเอียดเป็น 8 กลุ่ม ได้แก่ 1) การเริ่มต้นโครงการ (Inception) 2) การออกแบบและพัฒนา (Design and Development) 3) การตรวจสอบและตรวจยืนยัน (Verification and Validation) 4) การปรับใช้/ติดตั้ง (Deployment) 5) การดำเนินงานและเฝ้าติดตาม (Operation and Monitoring) 6) การตรวจยืนยันอย่างต่อเนื่อง (Continuous Validation) 7) การประเมินซ้ำ (Re-evaluation) และ 8) การปลดประจำการ (Retirement) โดยกระบวนการที่ 6 จะทำงานร่วมกับกระบวนการที่ 5 เพื่อปรับปรุงการดำเนินงานอย่างต่อเนื่อง และกระบวนการที่ 7 จะนำไปสู่การออกแบบระบบใหม่ในกระบวนการที่ 1 และ 2 (International Organization for Standardization & International Electrotechnical Commission 2023)

สำหรับในบทความนี้เราจะใช้การแบ่งแบบง่ายตามแผนภาพในภาพที่ 4 ซึ่งเพียงพอจะทำให้เห็นภาพรวมของวงจรชีวิต รวมถึงความสัมพันธ์ของการบันทึกข้อมูลเพื่อสนับสนุนความโปร่งใสและความรับผิดได้ในแต่ละช่วง



ภาพที่ 4 วงจรชีวิตของระบบปัญญาประดิษฐ์และผู้เกี่ยวข้องบางส่วนในแต่ละช่วง

(ที่มา: AI, Data and Robotics Association-ecosystem (Adra-e))

ในแผนภาพดังกล่าวเราจะเห็นผู้มีส่วนเกี่ยวข้องหลัก 3 ประเภทคือ *ผู้จัดหา (Provider)*, *ผู้ปรับใช้ (Deployer)*, และ*ผู้ใช้ปลายทาง (End-user)* ซึ่งเป็นการแบ่งตามที่ EU AI Act ใช้ (คำภาษาไทยใช้ตาม พีรพัฒ โชคสุวัฒนสกุล และคณะ 2565)

ความสัมพันธ์ระหว่างผู้เกี่ยวข้อง 3 ประเภทนี้ อาจพอทำให้เห็นภาพได้ผ่านตัวอย่างของการผลิตระบบที่ใช้เครื่องยนต์ ระบบปัญญาประดิษฐ์ขั้นต้นอาจเปรียบได้กับเครื่องยนต์รุ่นหนึ่งซึ่งผลิตโดยผู้จัดหา ในขณะที่ผู้ปรับใช้เป็นผู้ศึกษาพฤติกรรมของผู้ใช้งานปลายทาง ความต้องการของตลาด และข้อจำกัดด้านกฎหมาย ประกอบกับข้อกำหนดทางเทคนิคของเครื่องยนต์ที่ผู้จัดหาจัดทำขึ้น แล้วตัดสินใจออกแบบรถยนต์ เรือยนต์ เครื่องจักรการเกษตร หรือเครื่องจักรอุตสาหกรรม โดยใช้เครื่องยนต์รุ่นดังกล่าว

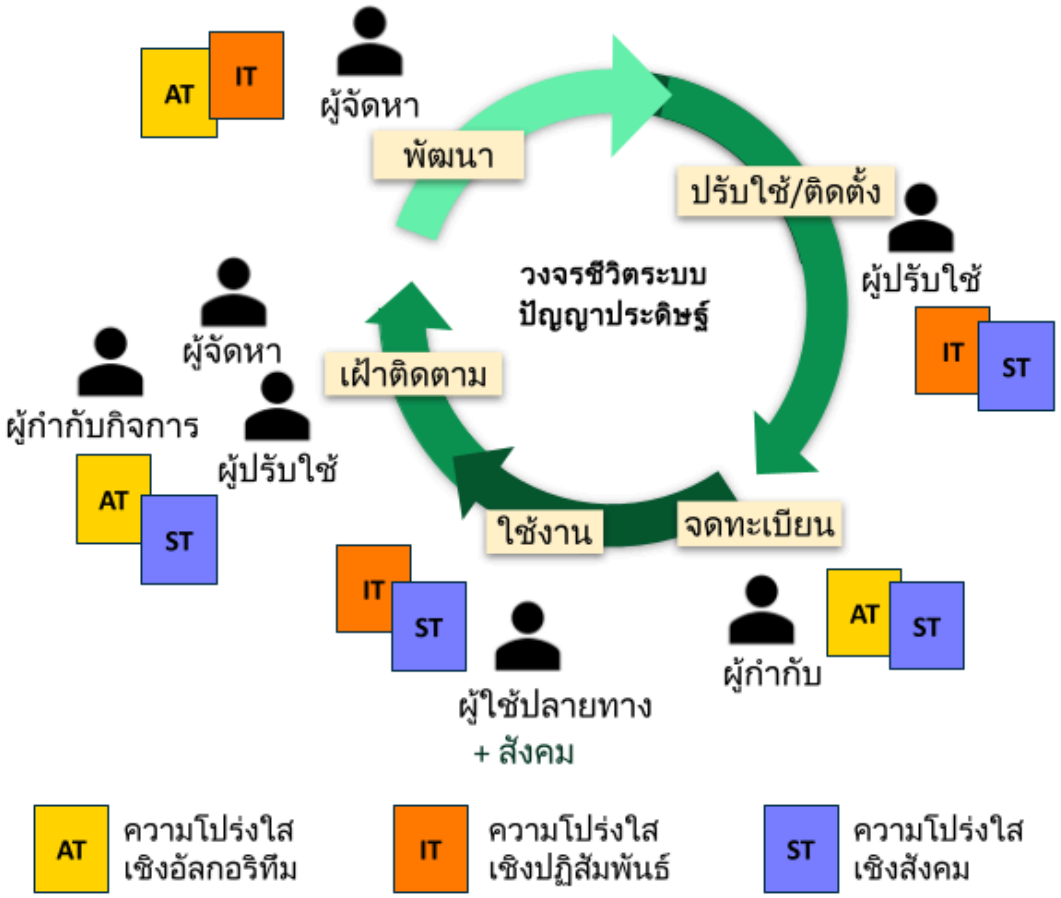
เนื่องจากผู้ใช้ปลายทางมีลักษณะแตกต่างกัน ใช้งานรถ เรือ หรือเครื่องจักรในสภาพแวดล้อมที่ต่างกัน จึงนำไปสู่การออกแบบส่วนประกอบอื่นๆ ที่ต่อเข้ากับเครื่องยนต์ที่ต่างกัน ซึ่งก็นำไปสู่ความเสี่ยงและข้อควรระวังที่แตกต่างกันด้วย ข้อมูลเกี่ยวกับเรื่องเหล่านี้เป็นส่วนที่ผู้จัดหาอาจพอมีภาพคร่าวๆ อยู่ เช่นผ่านการแลกเปลี่ยนข้อมูลกับผู้ปรับใช้ในฐานะคู่ค้า-หุ้นส่วน แต่ผู้ปรับใช้คือคนที่อยู่ใกล้ชิดผู้ใช้ปลายทางมากกว่า และมีข้อมูลของระบบที่ใช้งานจริงครอบคลุมมากกว่า เมื่อเปรียบเทียบกับผู้จัดหาที่มีข้อมูลเชิงลึกของเครื่องยนต์ แต่ก็มีเฉพาะส่วนเครื่องยนต์ ผู้ผลิตเครื่องยนต์เขียนคู่มือประกอบการใช้เครื่องยนต์ให้วิศวกรของบริษัทผู้ผลิตรถยนต์อ่าน ส่วนผู้ผลิตรถยนต์เขียนคู่มือประกอบการใช้รถยนต์ให้ผู้ขับรถทั่วไปอ่าน

สำหรับตัวอย่างที่ใกล้กับระบบปัญญาประดิษฐ์มากขึ้น เช่น ระบบรู้จำใบหน้าจากบริษัท A ซึ่งธนาคาร B นำมาปรับใช้กับระบบยืนยันตัวตนในแอปพลิเคชันธนาคารบนโทรศัพท์มือถือ ในที่นี้ A คือผู้จัดหา ส่วน B คือผู้ปรับใช้ ระบบรู้จำใบหน้าอันเดียวกันจากบริษัท A อาจถูกบริษัท C นำไปปรับใช้กับการบันทึกเวลาเข้าและออกงานได้ด้วย อย่างไรก็ตาม เนื่องจากบริษัท A ได้แจ้งไว้ในเอกสารประกอบการใช้งานว่า ชุดข้อมูลที่ใช้ในการสร้างระบบนั้นเกือบทั้งหมดเก็บตัวอย่างมาจากบุคคลในประเทศไทย ทำให้ระบบมีความแม่นยำสูงมากสำหรับการใช้งานที่ผู้ใช้ปลายทางมีลักษณะหน้าตา สีผม สีผิว รวมไปถึงเครื่องประดับบนใบหน้า ในแบบที่พบได้บ่อยในประเทศไทย แต่อาจมีความแม่นยำน้อยลงสำหรับผู้ใช้ปลายทางที่มีลักษณะหน้าตาที่ต่างออกไป สิ่งนี้ทำให้บริษัท D ที่กำลังพิจารณาเลือกระบบรู้จำใบหน้าสำหรับด่านตรวจคนเข้าเมืองที่สนามบินนานาชาติต้องทบทวนความจำเป็นในการเพิ่มชุดข้อมูลใบหน้าใหม่เข้าไปในระบบ

ผู้เกี่ยวข้องและข้อมูลที่สามารถบันทึกได้ในแต่ละช่วง อาจมีดังนี้

- **ช่วงพัฒนา**
    - *ผู้จัดหา (Provider)* ซึ่งเป็นผู้ออกแบบ พัฒนา และผลิตระบบปัญญาประดิษฐ์ขั้นต้น อยู่ในฐานะที่พร้อมที่สุด (และอาจเป็นผู้เดียวที่ทำได้) ในบรรดาผู้มีส่วนเกี่ยวข้องทั้งหมด ที่สามารถบันทึกถึงการตัดสินใจต่างๆ ที่เกี่ยวกับการออกแบบและพัฒนา เช่น สถาปัตยกรรมทางซอฟต์แวร์ที่เลือกใช้และเหตุผล ชุดข้อมูลที่ใช้ในการพัฒนาระบบและข้อจำกัดของชุดข้อมูลดังกล่าว ชุดข้อมูลที่ใช้มีข้อมูลส่วนบุคคลหรือข้อมูลอันมีลิขสิทธิ์หรือไม่ ความแม่นยำเมื่อวัดจากการทดสอบมาตรฐาน ในการสร้างระบบใช้พลังงานไปเท่าใด นอกจากนี้ ผู้จัดหายังสามารถเขียนเอกสารประกอบการใช้งาน รวมถึงแนะนำรุ่นของซอฟต์แวร์และฮาร์ดแวร์ที่เหมาะสมกับระบบปัญญาประดิษฐ์ที่ตนผลิต
- **ช่วงปรับใช้/ติดตั้ง**
    - *ผู้ปรับใช้ (Deployer)* ซึ่งนำระบบปัญญาประดิษฐ์จากผู้จัดหามาปรับใช้ประยุกต์ให้เข้ากับงาน รู้จักและอยู่ใกล้ผู้ใช้ปลายทาง อยู่ในฐานะที่เข้าใจลักษณะการใช้งานและผู้ใช้ปลายทางมากกว่าผู้จัดหา จึงสามารถถ่ายทอดข้อมูลที่ได้รับจากผู้จัดหาไปสู่ผู้ใช้ปลายทางในรูปแบบที่เหมาะสมกับผู้ใช้ปลายทางได้มากกว่าและทันกาลกว่า เช่นการแจ้งผ่านส่วนติดต่อผู้ใช้ (user interface) บนหน้าจอบนซอฟต์แวร์หรือผ่านการส่งสัญญาณอื่นบนอุปกรณ์
- **ช่วงใช้งาน-เฝ้าติดตาม**
    - *ผู้ใช้ปลายทาง (End-user)* ใช้งาน และแจ้งเหตุผิดปกติหรือให้ข้อมูลเพิ่มเติมกับผู้ปรับใช้ เพื่อให้ผู้ปรับใช้สามารถปรับปรุงระบบ หรือแจ้งต่อให้กับผู้จัดหาได้ทราบ
    - *ผู้ปรับใช้*และ*ผู้จัดหา* เฝ้าติดตามว่าระบบยังทำงานได้ตามที่คาดหวังหรือไม่ บันทึกข้อมูลการทำนายผิด-ทำนายถูกเพื่อดูว่าความแม่นยำของระบบยังอยู่ในเกณฑ์ที่รับได้อยู่หรือไม่ หรือจำเป็นต้องมีการปรับปรุงระบบให้เหมาะสมกับสภาพแวดล้อมที่เปลี่ยนไป

การบันทึกข้อมูลและจัดทำเอกสารต่างๆ เพื่อสนับสนุนความโปร่งใสและความรับผิดได้ดังกล่าว อาจวาดเป็นแผนผังได้ดังภาพที่ 5 ซึ่งมีผู้กำกับกิจการเพิ่มเข้ามาในภาพด้วย



ภาพที่ 5 วงจรชีวิตของระบบปัญญาประดิษฐ์ ความสัมพันธ์กับผู้กำกับกิจการ และชนิดของความโปร่งใสที่เกี่ยวข้องในแต่ละช่วง (ที่มา: ผู้แต่ง ดัดแปลงจาก Adra-e)

นอกจากนี้ในภาพที่ 5 มีขั้นตอนเพิ่มมาอีกขั้นตอนหนึ่งคือการ “จดทะเบียน” ซึ่งเป็นขั้นตอนที่เปิดโอกาส รวมถึงรับประกัน ให้ข้อมูลและเอกสารจากช่วงพัฒนาและช่วงปรับใช้/ติดตั้ง สามารถถูกจัดเก็บและเผยแพร่สู่ผู้เกี่ยวข้องหรือสู่สาธารณะ อีกทั้งสืบค้นได้เมื่อจำเป็น ซึ่งเป็นสิ่งที่มาตรา 49 (Registration) และมาตรา 71 (EU Database) ของ EU AI Act พยายามทำ (ดูการอภิปรายที่ท้ายส่วนนี้)

ภาพที่ 5 ยังพยายามแสดงให้เห็นว่า ในแต่ละช่วงของวงจรชีวิต ผู้ที่เกี่ยวข้องน่าจะกำลังทำงานกับข้อมูลและเอกสารประเภทใดเป็นหลัก หากพิจารณาในเชิงความโปร่งใส โดยใช้วิธีการแบ่งชนิดความโปร่งใสตามที่ Kashyap Haresamudram, Stefan Larsson & Fredrik Heintz (2023) ได้แบ่งความโปร่งใสเป็น 3 ระดับ คือความโปร่งใสเชิงอัลกอริทึม เชิงปฏิสัมพันธ์ และเชิงสังคม

	ความโปร่งใส เชิงอัลกอริทึม	**ความโปร่งใส เชิงปฏิสัมพันธ์**	**ความโปร่งใส เชิงสังคม**

ความหมาย	ความสามารถในการเข้าถึงและตรวจสอบ-ตั้งคำถามต่อโค้ด, ชุดข้อมูล, และอุปกรณ์ที่ประกอบเข้าด้วยกันเป็นระบบ AI	ความสามารถในการเข้าใจสิ่งที่ระบบ AI ทำได้ดีและสิ่งที่ทำได้จำกัด ซึ่งได้มาจากการแลกเปลี่ยนความรู้ระหว่างตัวระบบและผู้ใช้	ความสามารถทางกฎหมายและทางวัฒนธรรม ของ(สถาบันทาง)สังคมในการเข้าใจและหาหนทางตอบสนองกับการใช้งานระบบ AI
จุดแข็ง	วิธีในการอธิบาย[5] มีมาตรฐานทางเทคนิคที่ค่อนข้างชัดเจน มีแนวโน้มเป็นวัตถุวิสัย เข้าใจร่วมกันได้ระหว่างผู้เชี่ยวชาญเฉพาะเรื่อง ผู้ตรวจสอบ และผู้กำกับ	อุปลักษณ์ (metaphor) ที่จับต้องได้-ฝังอยู่ในประสบการณ์การใช้งาน ทำให้ผู้ใช้ปลายทางเข้าใจสภาพแวดล้อมและวิธีคิดในการออกแบบระบบ	วิธีการเพื่อความโปร่งใสถูกผนวกอยู่ในการทำงานของสถาบัน (เช่น ความปลอดภัยในอุตสาหกรรมอาหาร การบิน) รวมถึงองค์กรวิชาชีพ
จุดอ่อน	อาจเป็นการลำบากสำหรับผู้ที่ไม่มีภูมิหลังทางปัญญาประดิษฐ์หรือความรู้ในกิจการดังกล่าวในการจะเข้าใจข้อมูล	ความรู้หรือคำอธิบายนี้ เป็นสิ่งที่ระบบและผู้ใช้สร้างขึ้นร่วมกัน ทำให้แต่ละคนอาจมีไม่ตรงกัน	วิธีการที่เสนอข้อมูลหรือ “ทางเลือก” ให้กับผู้ใช้จนเกินรับไหว (เช่น กล่องข้อความขอความยินยอมให้เก็บคุกกี้)
ตัวอย่าง	เอกสารการออกแบบ, ค่าความน่าจะเป็น, แผนภูมิแสดงความเชื่อมโยงผลลัพธ์	การออกแบบหน้าจอ, การบันทึกข้อมูลเหตุการณ์, ศูนย์รับเรื่องร้องเรียน	ฉลากผลิตภัณฑ์, ฐานข้อมูลระบบปัญญาประดิษฐ์, ฐานข้อมูลแจ้งเหตุ

ตารางที่ 3 ความโปร่งใสแต่ละระดับ และจุดแข็ง-จุดอ่อน (ที่มา: ผู้เขียนเรียบเรียงจาก Kashyap Haresamudram, Stefan Larsson & Fredrik Heintz 2023)

ตัวอย่างชนิดของข้อมูลที่กฎหมายปัญญาประดิษฐ์ของสหภาพยุโรปกำหนดให้บันทึกในแต่ละช่วงชีวิตของระบบ และมาตรฐานทางเทคนิคที่เกี่ยวข้อง

ในบริบทความโปร่งใสและความรับผิดได้ตาม EU AI Act เอกสารหลักอันหนึ่งซึ่งจัดทำใน*ช่วงการพัฒนา*ระบบปัญญาประดิษฐ์ก็คือ “เอกสารทางเทคนิค” ตามมาตรา 11 (Technical documentation) รวมถึงข้อกำหนดตามมาตรา 13 และมาตราที่เกี่ยวข้องที่ผู้จัดหาระบบปัญญาประดิษฐ์ที่มีความเสี่ยงสูงมีภาระหน้าจะต้องจัดหาข้อมูลให้กับผู้ปรับใช้ได้ทราบ

จากนั้น เมื่อกำลังจะพ้น*ช่วงการปรับใช้* และจะวางตลาดหรือเปิดให้บริการระบบ ทั้งผู้จัดหาและผู้ปรับใช้ต้องจดทะเบียนระบบปัญญาประดิษฐ์เข้าฐานข้อมูล ตามหน้าที่ในมาตรา 49 และรายละเอียดในภาคผนวก 8 (Annex VIII) ภาคผนวกดังกล่าวแบ่งเป็น 3 ส่วน สองส่วนแรก (Section A และ Section B) เป็นข้อมูลเกี่ยวกับตัวระบบและเอกสารการทดสอบ ซึ่งเป็นหน้าที่ของผู้จัดหาในการแจ้ง โดยส่วนหนึ่งใช้ข้อมูลจากเอกสารทางเทคนิคข้างต้น และส่วนที่สาม (Section C) โดยหลักคือข้อมูลการ

[5] เช่นวิธี Local Interpretable Model-agnostic Explanations (LIME) (Marco Tulio Ribeiro, Sameer Singh & Carlos Guestrin 2016) และ SHapley Additive exPlanations (SHAP) (Scott M. Lundberg & Su-In Lee 2017)

ประเมินผลกระทบด้านสิทธิพื้นฐาน (fundamental rights impact assessment) และการประเมินผลกระทบด้านการคุ้มครองข้อมูล (data protection impact assessment) ซึ่งการประเมินทั้งสองเป็นเรื่องที่มีเฉพาะผู้ปรับใช้ที่รู้จักผู้ใช้ปลายทางและลักษณะการใช้งานที่สามารถประเมินได้

เมื่อเข้าสู่*ช่วงการใช้งาน* ข้อมูลที่สำคัญต้องเก็บคือ “บันทึกเหตุการณ์” ที่เกิดขึ้นในระบบ (log) ซึ่งมาตรา 12 (Record-keeping) กำหนดให้ระบบจะต้องบันทึกข้อมูลดังกล่าวโดยอัตโนมัติตลอดอายุการใช้งานของระบบ โดยมีจุดประสงค์เพื่อเฝ้าติดตาม(และบรรเทา)ความเสี่ยงที่อาจเกิดขึ้น รวมถึงสนับสนุนการติดตามว่าตัวระบบยังทำตามมาตรฐานที่กำหนดอย่างต่อเนื่องหลังจากได้วางตลาดแล้ว ตามมาตรา 72 (Post-market monitoring by providers and post-market monitoring plan for high-risk AI systems) นอกจากนี้ยังมีการแบ่งปันข้อมูลจากการแจ้งเหตุตามมาตรา 73 (Reporting of serious incidents) ซึ่งกำหนดหน้าที่ในการแบ่งปันข้อมูล “รายงานการเกิดเหตุ” (incident report) ร่วมกันระหว่างผู้จัดหา ผู้ปรับใช้ และผู้กำกับกิจการ เมื่อเกิดเหตุร้ายแรงกับระบบ

จะเห็นว่า เรามีเอกสารหรือข้อมูลอย่างน้อย 3 ชนิด จากตัวอย่างที่อภิปรายข้างต้น คือ เอกสารทางเทคนิค, บันทึกเหตุการณ์, และรายงานการเกิดเหตุ ซึ่งความท้าทายในการบังคับใช้กฎหมายก็คือ จะจัดการอย่างไรกับข้อมูลที่ทั้งซับซ้อนและทั้งมีปริมาณมาก ได้อย่างถูกต้องรวดเร็ว ไม่เป็นอุปสรรคต่อผู้ประกอบการ รับมือกับเหตุร้ายแรงได้ทันเวลา และดูแลความยุติธรรมให้กับผู้ได้รับความเสียหายได้โดยใช้เวลาเหมาะสม เอกสารทางเทคนิคมีแนวโน้มจะซับซ้อนสูง บันทึกเหตุการณ์สำหรับระบบปัญญาประดิษฐ์ของระบบสาธารณูปโภคจะมีจำนวนมหาศาลในแต่ละนาที รายงานการเกิดเหตุอาจเกี่ยวข้องกับซอฟต์แวร์หลายส่วน อุปกรณ์หลากชนิด และผู้จัดหา-ผู้ปรับใช้มากราย วิธีหนึ่งในความพยายามจัดการสิ่งเหล่านี้ คือมาตรฐานการจัดเก็บและแลกเปลี่ยนข้อมูล ซึ่งจะทำให้ข้อมูลจากทุกผู้จัดหาและผู้ปรับใช้ อยู่ในรูปแบบมาตรฐานเดียวกัน สามารถประมวลผลได้โดยคอมพิวเตอร์ ผู้กำกับกิจการสามารถสืบค้นหรือใช้ซอฟต์แวร์จัดระเบียบและเปรียบเทียบได้

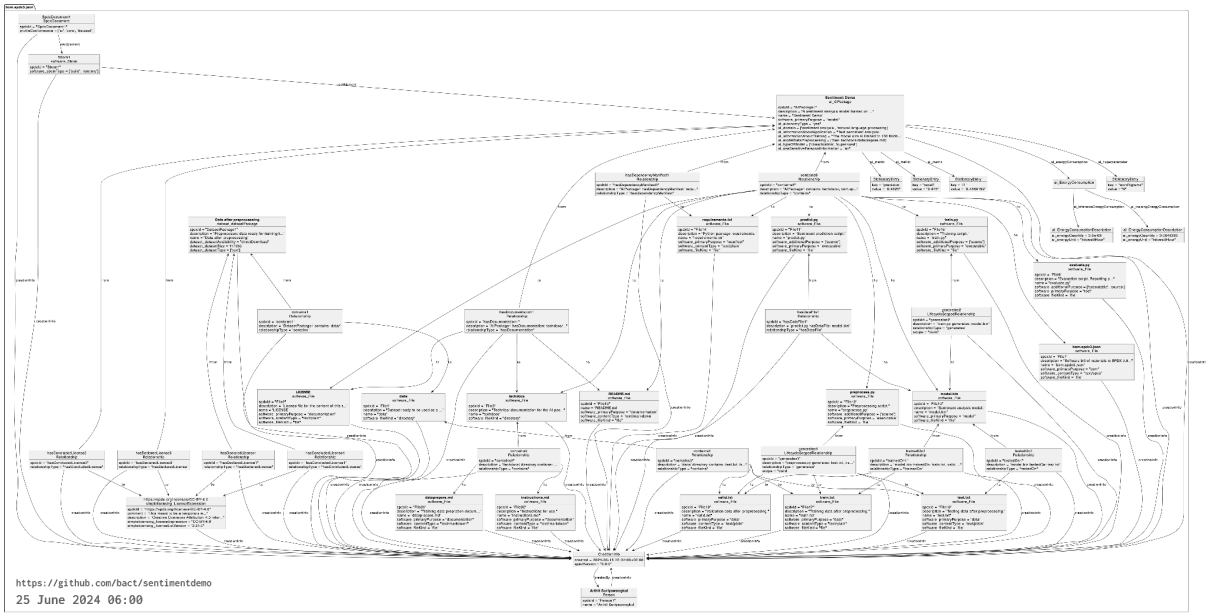
มาตรฐานอุตสาหกรรมชนิดหนึ่งที่สามารถใช้กับเอกสารทางเทคนิคของระบบปัญญาประดิษฐ์ได้ ก็คือ bill of materials (BOM) หรือในภาษาไทยเรียกในชื่อ รายการวัสดุ สูตรการผลิต หรือ โครงสร้างสินค้า ซึ่งเป็นรายการของวัตถุดิบหรือส่วนประกอบ ที่นำมารวมกันหรือประกอบกันจนเป็นผลิตภัณฑ์ เช่น รายการของชิ้นส่วนที่นำมาประกอบเป็นเครื่องยนต์หรือรถยนต์ (ซึ่งจะเกี่ยวข้องกับการตรวจสอบตามมาตรฐานความปลอดภัยของยานยนต์หรือตามมาตรฐานทางสิ่งแวดล้อม) หรือรายการส่วนประกอบของอาหาร (ซึ่งจะเกี่ยวข้องกับสิ่งที่ต้องแสดงบนฉลากในภาชนะบรรจุตามกฎหมายอาหารและยา)

ในอุตสาหกรรมซอฟต์แวร์ มี “รายการส่วนประกอบซอฟต์แวร์” (software bill of materials) หรือ SBOM ซึ่งบันทึกส่วนประกอบต่างๆ ของซอฟต์แวร์ ทั้งในส่วนของต้นรหัส (source code) ข้อมูล การตั้งค่า ฯลฯ ใช้กันมาระยะหนึ่งแล้ว (NTIA Multistakeholder Process on Software Component Transparency Framing Working Group 2021) โดยงานที่ใช้กันแพร่หลายเช่น งานทางทรัพย์สินทางปัญญา และงานทางความมั่นคงปลอดภัยของห่วงโซ่อุปทานซอฟต์แวร์ (software supply chain security) ตัวอย่างของงานทางทรัพย์สินทางปัญญาหรือลิขสิทธิ์นั้นอยู่ในบริบทที่ว่า ระบบซอฟต์แวร์หนึ่งอาจมีส่วนประกอบจำนวนมากจากหลายผู้ผลิต ซึ่งแต่ละส่วนอาจมีสัญญาอนุญาตและเงื่อนไขในการใช้งานแตก

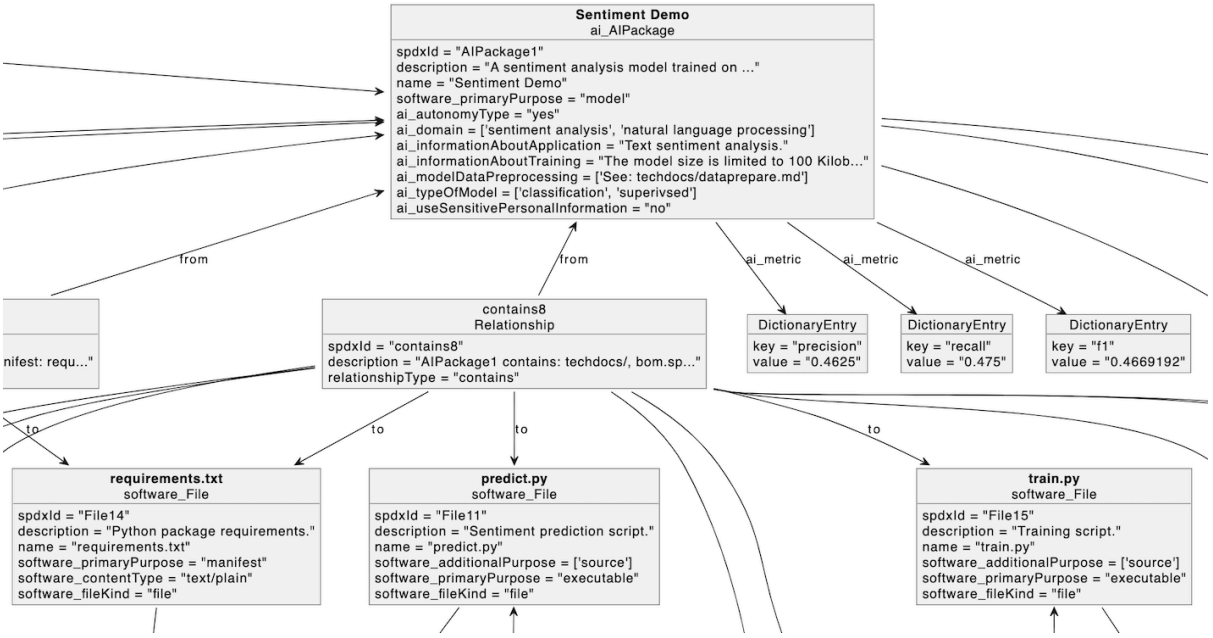
ต่างกัน การนำระบบซอฟต์แวร์ไปปรับใช้ในงานต่างๆ อาจต้องตรวจสอบเงื่อนไขการใช้งานของทุกชิ้นส่วน เพื่อหลีกเลี่ยงปัญหาทางกฎหมายในภายใน ส่วนตัวอย่างของงานด้านความมั่นคงปลอดภัยของซอฟต์แวร์ เช่น เมื่อมีการประกาศว่าซอฟต์แวร์ A รุ่น 5.4 มีรูรั่วความปลอดภัย เราจะทราบได้อย่างไรว่าระบบของเราใช้ซอฟต์แวร์รุ่นดังกล่าวหรือไม่ การมีรายการส่วนประกอบของซอฟต์แวร์ในระบบ จะช่วยตอบคำถามดังกล่าวได้

กระทรวงพาณิชย์และสำนักงานโทรคมนาคมและสารสนเทศแห่งชาติของสหรัฐอเมริกา (National Telecommunications and Information Administration) ได้รวมกันจัดทำข้อกำหนดขั้นต่ำของสิ่งที่จะต้องมีในรายการส่วนประกอบซอฟต์แวร์ (The Minimum Elements For a Software Bill of Materials) เพื่อสนับสนุนการปฏิบัติตามคำสั่งประธานาธิบดีที่ 14028 (Executive Order 14028) ว่าด้วยการปรับปรุงความมั่นคงปลอดภัยไซเบอร์แห่งชาติ (US Department of Commerce 2021) นอกจากนี้ยังมีมาตรฐานระหว่างประเทศ ISO/IEC 5962:2021 ที่กำหนดรูปแบบข้อมูลสำหรับรายการส่วนประกอบซอฟต์แวร์ (International Organization for Standardization and International Electrotechnical Commission 2021) ซึ่งมาตรฐานดังกล่าวถูกอ้างอิงโดยเอกสารเช่น แนวทางทางเทคนิค TR-03183 ซึ่งเป็นข้อกำหนดของรัฐบาลกลางเยอรมนีสำหรับผู้ผลิตและผลิตภัณฑ์ซอฟต์แวร์ (Federal Office for Information Security 2024)

ปัจจุบันมีการพัฒนาต่อยอดรายการส่วนประกอบซอฟต์แวร์ จนเป็นรายการส่วนประกอบระบบปัญญาประดิษฐ์ (AI bill of materials - AI BOM) ซึ่งเพิ่มข้อมูลเช่น ความสัมพันธ์ระหว่างชุดข้อมูลและตัวแบบปัญญาประดิษฐ์ (AI model) ความแม่นยำและวิธีการวัด อคติของข้อมูล ตัวอย่างของมาตรฐาน AI BOM เช่น CycloneDX Machine Learning Bill of Materials (ML-BOM) (OWASP Foundation n.d.) และ AI Profile ของ System Package Data Exchange (SPDX) 3.0 (The Linux Foundation and its Contributors 2024) (Karen Bennet, Gopi Krishnan Rajbahadur, Arthit Suriyawongkul & Kate Stewart 2024) โดยมาตรฐานตัวหลังกำลังอยู่ในกระบวนการประกาศเป็นรุ่นถัดไปของมาตรฐาน ISO/IEC 5962 ซึ่งคาดว่าจะออกได้ราวต้นปี 2568



ภาพที่ 6.1 แผนผังที่วาดขึ้นจากข้อมูลในรายการวัสดุของระบบปัญญาประดิษฐ์ (AI bill of materials) แสดงให้เห็นภาพรวมของความสัมพันธ์ระหว่างชิ้นส่วนต่างๆ ในระบบ ทำให้ส่วนประกอบที่อาจไม่เคยสังเกตเห็นได้ถูกเห็นชัดขึ้น แผนผังนี้วาดขึ้นโดยอัตโนมัติจากข้อมูล AI BOM ที่บันทึกด้วยมาตรฐานแลกเปลี่ยนข้อมูล SPDX 3.0 (Arthit Suriyawongkul 2024)



ภาพที่ 6.2 ขยายภาพที่ 6.1 ให้เห็นช่องข้อมูลที่ตัวรายการได้บันทึกไว้ เช่น ชนิดของตัวแบบปัญญาประดิษฐ์ มีใช้ข้อมูลส่วนบุคคลหรือไม่ และความแม่นยำ (Arthit Suriyawongkul 2024)

ในส่วนของการบันทึกเหตุการณ์ (ซึ่งมีบางส่วนคล้ายกับการเก็บข้อมูลจราจรทางคอมพิวเตอร์ที่ได้ยกตัวอย่างไปก่อนหน้านี้) มาตรฐาน ISO/IEC 42001:2023 ที่ว่าด้วยระบบบริหารจัดการปัญญาประดิษฐ์ ในข้อที่ B.6.2.8 (AI system recording of event logs) มีแนวทางเกี่ยวกับข้อมูลที่ควรบันทึก เช่น การตรวจวัดสมรรถนะของระบบปัญญาประดิษฐ์เมื่อถูกให้ทำงานนอกเหนือสภาพแวดล้อมที่ตั้งใจออกแบบมา วันเวลาที่ระบบถูกใช้งาน ผลลัพธ์ที่ตกอยู่นอกช่วงข้อมูลที่ถูกออกแบบไว้ อย่างไรก็ตามยังไม่พบว่ามีมาตรฐานรูปแบบข้อมูลที่เจาะจงสำหรับบันทึกชนิดนี้

สำหรับรายงานการเกิดเหตุ (Violet Turri & Rachel Dzombak 2023; Avinash Agarwal & Manisha J. Nene 2024) ข้อมูลประเภทหนึ่งในรายงานดังกล่าว คือข้อมูลชนิดความเสียหายหรือความเสี่ยง ซึ่งหลายหน่วยงานมีความพยายามในการกำหนดวิธีจำแนกหรืออนุกรมวิธาน (taxonomy) เช่น Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST AI 100-2e2023) จากสถาบันมาตรฐานและเทคโนโลยีแห่งชาติของสหรัฐอเมริกา (Apostol Vassilev, Alina Oprea, Alie Fordyce & Hyrum Anderson 2024), AI Risks Taxonomy จากสถาบันวิจัยเพื่อการลดอาวุธแห่งสหประชาชาติ (Ioana Puscas 2023), และ AI, Algorithmic and Automation Harms Taxonomy จากกลุ่ม AIAAIC (Gavin Abercrombie et al. 2024) ซึ่งจะช่วยให้การรายงานการเกิดเหตุมีการบันทึกข้อมูลที่สม่ำเสมอกัน อ้างอิงถึงเหตุการณ์ประเภทเดียวกันด้วยคำศัพท์เดียวกัน

บทสรุป: บันทึกข้อมูลเพื่อประโยชน์อะไร

เราจะบันทึกข้อมูลตามมาตรฐานที่ประมวลผลได้โดยง่ายด้วยคอมพิวเตอร์ ทำให้เกิดความโปร่งใส และรับประกันว่าจะมีความรับผิดได้ไปเพื่ออะไร อะไรคือประโยชน์ของการทำสิ่งเหล่านี้ Bovens (2007) เสนอว่าวัตถุประสงค์ของความรับผิดได้ ในบริบทของหน่วยงานที่ให้บริการสาธารณะหรือใช้อำนาจสาธารณะนั้น อาจแบ่งได้ 3 มุมมองเพื่อการตั้งคำถามคือ 1) มุมมองเชิงประชาธิปไตย (democratic) ซึ่งหมายความว่าประชาชนเป็นผู้ควบคุมการใช้อำนาจ 2) มุมมองเชิงรัฐธรรมนูญ (constitutional) ซึ่งว่าด้วยการป้องกันการทุจริตและการใช้อำนาจในทางที่ผิด และ 3) มุมมองเชิงการเรียนรู้ (learning) ซึ่งว่าด้วยการขยายประสิทธิผลของบริการสาธารณะ กิจกรรมใดๆ ที่จะทำไปเพื่อความรับผิดได้ รวมถึงการบันทึกข้อมูลหรือจัดทำเอกสาร ควรต้องตอบจุดประสงค์อย่างน้อยอย่างใดอย่างหนึ่งดังกล่าว

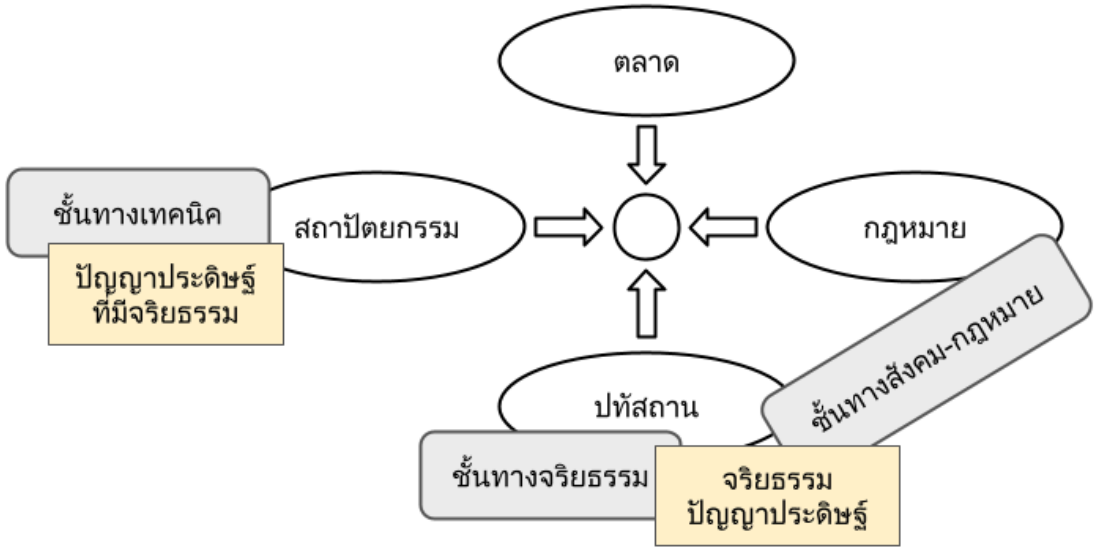
วัตถุประสงค์ของความรับผิดได้	เชิงประชาธิปไตย: ประชาชนควบคุม	เชิงรัฐธรรมนูญ: ป้องกันการลุแก่อำนาจ	เชิงการเรียนรู้: ทำงานดีขึ้น ได้ผลมากขึ้น
ตัวอย่างชนิดเอกสารและสิ่งบันทึก	เอกสารทางเทคนิคว่าด้วยการออกแบบระบบ, รายการส่วนประกอบซอฟต์แวร์ (SBOM)	บันทึกเหตุการณ์ (log)	รายงานการเกิดเหตุ
ตัวอย่างเครื่องมือสนับสนุน	วิธีการอธิบายการตัดสินใจหรือคำทำนายของระบบปัญญาประดิษฐ์, มาตรฐานรายการส่วนประกอบซอฟต์แวร์	ระบบเฝ้าติดตามอัตโนมัติ, การตรวจวัดสมรรถนะของระบบ, การตรวจจับอคติในผลลัพธ์	การวิเคราะห์ทบทวนรายงานการเกิดเหตุ เพื่อปรับปรุงการทำงานของระบบ
สิ่งที่คาดหวัง	ผู้เกี่ยวข้องเข้าใจการทำงานของระบบเพียงพอที่จะตรวจสอบมันได้	หากมีความผิดพลาดหรือความเสี่ยง ระบบสามารถพบและปรับปรุงตัวเองเพื่อลดโอกาสพลาด	การเรียนรู้ในเชิงองค์กร องค์กรไม่ทำพลาดซ้ำ รับมือกับเหตุได้ดีขึ้น

ตารางที่ 4 ความสัมพันธ์ระหว่างข้อมูลที่จัดเก็บหรือเอกสารที่จัดทำ กับวัตถุประสงค์ของความรับผิดได้

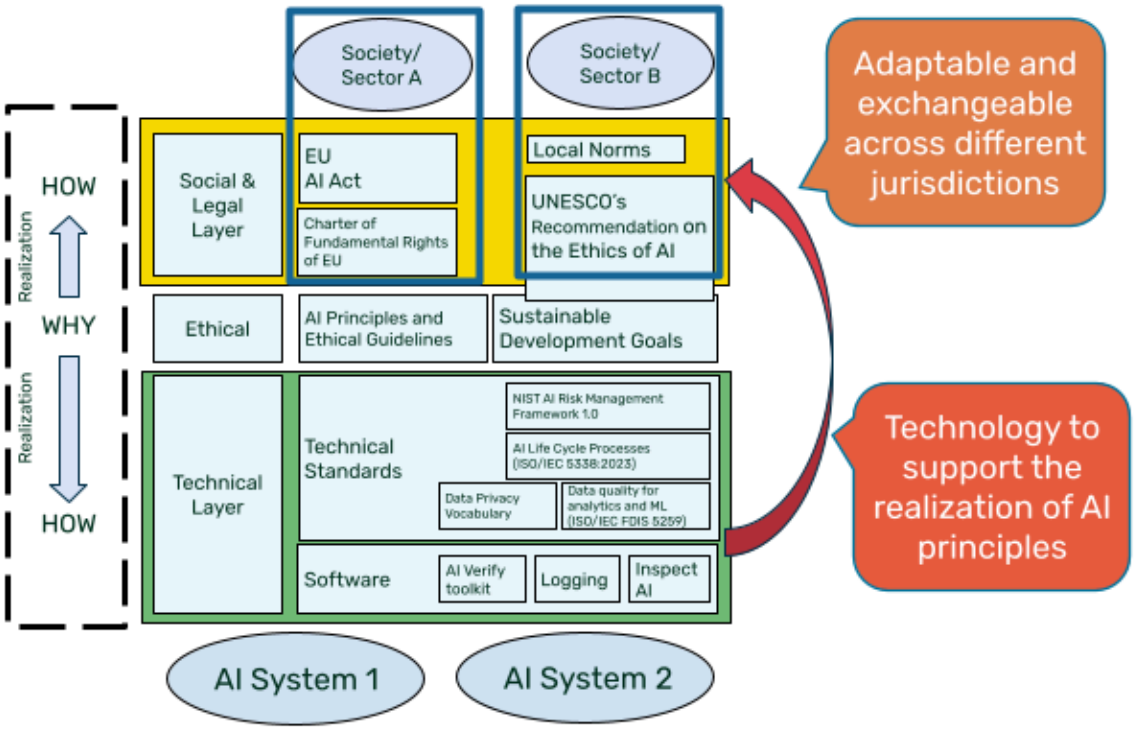
เมื่อกลับไปพิจารณาการแบ่งชั้นในการกำกับปัญญาประดิษฐ์ที่เสนอโดย *Gasser & Almeida (2017)* เทียบกับการจำแนกหลักการจริยธรรมปัญญาประดิษฐ์ของ Zhou et al. (2020) ชั้นทางเทคนิคอาจเทียบได้กับกลุ่ม “ปัญญาประดิษฐ์ที่มีจริยธรรม” และชั้นทางจริยธรรมกับชั้นทางสังคม-กฎหมายอาจเทียบได้กับกลุ่ม “จริยธรรมปัญญาประดิษฐ์” ขณะที่หากเทียบกับแผนภาพตามทฤษฎี “จุดที่น่าสงสาร” (pathetic dot) ของ Lessig (2006:120-137) ที่เสนอ “แรง” 4 ประเภทที่จะส่งผลต่อเรื่องในกำกับ (หรือตัว “จุด”) อันได้แก่ กฎหมาย (Law), ปทัสถาน (Norms), สถาปัตยกรรม (Architecture), และตลาด

(Market) จะพบว่าชั้นทางสังคม-กฎหมายอาจเทียบได้กับแรง “ปทัสถาน” และแรง “กฎหมาย” ส่วนชั้นทางเทคนิคอาจเทียบได้กับแรง “สถาปัตยกรรม” การเทียบกรอบการมองนี้ทำให้เห็นได้ว่าเครื่องมือการกำกับที่ถูกพูดถึงมากในการอภิปรายเกี่ยวกับการกำกับปัญญาประดิษฐ์ คือเครื่องมือด้านจริยธรรมและด้านกฎหมาย ขณะที่เครื่องมือทางชั้นเทคนิคนั้นถูกพูดถึงรองลงมา และเครื่องมือกำกับด้านตลาด ทั้งกลไกตลาดเองและการกำกับตลาด ถูกพูดถึงน้อยที่สุด

บทความนี้เสนอการพิจารณามาตรฐานทางเทคนิคสำหรับการบันทึกและแลกเปลี่ยนข้อมูลเพื่อสนับสนุนหลักการความรับผิดได้ (accountability) โดยยกตัวอย่างเอกสาร 3 ชนิดคือ เอกสารทางเทคนิค (technical documentation) ที่แสดงรายการส่วนประกอบในระบบปัญญาประดิษฐ์, บันทึกเหตุการณ์ (log) และรายงานการเกิดเหตุ (incident report) ซึ่งจะถูกบันทึกในช่วงที่แตกต่างกันในวงจรชีวิตของระบบปัญญาประดิษฐ์ และตอบโจทย์วัตถุประสงค์ของความรับผิดได้ที่ต่างกัน พร้อมกับตัวอย่างมาตรฐานคอมพิวเตอร์สำหรับการบันทึกและแลกเปลี่ยนข้อมูล โดยหวังว่าการพัฒนามาตรฐานที่เกี่ยวข้องที่คำนึงถึงบริบทของประเทศไทย จะเป็นส่วนหนึ่งที่ทำให้การกำกับระบบปัญญาประดิษฐ์ในประเทศไทยมีความเหมาะสม ส่งเสริมเศรษฐกิจ คุ้มครองความปลอดภัยของสาธารณะ และสนับสนุนสิทธิของผู้เกี่ยวข้องทั้งหมด



ภาพที่ 7 แผนภาพทฤษฎี “จุดที่น่าสงสาร” (pathetic dot) (Lessig 2006) เทียบกับการจำแนกหลักการปัญญาประดิษฐ์ที่มีจริยธรรม-จริยธรรมปัญญาประดิษฐ์โดย Zhou et al. (2020) และเทียบกับการจำแนกเครื่องมือในการกำกับปัญญาประดิษฐ์แบบแบ่งชั้นโดย Gasser & Almeida (2017) ทำให้เราเห็นได้ว่า ในการศึกษาด้านการกำกับปัญญาประดิษฐ์ เครื่องมือด้านตลาดยังเป็นสิ่งที่ยังได้รับความสนใจน้อยอยู่โดยเปรียบเทียบ



ภาพที่ 8 มาตรฐานทางเทคนิคในฐานะส่วนประกอบหนึ่งของชั้นทางเทคนิค ซึ่งจะสนับสนุนการบังคับใช้กฎหมายและปทัสถานในชั้งทางสังคม-กฎหมาย รวมถึงสนับสนุนการแลกเปลี่ยนข้อมูล-ร่วมมือในการกำกับข้ามเขตอำนาจศาล

บรรณานุกรม

ภาษาไทย

กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม. 2562. 'หลักการและแนวทางจริยธรรมปัญญาประดิษฐ์ของประเทศไทย'.

คณะกรรมการจริยธรรมปัญญาประดิษฐ์ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ. 2564. 'แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ของสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ'.

ทรู คอร์ปอเรชั่น. 2567. 'ทรู คอร์ปอเรชั่น นำร่องแผน Responsible AI มุ่งใช้ด้วยความรับผิดชอบเป็นรายแรกในไทย'. *True Blog*. เข้าถึง 4 ตุลาคม 2567 (https://www.true.th/blog/responsible-ai/).

พีรพัฒ โชคสุวัฒนสกุล, ปิยะบุตร บุญอร่ามเรือง, พัฒนาพร โกวพัฒนกิจ, ชวิน อุ่นภัทร, ฐิติรัตน์ ทิพย์สัมฤทธิ์กุล, และเยาวลักษณ์ ชาติบัญชาชัย. 2565. *แนวปฏิบัติเกี่ยวกับมาตรฐานการใช้ปัญญาประดิษฐ์ (Thailand Artificial Intelligence Guidelines 1.0)*. Bangkok: ศูนย์วิจัยกฎหมายและการพัฒนา คณะนิติศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

อาทิตย์ สุริยะวงศ์กุล. 2567. 'รวมเอกสารข้อเสนอการกำกับกิจการ AI ของไทย (ต.ค. 2565 / มี.ค. 2567)'. *Bact' Is a Name*. เข้าถึง 4 ตุลาคม 2567 (https://bact.cc/2022/thailand-ai-regulations/).

ภาษาอังกฤษ

Abercrombie, Gavin, Djalel Benbouzid, Paolo Giudici, Delaram Golpayegani, Julio Hernandez, Pierre Noro, Harshvardhan Pandit, Eva Paraschou, Charlie Pownall, Jyoti Prajapati, Mark A. Sayre, Ushnish Sengupta, Arthit Suriyawongkul, Ruby Thelot, Sofia Vei, and Laura Waltersdorfer. 2024. 'A Collaborative, Human-Centred Taxonomy of AI, Algorithmic, and Automation Harms'. arXiv.Org. Retrieved 8 October 2024 (https://arxiv.org/abs/2407.01294)

Agarwal, Avinash, and Manisha J. Nene. 2024. 'Addressing AI Risks in Critical Infrastructure: Formalising the AI Incident Reporting Process'. Pp. 1–6 in *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*.

Association of Southeast Asian Nations. 2024. 'ASEAN Guide on AI Governance and Ethics'.

Arthit Suriyawongkul. 2024. 'Bact/Sentimentdemo: A Simple Sentiment Analysis Application.'

Bennet, Karen, Gopi Krishnan Rajbahadur, Arthit Suriyawongkul and Kate Stewart. 2024. *Implementing AI Bill of Materials (AI BOM) with SPDX 3.0: A Comprehensive Guide to Creating AI and Dataset Bill of Materials.* The Linux Foundation.

Bovens, Mark. 2007. 'Analysing and Assessing Accountability: A Conceptual Framework1'. *European Law Journal* 13(4):447–68. doi: 10.1111/j.1468-0386.2007.00378.x.

Bovens, Mark, Thomas Schillemans, and Robert E. Goodin. 2014. 'Public Accountability'. Pp. 1–20 in *The Oxford Handbook of Public Accountability*. Oxford; New York: Oxford University Press.

European Parliament and Council of the European Union. 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828*.

Federal Office for Information Security. 2024. 'Technical Guideline TR-03183: Cyber Resilience Requirements for Manufacturers and Products - Part 2: Software Bill of Materials (SBOM)'.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles*

for AI. SSRN Scholarly Paper. ID 3518482. Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.3518482.

Gasser, Urs, and Virgilio A. F. Almeida. 2017. 'A Layered Model for AI Governance'. *IEEE Internet Computing*. doi: 10.1109/MIC.2017.4180835.

GSMA Press Office. 2024. 'GSMA Launches Maturity Roadmap as Telecoms Industry Leads the Way in the Deployment of Responsible AI'. *Newsroom*. Retrieved 4 October 2024 (https://www.gsma.com/newsroom/press-release/gsma-launches-maturity-roadmap-as-telecoms-industry-leads-the-way-in-the-deployment-of-responsible-ai/).

Gundersen, Odd Erik. 2023. *Improving Reproducibility of Artificial Intelligence Research to Increase Trust and Productivity*. Paris: OECD. doi: 10.1787/3f57323a-en.

Haresamudram, Kashyap, Stefan Larsson, and Fredrik Heintz. 2023. 'Three Levels of AI Transparency'. *Computer* 56(2):93–100. doi: 10.1109/MC.2022.3213181.

International Organization for Standardization and International Electrotechnical Commission. 2021. 'ISO/IEC 5962:2021 Information Technology — SPDX Specification V2.2.1'.

International Organization for Standardization and International Electrotechnical Commission. 2023. 'ISO/IEC 5338:2023 Information Technology — Artificial Intelligence — AI System Life Cycle Processes'.

International Organization for Standardization and International Electrotechnical Commission. 2023. 'ISO/IEC 42001:2023 Information Technology — Artificial Intelligence — Management System'.

Lessig, Lawrence. 2006. *Code: And Other Laws of Cyberspace, Version 2.0*. New York: Basic Books.

Lundberg, Scott M., and Su-In Lee. 2017. 'A Unified Approach to Interpreting Model Predictions'. in *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Mbiazi, Dave, Meghana Bhange, Maryam Babaei, Ivaxi Sheth, and Patrik Joslin Kenfack. 2023. 'Survey on AI Ethics: A Socio-Technical Perspective'. *arXiv.Org*. Retrieved 4 October 2024 (https://arxiv.org/abs/2311.17228v1).

National Institute of Standards and Technology. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology (U.S.). doi: 10.6028/NIST.AI.100-1.

NTIA Multistakeholder Process on Software Component Transparency Framing Working Group. 2021. *Framing Software Component Transparency: Establishing a Common Software Bill of Materials (SBOM)*. National Telecommunications and Information Administration.

Organisation for Economic Cooperation and Development. 2022. *OECD Framework for the Classification of AI Systems*. Paris: OECD. doi: 10.1787/cb6d9eca-en.

Organisation for Economic Cooperation and Development. n.d. 'How Artificial Intelligence Works'. *OECD.AI*. Retrieved 8 October 2024 (https://oecd.ai/en/inside-artificial-intelligence).

OWASP Foundation. n.d. 'CycloneDX - Machine Learning Bill of Materials (ML-BOM)'. Retrieved 10 October 2024 (https://cyclonedx.org/capabilities/mlbom/).

Piyatumrong, Apivadee. 2024. 'Building a Responsible AI Ecosystem: Thailand's Journey Towards Ethical AI'. Pp. 9–11 in *Human Choice and Computers*, edited by R. M. Davison and D. Kreps. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-67535-5_2

Puscas, Ioana. 2023. *AI Risks Taxonomy: Paving the Path for Confidence-Building Measures*. UNIDIR.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. '"Why Should I Trust You?": Explaining the Predictions of Any Classifier'. Pp. 1135–44 in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *KDD '16*. New York, NY, USA: Association for Computing Machinery.

Roy, Alka. 2021. 'LF AI & Data Announces Principles for Trusted AI – LFAI & Data'. Retrieved 5 October 2024 (https://lfaidata.foundation/blog/2021/02/08/lf-ai-data-announces-principles-for-trusted-ai/).

The Linux Foundation and its Contributors. 2024. 'SPDX Specification 3.0'. Retrieved 10 October 2024 (https://spdx.github.io/spdx-spec/v3.0/).

Turri, Violet, and Rachel Dzombak. 2023. 'Why We Need to Know More: Exploring the State of AI Incident Documentation Practices'. Pp. 576–83 in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, *AIES '23*. New York, NY, USA: Association for Computing Machinery.

U.S. Department of Commerce. 2021. 'The Minimum Elements For a Software Bill of Materials (SBOM)'.

U.S. General Services Administration. n.d. ‘Understanding and Managing the AI Lifecycle’. *GSA - IT Modernization Centers of Excellence*. Retrieved 8 October 2024 (https://coe.gsa.gov/coe/ai-guide-for-government/understanding-managing-ai-lifecycle/).

Vassilev, Apostol, Alina Oprea, Alie Fordyce, and Hyrum Anderson. 2024. *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*. NIST AI 100-2e2023. Gaithersburg, MD: National Institute of Standards and Technology (U.S.). doi: 10.6028/NIST.AI.100-2e2023.

Zhou, Jianlong, Fang Chen, Adam Berry, Mike Reed, Shujia Zhang, and Siobhan Savage. 2020. ‘A Survey on Ethical Principles of AI and Implementations’. Pp. 3010–17 in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*.