# รายงานการศึกษารูปแบบการทำงานของเทคโนโลยีปัญญาประดิษฐ์

# และ

# (ร่าง) หลักเกณฑ์การตรวจประเมินการทำงานของโปรแกรมที่มีการใช้เทคโนโลยีปัญญาประดิษฐ์

**เสนอ**

**สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์**

**โดย**

**ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ**

## สารบัญ

**สารบัญรูป**

# บทที่ 1 รายงานการศึกษารูปแบบการทำงานของเทคโนโลยีปัญญาประดิษฐ์

## ความหมายของปัญญาประดิษฐ์

ปัญญาประดิษฐ์ (Artificial intelligence: AI) เป็นคำที่ถูกพูดถึงในวงกว้างมามากกว่า 2 ทศวรรษ และได้รับความสนใจเป็นอย่างมากทั่วโลก ปัญญาประดิษฐ์หมายถึงระบบคอมพิวเตอร์ประเภทหนึ่ง ที่มีกลไกการทำงานอย่างชาญฉลาด มีความสามารถคิด วิเคราะห์ วางแผน และตัดสินใจได้ โดยมีการให้คำนิยามของปัญญาประดิษฐ์ใน 2มิติกว้าง ๆ คือ ระบบที่คิดหรือทำได้เหมือนมนุษย์ (Systems that think like humans) หรือ ระบบที่คิดหรือทำได้อย่างมีเหตุผล (Systems that think rationally) ปัญญาประดิษฐ์เป็นเรื่องที่ใกล้ตัว และได้ใช้งานอยู่ในชีวิตประจำวันกัน เช่น ระบบนำทาง ระบบแนะนำภาพยนตร์หรือสินค้า ระบบสั่งงานด้วยเสียง เป็นต้น ปัญญาประดิษฐ์เสามารถแบ่งออกเป็นระดับตามความสามารถหรือความฉลาด ได้ออกเป็น 3 ระดับ คือ

1. ปัญญาประดิษฐ์เชิงแคบ (**Artificial Narrow Intelligence: ANI**) หรือ ปัญญาประดิษฐ์แบบอ่อน (**Weak AI**) หมายถึง ปัญญาประดิษฐ์ที่มีความสามารถและความเชี่ยวชาญในเฉพาะด้าน

2. ปัญญาประดิษฐ์ทั่วไป (**Artificial General Intelligence: AGI**) หมายถึง ปัญญาประดิษฐ์ ที่มีความสามารถและความฉลาดเทียบเท่ามนุษย์ ทั้งความสามารถการคิดเชิงเหตุผล การวางแผนและแก้ปัญหาการคิดในเชิงซับซ้อน และสามารถเรียนรู้ได้จากประสบการณ์ โดยมีความเชี่ยวชาญในหลาย ๆ ด้าน การพัฒนา ปัญญาประดิษฐ์ให้มีความสามารถในระดับนี้ มีความยากกว่าการพัฒนาปัญญาประดิษฐ์เชิงแคบเป็นอย่างมาก และในปัจจุบัน ความสามารถของปัญญาประดิษฐ์ในการเชื่อมโยงความรู้หลายสาขาและทำงานได้เฉกเช่นเดียวกับมนุษย์ยังเป็นโจทย์ที่ท้าทาย

3. ปัญญาประดิษฐ์ขั้นสูงสุด (**Artificial Superintelligence: ASI**) หมายถึง ปัญญาประดิษฐ์ที่มีการพัฒนาเหนือขีดจำกัดของมนุษย์ที่มีทั้งสติปัญญาและความสามารถเหนือกว่ามนุษย์ในทุกสาขา รวมถึงความคิดเชิงสร้างสรรค์ทางวิทยาศาสตร์ ความรู้เชิงภูมิปัญญา และมีความรู้สึกนึกคิดเช่นเดียวกับมนุษย์ โดยปัญญาประดิษฐ์ในระดับนี้ยังคงเป็นเพียงทฤษฎี

จุดประสงค์ของปัญญาประดิษฐ์ คือการช่วยให้มนุษย์สามารถที่จะช่วยตัดสินใจในสิ่งที่ซับซ้อนได้โดยอ้างอิงจากข้อมูลที่ปัญญาประดิษฐ์สรุปหรือสังเคราะห์ออกมา การพัฒนาเทคโนโลยีปัญญาประดิษฐ์ในประเทศไทย ได้มีการวิจัย พัฒนา และนำปัญญาประดิษฐ์มาประยุกต์ใช้เป็นเวลามากกว่าสิบปี ผ่านการวิจัยพัฒนาขององค์กรภาครัฐ ภาคเอกชน และความร่วมมือกับบริษัทปัญญาประดิษฐ์ทั้ง ในประเทศไทยและต่างประเทศ โดยวันที่ 25 กุมภาพันธ์ 2564 คณะทำงานจัดทำแผนแม่บทปัญญาประดิษฐ์แห่งชาติเพื่อการพัฒนาประเทศไทย เริ่มเดินหน้าจัดทำแผนแม่บทปัญญาประดิษฐ์แห่งชาติ เพื่อการพัฒนาประเทศไทย ในปัจจุบันปัญญาประดิษฐ์สามารถถูกนำไปประยุกต์ใช้งานในหลากหลายอุตสาหกรรม อาทิเช่น ด้านการเงินและธนาคาร งานบริการภาครัฐ ด้านสุขภาพ และด้านอสังหาริมทรัพย์ เป็นต้น

## เทคนิคในการพัฒนาปัญญาประดิษฐ์

การพัฒนาปัญญาประดิษฐ์มีหลากหลายแนวทาง โดยงานวิจัยด้านปัญญาประดิษฐ์ในช่วงเริ่มต้น อิงจากการให้เหตุผลแบบทีละขั้น ๆ เป็นการให้เหตุผลแบบเดียวกับที่มนุษย์ใช้ในการไขปัญหาหรือหาข้อสรุปทางตรรกศาสตร์จากฐานความรู้ที่จัดเก็บในเครื่อง ถัดมางานวิจัยในแวดวงปัญญาประดิษฐ์ได้ให้ความสำคัญกับการเรียนรู้ของเครื่อง (machine learning) ซึ่งเป็นการพัฒนาและศึกษาอัลกอริทึมคอมพิวเตอร์ที่ขั้นตอนวิธีจะถูกปรับปรุงอย่างอัตโนมัติผ่านการเรียนรู้จากประสบการณ์ จนในปัจจุบันได้พัฒนามาจนถึงการเรียนรู้เชิงลึกที่เป็นแนวทางที่ได้รับความนิยมอย่างแพร่หลายในการประยุกต์ใช้อุตสาหกรรมต่าง ๆ

### 1. ระบบฐานความรู้ (Knowledge Base)

ฐานความรู้ (knowledge base : KB) เป็นฐานข้อมูลชนิดพิเศษสำหรับการจัดการความรู้ ฐานความรู้เป็นแหล่งเก็บสารสนเทศที่มีวิธีการรวบรวม จัดการ แบ่งปัน สืบค้น และนำสารสนเทศมาใช้ให้เป็นประโยชน์ โดยฐานความรู้สำหรับระบบปัญญาประดิษฐ์ เป็นฐานความรู้ที่เครื่องอ่านได้โดยเฉพาะ เพื่อจุดประสงค์ของการใช้เหตุผลนิรนัยอัตโนมัติ (Automated Inference)

ฐานความรู้ที่เครื่องอ่านได้ จัดเก็บข้อมูลที่อธิบายความรู้ในแบบเชื่อมโยงกันเชิงตรรกศาสตร์ ด้วยโครงสร้างข้อมูล โดยฐานความรู้ที่เป็นที่นิยม คือ ออนโทโลยี (Ontology) ที่เก็บนิยามของสรรพสิ่งที่เกี่ยวข้องเข้าเป็นโครงสร้างของข้อมูล และใช้เป็นฐานความรู้สำหรับ ระบบผู้เชี่ยวชาญ (Expert system) ระบบแนะนำ (Recommendation system) และ ระบบสืบค้นเชิงความหมาย (Semantic search) ทั้งนี้ ออนโทโลยี

สามารถใช้งานร่วมกับกลไกการอนุมาน (Inference Engine) ผ่านตัวดำเนินการทางตรรกศาสตร์ต่าง ๆ อย่างเช่น แอนด์ (การเชื่อม) ออร์ (การเลือก) การมีเงื่อนไข และนิเสธ เพื่อหาข้อสรุปเชิงตรรกะ สำหรับถูกนำไปใช้กับระบบปัญญาประดิษฐ์

## 1.1 เทคโนโลยีที่เครือข่ายความหมาย

เทคโนโลยีที่เครือข่ายความหมาย หรือ เว็บเชิงความหมาย (Semantic web) คือการสร้างเครือข่ายของ สิ่งที่เกี่ยวข้องกันแบบมีแผนโครงสร้างข้อมูลที่เป็นความหมายของคำมีความสัมพันธ์กันในส่วนโครงสร้างใน รูปแบบกราฟ เพื่อเป็นกราฟความรู้ (Knowledge graph) ให้กับระบบสารสนเทศ (Information technology) เทคโนโลยีเครือข่ายความหมายนี้เป็นเทคโนโลยีที่ใช้ในการจัดเก็บและนำเสนอเนื้อหาข้อมูล แบบมีโครงสร้างในลักษณะของการเชื่อมโยงความสัมพันธ์ของข้อมูล ในระดับเมทาเดทา (Metadata) ตาม หลักการวิศวกรรมความรู้ (Knowledge engineering) ออกมาเป็น โครงสร้างตัวแทนความรู้แบบออนโทโลยี (Ontology knowledge representation)

เครือข่ายความหมายเป็นทิศทางการพัฒนาเทคโนโลยีเว็บหรือการจัดเก็บข้อมูลตามเทคโนโลยีเว็บ 3.0 ซึ่งคาดหวังให้ข้อมูลมีการเชื่อมโยงกันมากยิ่งขึ้น ในลักษณะของเครือข่ายเชิงความหมาย (Semantic Network) เพื่อนำไปสู่การพัฒนาโปรแกรมประยุกต์ที่มีความชาญฉลาดและเข้าใจถึงข้อมูลต่างๆ ตามนัยความหมายเฉกเช่นที่มนุษย์เข้าใจ โดยมีหน่วยงาน W3C (http://www.w3.org/) เป็นองค์กรสากลที่เป็นผู้ กำหนดแนวทางการพัฒนาและมาตรฐานสำหรับข้อมูลเครือข่ายเชิงความหมาย

เครือข่ายความหมายเป็นเทคโนโลยีที่นิยมใช้ในการจัดการความรู้ (Knowledge management) โดยเฉพาะอย่างยิ่งความรู้เชิงลึกภายในองค์กร เนื่องจากความรู้ (knowledge) คือ ผลสรุปของการสังเคราะห์ สารสนเทศ (information) และข้อมูล (Data) โดยพิจารณาถึงความสัมพันธ์ของข้อมูลตามหลักการความหมาย (semantic) ความเชื่อมโยง ความเหมือน และความเฉพาะเจาะจงของข้อมูลนั้น จนได้ผลสรุปที่ชัดเจนถูกต้อง และความสัมพันธ์ของข้อมูลเหล่านี้สามารถนำแสดงไว้ในรูปแบบเครือข่ายความหมาย ที่มนุษย์และคอมพิวเตอร์อ่านและทำความเข้าใจได้ร่วมกัน จนนำไปประยุกต์ใช้ในกิจกรรมต่างๆ ต่อไปได้อย่างเหมาะสม

การจัดการความรู้โดยเครือข่ายความหมาย (Semantic Knowledge Management) จึงเป็นรูปแบบ การจัดการความรู้ที่มุ่งเน้นการจัดเก็บองค์ความรู้ที่สามารถนำไปใช้งานในโปรแกรมคอมพิวเตอร์ได้ในรูปแบบ ของฐานความรู้สำหรับโปรแกรมคอมพิวเตอร์ในรูปแบบโครงสร้างตัวแทนความรู้แบบออนโทโลยี (Ontology) การพัฒนาโครงสร้างตัวแทนความรู้แบบออนโทโลยีอาศัยการสกัดความรู้จากแหล่งความรู้ที่มีอยู่ ทั้งที่อยู่ในรูปแบบของเอกสารอ้างอิง (Reference documents) และจากผู้เชี่ยวชาญเฉพาะสาขา (Domain experts) ดังนั้นการพัฒนาออนโทโลยีจึงต้องมีการผสมผสานทั้งการสังเคราะห์ความรู้ที่ชัดแจ้ง (Explicit Knowledge) และการจัดการความรู้ที่อยู่ในตัวบุคคล (Tacit Knowledge) เข้าด้วยกันเพื่อให้จัดเก็บอยู่ในโครงสร้างที่เป็นเครือข่ายของตัวแทนความคิด (Concept Network) ที่มีความสามารถอธิบายความสัมพันธ์ภายใน ในระดับสกีมา (schema) ได้อย่างเป็นเหตุเป็นผลและไม่กำกวม

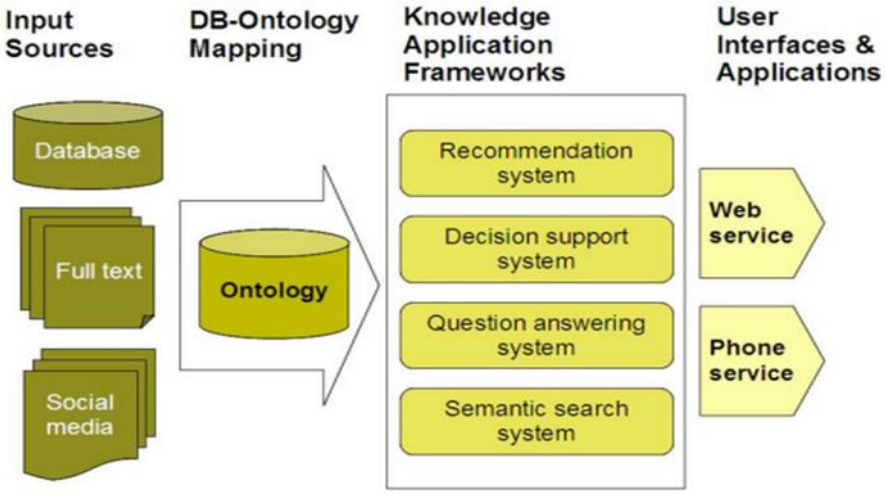
## 1.2 ออนโทโลยี

ออนโทโลยี (Ontology) เป็นรูปแบบองค์ความรู้เฉพาะสาขา (Domain Knowledge) ที่เกิดจากการ วิศวกรความรู้ (Knowledge engineers) ร่วมกับผู้เชี่ยวชาญเฉพาะทาง (Domain experts) โดยมี วัตถุประสงค์หลักคือเพื่อให้สามารถนำความรู้เฉพาะทาง ไปประยุกต์ใช้ในโปรแกรมคอมพิวเตอร์ได้ หลากหลายชนิด รวมถึงการจัดเก็บอย่างเป็นระบบ และแบ่งปันองค์ความรู้จากผู้เชี่ยวชาญเฉพาะทางให้สามารถนำไปใช้งานได้ในโปรแกรมและระบบคอมพิวเตอร์ต่างๆ ให้สามารถทำงานได้อย่างชาญฉลาดและมีความเป็นอัตโนมัติ ออนโทโลยีจัดเก็บความเชื่อมโยงของสิ่งต่าง ๆ ในรูปแบบกราฟ กราฟของออนโทโลยีจะแสดงความเชื่อมโยงของคอนเซ็ปต์ต่างๆ เพื่อช่วยให้คอมพิวเตอร์เข้าใจ ความหมายของคอนเซ็ปต์ในแง่ของความสัมพันธ์ระหว่างคอนเซ็ปต์

ออนโทโลยีนั้นเป็นโครงสร้างโครงข่ายในระดับสกีมา (schema level) กล่าวคือ เป็นโครงสร้างใน ระดับความคิด ดังนั้นจึงต้องนำข้อมูลแท้ (Instance) มาผูกกับโนด (node) ต่างๆ ของออนโทโลยีเพื่อเป็นตัวแทนในระดับข้อมูล (Data level) ความสัมพันธ์จากออนโทโลยีก็จะส่งมอบมาที่ระดับข้อมูล ทำให้ข้อมูลมีความเชื่อมโยงกันอย่างมีนัยยะตามหลักความเป็นจริง (Fact) และหลักตรรกะ (Logic) ซึ่งสามารถที่จะช่วยวิเคราะห์จำแนกและจัดแบ่งได้ว่า ข้อมูลที่ปรากฏนั้น มีความสัมพันธ์กับข้อมูลอื่นอย่างไร โดยการประมวลผลได้โดยอัตโนมัติจากการจัดกลุ่ม โดยใช้วิธีการแยกความ แตกต่างจากคุณลักษณะ และความหมายของคำ ทำให้ง่ายต่อการสืบค้นอ้างอิง โดยเพิ่มศักยภาพในการชี้ไปถึง สิ่งที่คล้ายกันถึงแม้จะใช้คำต่างกันหรือมีความแตกต่างทางกายภาพได้นอกจากนั้นยังสามารถนำไปใช้ในการ อนุมานหาคำตอบตามหลักการทางตรรกศาสตร์ (Logic

Inference) หรือการคาดคะเนตามหลักเหตุผลเพื่อหา ข้อสรุปจากหลักเกณฑ์และข้อเท็จจริงที่มีอยู่ได้อย่างมีประสิทธิภาพ เป็นกลาง และไม่มีอคติ

ปัจจุบันมีการนำออนโทโลยีมาใช้ในโปรแกรมประยุกต์หลายประเภท ได้แก่ ระบบค้นหาเชิงความหมาย (Semantic search) ระบบแนะนำ (Recommendation) ระบบถามตอบ (Question answering) และระบบสนับสนุนการตัดสินใจ (Decision-support) โดยการออนโทโลยีมาใช้จะมี สถาปัตยกรรมพื้นฐานดังรูปที่ 4 คือมีการเชื่อมโยงระหว่างข้อมูลแท้กับออนโทโลยีเพื่อให้เป็นฐานความรู้ให้กับ ระบบ และเชื่อมต่อกับผู้ใช้งานผ่านทางหน้าจอผู้ใช้งาน (User interface)




*รูป 1 สถาปัตยกรรมพื้นฐานของระบบที่ใช้ออนโทโลยีเป็นฐานความรู้สำหรับปัญญาประดิษฐ์*

การสกัดความรู้สำหรับการพัฒนาออนโทโลยี ประกอบไปด้วยวามรู้ 2 ประเภท คือ ความรู้ชัดแจ้ง (Explicit Knowledge) และ ความรู้ซ่อนเร้น (Tacit Knowledge) โดย ความรู้ชัดแจ้ง เป็นความรู้ที่สามารถบันทึกได้ ในรูปแบบที่เป็นเอกสาร หรือ วิชาการ อยู่ในตำรา คู่มือปฏิบัติงาน ซึ่งสามารถถ่ายทอด และเรียนรู้เข้าใจได้ง่าย ส่วนความรู้ซ่อนเร้นเป็นความรู้ที่เกิดจาก ประสบการณ์ สัญชาตญาณ ความชำนาญที่เกิดจากการสั่งสมมาเป็นระยะเวลานาน และไม่สามารถสามารถบันทึกได้หรือบันทึกได้ไม่หมด กล่าวคือ ความรู้ที่ทำให้บุคคลมีความเชี่ยวชาญเฉพาะ ดังนั้นการสกัดความรู้ซ่อนเร้น จำเป็นต้องมีการปฏิสัมพันธ์ที่ใกล้ชิด เพื่อซักถามและสังเกต

องค์ประกอบของออนโทโลยี จะเน้นไปที่คลาส ซึ่งอธิบายถึงคอนเซ็ปต์ของสิ่งใดสิ่งหนึ่งในขอบเขตที่ควรจะเป็น และจัดแยกหมวดหมู่ให้คลาสเป็นกลุ่ม ที่เชื่อมโยงกัน ในรูปแบบโครงสร้างข้อมูลแบบลำดับชั้นและหมวดหมู่ นอกจากความสัมพันธ์แบบลำดับชั้น ออนโทโลยีอนุญาตให้มีความสัมพันธ์ระหว่างคลาสแบบอื่นๆ ได้แก่ การเป็นส่วนประกอบ (Part-of relation) การมีคุณสมบัติ (Attribute-of relation) เพื่ออธิบายความสัมพันธ์ตามข้อเท็จจริงในโลก โดยความสัมพันธ์เหล่านี้ สามารถช่วยให้ระบบปัญญาประดิษฐ์ประมวลผล อ่านและทำการวิเคราะห์ข้อมูลได้ตามข้อเท็จจริง

ในปัจจุบันได้กำหนดภาษามาตรฐานที่ใช้จำลองและออกแบบโครงสร้างของเอกสารเอกซ์เอ็มแอล (XML) เรียกว่า OWL และ RDF/XML โดยใช้นิยามแนวคิดให้อยู่ในรูปของกฎ คลาส ความสัมพันธ์ระหว่างคลาส และสมบัติของคลาส หนึ่งในวิธีการจัดเก็บองค์ความรู้และข้อมูลที่ใช้กันแพร่หลายคือการเข้ารหัสข้อมูลในรูปแบบที่มนุษย์และคอมพิวเตอร์สามารถอ่านได้ด้วยภาษาทีเรียกว่า RDF/XML (Resource Description Framework - Extensible Markup Language) ภาษานี้ใช้ในการจัดเก็บชุดกลุ่มข้อมูลเพื่ออธิบายข้อความในเอกสารดิจิทัล เพื่อสร้างมาตรฐานในการจัดเก็บองค์ความรู้และข้อมูลโดยภาษา XML องค์กรระดับนานาชาติที่มีเป้าหมายในการพัฒนาเทคโนโลยีเว็บ (World Wide Web Consortium - W3C) ได้กำหนดมาตรฐานกรอบคำอธิบายทรัพยากรข้อมูล (Resource Description Framework - RDF) เมื่อนำวิธีการจัดเก็บและมาตรฐานมาใช้ร่วมกัน จะเรียกว่า RDF/XML ที่สามารถใช้แลกเปลี่ยนระหว่างคอมพิวเตอร์ต่างประเภทกัน, ระบบปฏิบัติการที่ต่างกัน หรือใช้แอพพลิเคชั่นที่ใช้ภาษาต่างกันได้ ตัวอย่างการใช้งาน RDF/XML มีดังรูปที่ 2

```
<rdf:Description rdf:about="https://lst.nectec.or.th">

  <lst:title>Language and Semantic Technology Laboratory</lst:title>

  <lst:author>Akkharawoot Takhom</lst:author>

</rdf:Description>
```

*รูป 2 ตัวอย่างมาตรฐานกรอบคำอธิบายทรัพยากรข้อมูล RDF/XML สำหรับจัดเก็บ*

การพัฒนาออนโทโลยี นิยมใช้โปรแกรมเครื่องมือสำหรับพัฒนาออนโทโลยี (Ontology Editor) เพื่อลดภาระในการกำกับข้อมูลที่มีความยุ่งยากสำหรับการเชื่อมโยงความรู้ โดยในปัจจุบัน โปรแกรมเครื่องมือสำหรับพัฒนาออนโทโลยีที่ได้รับความนิยม มีหลายเครื่องมือ เช่น โปรแกรม Protege ซึ่งพัฒนาโดย มหาวิทยาลัยสแตนฟอร์ด (Stanford University) และ โปรแกรมHozo ซึ่งพัฒนาโดยมหาวิทยาลัยโอซากา (Osaka University) เป็นต้น โดยเครื่องมือเหล่านี้เป็นเครื่องมือสนับสนุนกระบวนการวิศวกรรมความรู้ ที่ช่วยให้ผู้ใช้ที่เป็นวิศวกรความรู้ หรือผู้เชี่ยวชาญเฉพาะสาขา สามารถถ่ายทอดและจัดเก็บองค์ความรู้ในรูปแบบของออนโทโลยีได้สะดวก และง่ายดายยิ่งขึ้น

### 1.3 ระบบผู้เชี่ยวชาญ

ระบบผู้เชี่ยวชาญ (Expert system) เป็นระบบคอมพิวเตอร์ที่ช่วยในการหาคำตอบและอธิบายความไม่ชัดเจนผ่านแบบจำลองการตัดสินใจของมนุษย์ ซึ่งใช้ความรู้เฉพาะทางที่สกัดได้จากผู้เชี่ยวชาญในแต่ละสาขาเพื่อตอบคำถามนั้นตามหลักการทางตรรกะ โดยระบบผู้เชี่ยวชาญเป็นส่วนหนึ่งของปัญญาประดิษฐ์ในรุ่นบุกเบิกโดยอาศัยฐานความรู้ (knowledge-based system) และกลไกการอนุมาน (inference engine) เป็นองค์ประกอบหลักในการทำงาน ทั้งนี้ ในขั้นตอนการพัฒนาระบบผู้เชี่ยวชาญ ต้องมีการจัดเตรียมข้อมูลในรูปแบบฐานความรู้ให้พร้อมสำหรับนำไปใช้งาน ครอบคลุมความรู้ที่ต้องใช้ตามวัตถุประสงค์ของระบบ และมักจะถูกพัฒนาให้สามารถตอบสนอง ต่อปัญหาในทันทีที่เกิดความต้องการและสามารถปฏิบัติงานแทนผู้เชี่ยวชาญ อย่างมีประสิทธิภาพ กล่าวคือระบบผู้เชี่ยวชาญ จะสามารถให้ผลลัพธ์ ทั้งการตัดสินปัญหาและการหาคำตอบได้อย่างแน่นอน เนื่องจากระบบถูกพัฒนาให้สามารถปฏิบัติงานโดยปราศจากผลกระทบ ทางร่างกายและอารมณ์ที่มีอยู่ในตัวมนุษย์เช่น ความเครียด ความเจ็บ ป่วย เป็นต้น

การสร้างระบบผู้เชี่ยวชาญโดยทั่วไปนิยมวิธีการสร้างแบบวนซ้ำ (Interative process) โดยการเริ่มสร้างจากระบบที่มีขนาดเล็กก่อน แล้วจึงค่อยขยายขนาดของระบบ โดยเพิ่มจำนวนกฎการตัดสินใจเข้าไป และทำการทดสอบระบบและวนซ้ำ จนกว่าจะได้ระบบที่สมบูรณ์ ทั้งนี้การพัฒนาระบบงานปัญญาประดิษฐ์จำเป็นต้องทีผู้เชี่ยวชาญ (Expert) ซึ่งมีความรู้ความเข้าใจอย่างถ่องแท้ในความรู้ที่จะสร้างฐานความรู้มาทดแทน และมีวิศวกรความรู้ (Knowledge engineer) เพื่อแปลงความรู้จากผู้เชี่ยวชาญให้อยู่ในรูปแบบฐานความรู้และกฎการตัดสินใจ

ข้อดีของระบบผู้เชี่ยวชาญ ได้แก่

1. ระบบผู้เชี่ยวชาญ จะช่วยขยายขีดความสามารถในการตัดสินใจให้กับผู้รับบริการจำนวนมากพร้อมๆ กัน
2. ระบบผู้เชี่ยวชาญให้ผลลัพธ์สม่ำเสมอและไม่ขัดแย้งกัน

### 1.4 กลไกอนุมานเชิงตรรกะ

กลไกอนุมานเชิงตรรกะ (Inference engine) เป็นส่วนควบคุมการใช้ความรู้ในฐานความรู้ เพื่อวิเคราะห์และสรุปหาคำตอบให้กับปัญหาเฉพาะกรณี (case based issue) กล่าวคือ กลไกอนุมานเชิงตรรกะเป็นส่วนการใช้เหตุและผลร่วมกับฐานความรู้ โดยทำหน้าที่เป็นความรู้ในการตัดสินใจ (Operational Knowledge) ทั้งนี้ กลไกอนุมานเชิงตรรกะมักใช้ร่วมกับระบบผู้เชี่ยวชาญในการตัดสินและสรุปคำตอบผ่านการอนุมานทางตรรกะสำหรับแต่ละเหตุการณ์ ซึ่งมักจะอยู่ในลักษณะ กฎข้อแม้ และ ผลของกฎ ดังแสดงในรูปที่ 3

IF <condition$^1$, condition$^2$ ... condition$^n$>

THEN <consequence$^1$... consequence$^n$ >

*รูป 3 ตัวอย่างรูปแบบกฎการอนุมานสำหรับกลไกอนุมานเชิงตรรกะ*

การอนุมานเชิงตรรกะมีอยู่ 2 ประเภทหลักได้แก่

การอนุมานไปข้างหน้า (Forward Chaining Inference) คือการอนุมานโดยเริ่มการตรวจสอบข้อมูลกับกฎที่มีอยู่ในระบบจนกว่าจะสามารถหากฎเกณฑ์ที่สอดคล้องกับสถานการณ์แล้วจึงดำเนินงานต่อส่วนผล การอนุมานแบบนี้ อาจจะได้ผลลัพธ์มากกว่า 1 ผล หากข้อเท็จจริงที่ใส่ให้ตรงกับข้อแม้ของกฎหลายข้อ

การอนุมานย้อนหลัง (Backward Chaining Inference) คือการอนุมานโดยเริ่มจากเป้าหมาย (Goals) ที่ต้องการและทำงานย้อนกลับโดยการวินิจฉัยข้อเท็จจริงที่ได้รับร่วมกับกฎและโครงสร้างฐานความรู้ โดยการอนุมานในลักษณะนี้มักนำมาใช้ในการพัฒนาระบบความฉลาดให้มีความเข้าใจ และมีประสบการณ์ในการแก้ปัญหา เพื่อให้ระบบสามารถทำการอนุมานหาข้อสรุปของปัญหาที่เกิดขึ้นในอนาคต

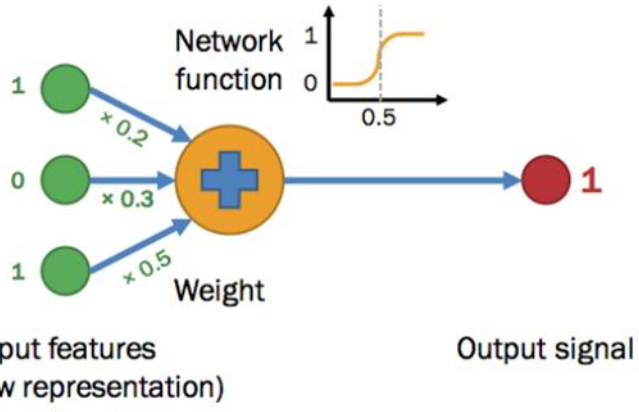
## 2 การเรียนรู้ในระดับลึก (Deep Learning)

การเรียนรู้ในระดับลึก (deep learning) เป็นเทคนิคการเรียนรู้ด้วยเครื่องจักรแบบหนึ่ง ซึ่งใช้วิธีการ สกัดคุณลักษณะเด่นจากข้อมูลปริมาณมากโดยอัตโนมัติเพื่อนำไปใช้ในการจำแนกประเภทของวัตถุหรือใช้ใน การตัดสินใจ เทคนิคนี้สามารถช่วยแก้ปัญหามิติของแบบจำลอง (the curse of dimensionality) ซึ่งเป็น ปัญหาพื้นฐานที่สำคัญในการประมวลผลภาษาธรรมชาติหากกล่าวโดยสรุปแล้ว มิติของแบบจำลองก็คือ จำนวนพารามิเตอร์ (parameter) ของแบบจำลองสถิติ (statistical model) นั่นเอง ยิ่งพารามิเตอร์ของ แบบจำลองมีมากเท่าใดก็ยิ่งทำให้แบบจำลองนั้นทำนายผลลัพธ์ได้ถูกต้องมากขึ้นเท่านั้น แต่การเพิ่มจำนวน พารามิเตอร์ของแบบจำลองจะทำให้ต้องใช้ข้อมูลจำนวนเพิ่มขึ้นแบบเท่าทวีคูณ เพื่อให้ข้อมูลมีปริมาณเพียงพอ และมีอุปกรณ์ประมวลผลสมรรถนะสูงเพียงพอที่จะปรับค่าพารามิเตอร์ของแบบจำลองได้นั่นทำให้การ ออกแบบพารามิเตอร์ของแบบจำลองตามกระบวนการมาตรฐานทางสถิติยังมีข้อจำกัดของมิติของแบบจำลอง ปริมาณข้อมูลที่ต้องใช้และอุปกรณ์ประมวลผลสมรรถนะสูง อย่างไรก็ตาม ด้วยการพัฒนาอย่างก้าวกระโดดของเทคโนโลยีอินเตอร์เน็ต เทคโนโลยีการเก็บข้อมูล ขนาดใหญ่บนกลุ่มเมฆ (cloud storage) และการประมวลผลข้อมูลด้วยเครือข่ายหน่วยประมวลผลขนาน ขนาดยักษ์ (massively parallel computing) ปัญหาเรื่องมิติของแบบจำลองจึงเริ่มบรรเทาลงและได้ทำให้การวิจัยในหลายด้านที่เกี่ยวข้องกับการประมวลผลสถิติโดยเฉพาะอย่างยิ่งการประมวลผลภาษาธรรมชาติเข้า สู่ยุคแห่งข้อมูลขนาดมหึมา (big data) นักวิจัยมีอิสระในการออกแบบแบบจำลองและคัดเลือกพารามิเตอร์ที่ เหมาะสมกับงานแต่ละประเภท เพื่อที่จะปรับปรุงประสิทธิภาพและความถูกต้องของระบบมากขึ้น นอกจากนี้ การประมวลผลด้วยเครือข่ายหน่วยประมวลผลขนาดยักษ์ยังช่วยลดเวลาในการปรับค่าพารามิเตอร์ของ แบบจำลองจากข้อมูลปริมาณมหาศาลอีกด้วย

แนวทางหนึ่งที่จะผสานประโยชน์ของทั้งข้อมูลขนาดมหึมาและเครือข่ายหน่วยประมวลผลขนาดยักษ์ ได้คือการเรียนรู้ในระดับลึก ในแนวทางนี้นักวิจัยจะออกแบบแบบจำลองสถิติให้เป็นลำดับชั้นของพารามิเตอร์ โดย

พารามิเตอร์ในแต่ละชั้นจะแทนความรู้เกี่ยวกับข้อมูลในระดับที่ต่างกัน ทำให้สามารถวิเคราะห์แบบจำลองสถิติได้ในระดับลึกได้ตามต้องการ วิธีการออกแบบแบบจำลองในลักษณะนี้จะแตกต่างจากวิธีการดั้งเดิมที่นำพารามิเตอร์ทุกตัวมารวมกันอยู่ในระดับเดียวกันซึ่งทำได้เพียงวิเคราะห์สถิติในระดับผิวเท่านั้น การเรียนรู้ในระดับลึกจึงเป็นแนวทางการออกแบบแบบจำลองสถิติที่ได้รับความนิยมในปัจจุบัน โดยมีเครือข่ายประสาทเทียมแบบเวียนซ้ำ (recurrent neural network) เป็นแบบจำลองที่ใช้พารามิเตอร์เหล่านี้ในการประมวลผล

กลไกพื้นฐานของการเรียนรู้ระดับลึกคือเครือข่ายประสาทเทียมแบบวนซ้ำ (recurrent neural network) ซึ่งเลียนแบบการทำงานของเครือข่ายเซลล์ประสาทในสมองมนุษย์โดยเนื้อแท้แล้วเครือข่าย ประสาทเทียมประกอบไปด้วยหน่วยประสาทเทียม (perceptron) ที่รับชุดสัญญาณประสาทจากเซลล์ประสาท อื่นผ่านช่องสัญญาณเข้าดังรูปที่ 5 เราสามารถเทียบเคียงชุดสัญญาณเข้าเหล่านี้ให้เป็นเวกเตอร์คุณสมบัติ (feature vector) ที่ประกอบไปด้วยสัญญาณ 0 และ 1 ได้ช่องสัญญาณเข้าแต่ละช่องจะมีน้ำหนัก (weight) ต่างกันเป็นพารามิเตอร์ของแบบจำลอง เมื่อได้ถ่วงน้ำหนักให้สัญญาณเข้าแล้ว หน่วยประสาทเทียมก็จะรวม สัญญาณเหล่านั้นเข้าด้วยกันโดยใช้ฟังก์ชันเครือข่าย (network function) และปล่อยสัญญาณออกตาม ฟังก์ชันการกระตุ้น (activation function) ที่กำหนด หากค่าที่ได้จากฟังก์ชันเครือข่ายมีค่าเกินค่าขีดจำกัด (threshold) หน่วยประสาทเทียมก็จะปล่อยสัญญาณ 1 ออกมา

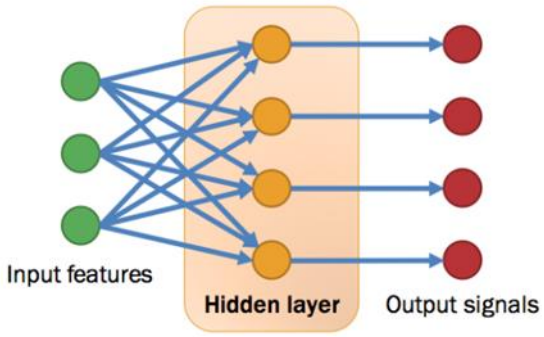



*รูป 4 แสดงหน่วยประสาทเทียม (perceptron)*

การเลือกน้ำหนักของช่องสัญญาณเข้าแต่ละช่องเป็นปัญหาที่ท้าทายอันหนึ่ง เพราะเป็นการปรับค่าพารามิเตอร์ของแบบจำลองซึ่งมีผลกระทบโดยตรงต่อความถูกต้องของการทำนายจากหน่วยประสาทเทียม กระบวนการปรับค่าพารามิเตอร์จะต้องใช้ข้อมูลสำหรับสอน (training data) ซึ่งเป็นรายการของเวกเตอร์

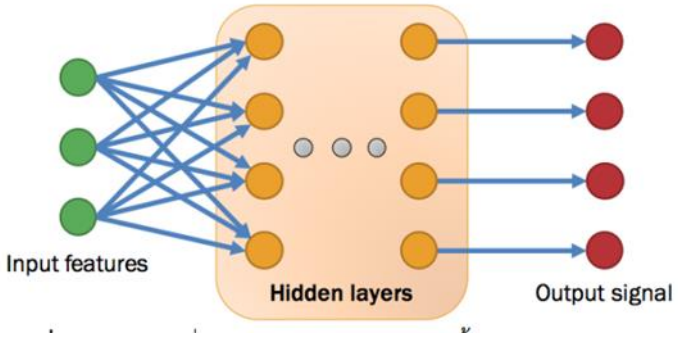
คุณสมบัติของสัญญาณเข้าพร้อมทั้งผลเฉลยเป็นสัญญาณออก ระบบจะปรับค่าน้ำหนักของช่องสัญญาณแต่ละช่องโดยเพิ่มค่าหรือลดค่าลงทีละน้อยตามค่าความชัน (gradient-descent approximation) เพื่อให้เมื่อรับสัญญาณเข้าชุดหนึ่งแล้วจะผลิตสัญญาณออกได้ตรงกับผลเฉลยของสัญญาณเข้าชุดนั้น กระบวนการปรับค่าพารามิเตอร์ดังกล่าวจะดำเนินไปเรื่อยๆ จนกว่าค่าพารามิเตอร์ทั้งหมดจะเริ่มไม่เปลี่ยนแปลง หรือที่เรียกว่าเข้าสู่ภาวะนิ่ง (converge)

เมื่อนำหน่วยประสาทเทียมมาต่อกันเป็นเครือข่ายเราจะได้เครือข่ายประสาทเทียม (artificial neural network หรือ ANN) ดังรูปที่ 6 เครือข่ายประสาทเทียมจะประกอบไปด้วยหน่วยประสาทเทียมจำนวนมากที่เรียงตัวกันเป็นหนึ่งชั้นเรียกว่า ชั้นซ่อนเร้น (hidden layer) หน่วยประสาทเทียมแต่ละตัวในชั้นซ่อนเร้นจะเชื่อมต่อกับช่องสัญญาณเข้าจากเวกเตอร์คุณสมบัติและผลิตสัญญาณตามหลักการของหน่วยประสาทเทียมออกมาเป็นคำทำนาย (prediction) ซึ่งมีลักษณะเป็นเวกเตอร์คุณสมบัติเช่นกัน ชั้นซ่อนเร้นนี้เปรียบเสมือนการกลั่นกรองหนึ่งชั้นเพื่อทำนายคำตอบจากเวกเตอร์คุณสมบัติของสัญญาณเข้านั่นเอง เนื่องจากมีชั้นซ่อนเร้นเพียงชั้นเดียว เราจึงเรียกการเรียนรู้แบบนี้ว่าการเรียนรู้ในระดับตื้น (shallow learning)



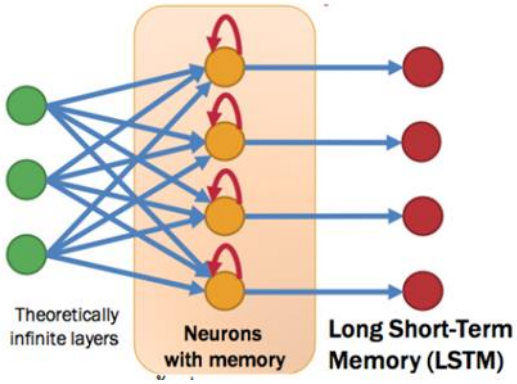

*รูป 5 แสดงเครือข่ายประสาทเทียม (artificial neural network หรือ ANN)*

นอกจากนี้เรายังสามารถเพิ่มชั้นซ่อนเร้นให้มีจำนวนมากขึ้น กลายเป็นเครือข่ายหน่วยประสาทเทียม แบบหลายชั้น (multilayer ANN) ดังรูปที่ 7 การเพิ่มจำนวนชั้นซ่อนเร้นจะช่วยให้เครือข่ายประสาทเทียมสามารถทำนายได้ถูกต้องมากขึ้น เพราะกระบวนการทำนายจะถูกกลั่นกรองโดยชั้นซ่อนเร้นหลายชั้นก่อนที่จะผลิตสัญญาณออก อย่างไรก็ตามการเพิ่มจำนวนชั้นซ่อนเร้นจะเป็นการเพิ่มจำนวนพารามิเตอร์ในเครือข่ายประสาทเทียม ทำให้ต้องใช้ข้อมูลปริมาณมากขึ้นในการปรับค่าพารามิเตอร์ตามไปด้วย โดยกระบวนการปรับค่าพารามิเตอร์จะเริ่มจากปรับค่าที่หน่วยประสาทเทียมในชั้นซ่อนเร้นที่อยู่ใกล้สัญญาณออกก่อน แล้วกวาดกลับเข้าไปหาหน่วยประสาทเทียมที่อยู่ในชั้นซ่อนเร้นที่ไกลออกไป ซึ่งเราเรียกกระบวนการนี้ว่า กระบวนการกวาดย้อนกลับ (backward propagation) กระบวนการกวาดย้อนกลับจะดำเนินการไปเรื่อยๆ จนกว่าค่าพารามิเตอร์ของหน่วยประสาทเทียมทุกจุดจะเริ่มเข้าสู่ภาวะนิ่ง




*รูป 6 แสดงเครือข่ายประสาทเทียมแบบหลายชั้น (multilayer ANN)*

หัวใจสำคัญของการเรียนรู้ระดับลึกคือการเพิ่มจำนวนชั้นซ่อนเร้นให้มีมากจนเป็นอนันต์ เพื่อให้ เทียบเคียงได้กับการทำงานของเครือข่ายเซลล์ประสาทในสมองมนุษย์แต่ด้วยข้อจำกัดที่ว่าการเพิมจำนวนชั้น ซ่อนเร้นก็จะทำให้ต้องใช้ข้อมูลสำหรับสอนเพิ่มขึ้น การเพิ่มชั้นซ ่อนเร้นจนเป็นอนันต์จึงเป็นไปไม่ได้ในทาง ปฏิบัติจึงมีการประดิษฐ์แบบจำลองหลายชนิดขึ้นมาเพื ่อแก้ปัญหาดังกล่าวในการเรียนรู้ในระดับลึก แบบจำลองหนึ่งที่เป็นที่นิยมในปัจจุบันคือเครือข่ายประสาทเทียมแบบวนซ้ำ (recurrent neural network) ซึ่งหน่วยประสาทเทียมแต่ละหน่วยจะนำสัญญาณออกมาใช้ซ้ำเป็นช่องสัญญาณรับเข้าด้วย จึงทำให้หน่วย ประสาทเทียมนี้เสมือนมีหน่วยความทรงจำที่สามารถจดจำสัญญาณออกที่ตัวเองเคยผลิตไว้ได้ด้วย แบบจำลอง นี้มีความเรียบง่ายและสามารถปรับค่าพารามิเตอร์ได้อย่างรวดเร็วแต่ก็มีข้อเสียคือบางครั้งหน่วยประสาท เทียมจะยึดติดกับความทรงจำเก่าจนทำให้ไม่สามารถเรียนรู้สิ่งใหม่ได้อีก เพื่อแก้ปัญหาดังกล่าว จึงมีการเพิ่ม วงจรการลืม (forgetting circuit) เข้าไปในหน่วยประสาทเทียมแต่ละตัว เพื่อให้ปิดกั้นช่องสัญญาณเข้าที่มา จากสัญญาณออกของตัวเอง เราเรียกเครือข่ายประสาทเทียมแบบวนซ้ำที่มีวงจรการลืมนี้ว่า แบบจำลอง ความจำระยะสั้นที่อยู่ระยะยาว (Long Short-Term Memory หรือ LSTM) ดังรูปที่ 8 ซึ่งแบบจำลองนี้เป็น แบบจำลองที่นิยมใช้กันมากที่สุดสำหรับการเรียนรู้ในระดับลึก



*รูป 7 แสดงแบบจำลองความจำระยะสั้นที่อยู่ระยะยาว (Long Short-Term Memory หรือ LSTM)*

จะเห็นได้ว่าแบบจำลองความจำระยะสั้นที่อยู่ระยะยาวมีพารามิเตอร์ 2 ชุด คือ ชุดหนึ่งมาจากหน่วย ประสาทเทียม และอีกชุดหนึ่งมาจากวงจรการลืม เราสามารถใช้กระบวนการกวาดย้อนกลับเพื่อปรับ ค่าพารามิเตอร์ทั้ง 2 ชุดนี้ได้แต่กระบวนการดังกล่าวจึงอยู่ขึ้นกับลำดับเวลาในข้อมูลสำหรับสอนเพราะต้อง ปรับค่าพารามิเตอร์ให้เหมาะสมกับช่วงเวลาที่จะเริ่มลืมความทรงจำดั้งเดิมด้วย

# บทที่ 2 (ร่าง) หลักเกณฑ์การตรวจประเมินการทำงานของโปรแกรมที่มีการใช้เทคโนโลยีปัญญาประดิษฐ์

## บทนำ

หลักเกณฑ์การประเมินการทำงานของโปรแกรมที่มีการใช้เทคโนโลยีปัญญาประดิษฐ์ฉบับนี้ มีเป้าหมายสำหรับหน่วยงานที่ต้องการทำหน้าที่เป็นหน่วยประเมินและทดสอบระบบที่ใช้เทคโนโลยีปัญญาประดิษฐ์ โดยครอบคลุมตามระบบคอมพิวเตอร์ และซอฟต์แวร์ตามนิยาม “AI-based system is a system that includes at least one AI component” [1] ทั้งนี้การประยุกต์ใช้เกณฑ์ฉบับนี้ จะเป็นไปในภาพรวมเพื่อประเมินผลิตภัณฑ์ที่มีการใช้ปัญญาประดิษฐ์ ในกรณีที่มีหน่วยงานกำกับดูแลเฉพาะของแต่ละภาคส่วน ซึ่งอาจกำหนดมาตรฐานที่เฉพาะเจาะจงที่ จำกัดหรือนอกเหนือจากเกณฑ์การตรวจประเมินนี้ ให้อ้างอิงกับข้อกำหนดของหน่วยงานกำกับดูแลเหล่านั้น ข้อเสนอแนะมาตรฐานฉบับนี้อ้างอิงข้อกำหนดเกี่ยวกับการทดสอบตามมาตรฐานสากล คือ ISO/IEC DIS 25059 , ISO/IEC 29119-11 :2020 , แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ (สวทช) และมาตรฐานการประเมิน ISO/IEC 25010 เป็นหลัก และนำข้อกำหนดดังกล่าวมาประยุกต์เป็นแนวทางการประเมินของประเทศไทยที่สอดคล้องกับมาตรฐานสากล

## ภาพรวมการประเมินคุณภาพผลิตภัณฑ์

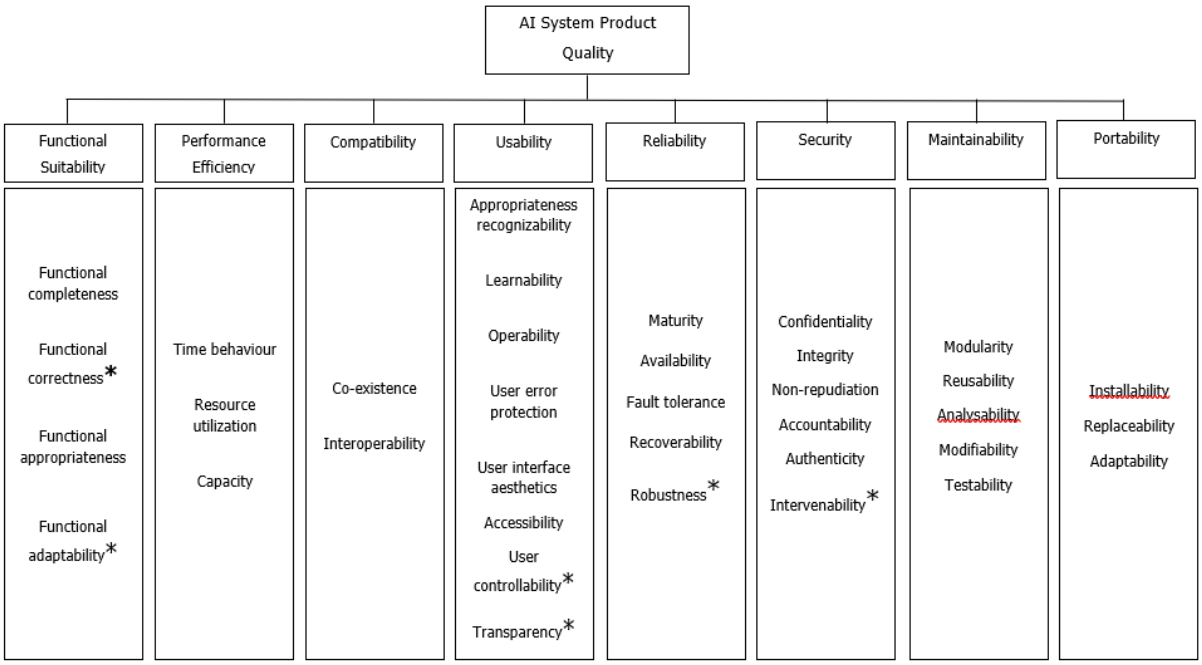
### ก. โมเดลคุณภาพผลิตภัณฑ์ (Product quality model)

การประเมินคุณภาพระบบคอมพิวเตอร์และซอฟต์แวร์แบบทั่วไปจะนำมาตรฐาน ISO/IEC 25000 – ISO/IEC 25040 ข้อกำหนดด้านระบบและคุณภาพซอฟต์แวร์และการประเมิน (Software Quality Requirements and Evaluation :SQuaRE) มาดำเนินการ โดยใช้โมเดลคุณภาพผลิตภัณฑ์ (Product Quality Model) ซึ่งกำหนดคุณลักษณะและคุณลักษณะย่อย (Characteristics / Sub Characteristics) ทั้ง 8 ด้าน ได้แก่ Functional Suitability, Performance efficiency, Compatibility, Usability, Reliability, Security, Maintainability, Portability ) ร่วมกับการใช้มาตรฐาน ISO/IEC/IEEE 29119 Series มาใช้ในการทดสอบนับตั้งแต่กระบวนการทดสอบ การบันทึกเอกสารทดสอบ เทคนิคที่ใช้ทดสอบ ไปจนถึงการรายงานผลการทดสอบ ในซอฟต์แวร์แบบทั่วไป

มาตรฐาน ISO/IEC DIS 25059 : 2022 ข้อกำหนดด้านระบบและคุณภาพซอฟต์แวร์และการประเมิน สำหรับระบบเทคโนโลยีปัญญาประดิษฐ์ ได้พัฒนาเพิ่มเติมจาก ISO/IEC 25010:2011 ในส่วนคุณลักษณะเฉพาะของ

ระบบเทคโนโลยีปัญญาประดิษฐ์ โดยเพิ่มคุณลักษณะย่อยใหม่ และคุณลักษณะย่อยที่ปรับปรุง ได้แก่ User Controllability, Functional correctness, Functional adaptability, Robustness, Transparency, Intervenability ดังแสดงในรูปที่ 1 โดยใช้สัญลักษณ์ * นี้

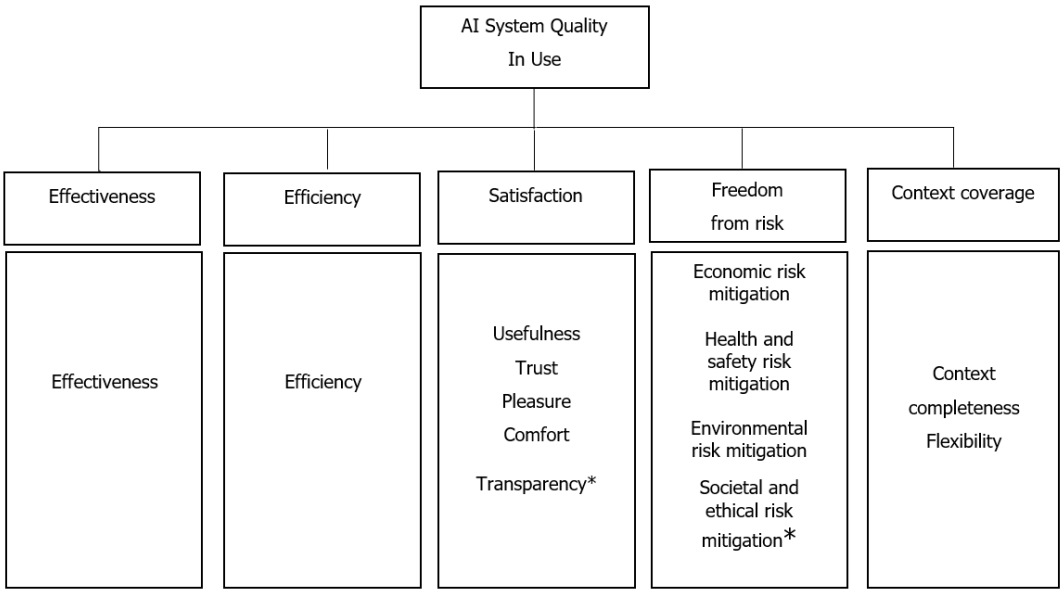
ทั้งนี้คุณสมบัติย่อยนี้จะถูกนำไปใช้การวัดในหัวข้อที่ 3




*รูป 8 AI System product quality model ISO/IEC DIS 25059*

### ข. โมเดลคุณภาพการใช้งาน (Quality in use model)

โมเดลคุณภาพการใช้งานของระบบเทคโนโลยีปัญญาประดิษฐ์ โดยอ้างอิงแบบจำลองคุณภาพทั่วไปใน ISO/IEC 25010 : 2011 ลักษณะย่อยใหม่แสดงในเครื่องหมาย * ได้แก่ Societal and ethical risk mitigation ดังแสดงในรูปที่ 9




*รูป 9 AI system quality in use model ISO/IEC DIS 25059*

แนวทางการลดความเสี่ยงทางสังคมและจริยธรรมเป็นคุณลักษณะย่อยที่เพิ่มขึ้นมา โดยในส่วนนี้จะเกี่ยวข้องกับกับมาตรฐานจริยธรรมด้านปัญญาประดิษฐ์ ISO/IEC 24368 :- ได้แก่ accountability, fairness and non-discrimination (Bias), transparency and explainability, professional responsibility, promotion of human values, privacy, human control of technology, community involvement and development, human centered design, respect for the rule of law, respect for international norms of behavior, labour practices ทั้งนี้ Transparency เป็นคุณลักษณะย่อยที่เพิ่มเข้ามาในส่วนของความพึงพอใจในการใช้ซึ่งมีความสอดคล้องกับโมเดลคุณภาพของผลิตภัณฑ์ข้างต้น

### ค. แนวทางการทดสอบระบบที่มีการใช้เทคโนโลยีปัญญาประดิษฐ์

ในเกณฑ์การตรวจประเมินฉบับนี้อ้างอิง มาตรฐาน ISO/IEC 29119-11 : 2020 แนวทางการทดสอบระบบที่มีการใช้เทคโนโลยีปัญญาประดิษฐ์ (ปัจจุบันมาตรฐานฉบับนี้อยู่ระหว่างปรับปรุง) โดยเกณฑ์ฉบับนี้ใช้นิยาม “AI-based system is a system that includes at least one AI component” ตามแนวทางมาตรฐานฉบับนี้ คุณลักษณะเฉพาะของเทคโนโลยีปัญญาประดิษฐ์มีทั้งในส่วน Functional และ Non-Functional เช่นเดียวกันกับที่ระบุไว้ในมาตรฐาน ISO/IEC 25010: 2011 แต่มีคุณลักษณะเฉพาะของเทคโนโลยีปัญญาประดิษฐ์ได้แก่ flexibility, adaptability, autonomy, evolution, bias, transparency/interpretability/explainability, complexity และ non-determinism เพิ่มขึ้นมา ซึ่งคุณลักษณะเชิงคุณภาพเหล่านี้สามารถใช้ทำเป็นรายการตรวจสอบตั้งต้นสำหรับวางแผนการทดสอบเพื่อ

ประเมินความเสี่ยง และระบุความเสี่ยงที่ต้องบรรเทาด้วยการทดสอบ คุณลักษณะเหล่านี้จะใช้ในการจัดทำเกณฑ์ในข้อ 3

### หลักสูตร Certified Tester AI Testing (CT-AI)

หลักสูตรใบรับรองมาตรฐานสากล การทดสอบระบบเทคโนโลยีปัญญาประดิษฐ์ เป็นหลักสูตรที่อยู่ภายใต้ International Software Testing Qualification Board :ISTQB มีเนื้อหาเพื่อสร้างความรู้ความเข้าใจเทคโนโลยีปัญญาประดิษฐ์ให้กับ tester ผู้สนใจและผู้เกี่ยวข้องในการทดสอบระบบ รวมถึงนำเสนอเทคนิคและวิธีการทดสอบ โดยในการพัฒนาเกณฑ์ในส่วนที่ 3 นี้ได้มีการอ้างอิงและประยุกต์ใช้ในส่วนของคุณลักษณะเฉพาะของระบบเทคโนโลยีปัญญาประดิษฐ์ ได้แก่ flexibility and adaptability, autonomy, evolution bias, ethics, side effects and reward hacking, transparency, interpretability, explainability safety and AI รวมถึงการกำหนดวัตถุประสงค์การทดสอบและเกณฑ์การทดสอบ ไปใช้อ้างอิงในการจัดทำเกณฑ์ในข้อ 3

## ง. แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์

ในประเทศไทยได้มีการจัดทำเอกสารแนวปฏิบัติจริยธรรมปัญญาประดิษฐ์ ซึ่งจัดทำโดยสำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ เพื่อให้ผู้กำกับดูแล ผู้วิจัย ผู้ออกแบบ ผู้พัฒนา ผู้ให้บริการปัญญาประดิษฐ์ และผู้ใช้ประโยชน์จากปัญญาประดิษฐ์ ใช้เป็นแนวทางสำหรับการกำกับการดำเนินงาน และให้ผู้รับบริการได้ทราบถึงสิทธิและตระหนักรู้ถึงความเสี่ยงของการใช้บริการปัญญาประดิษฐ์ หน่วยงานรัฐและหน่วยงานกำกับดูแลปัญญาประดิษฐ์ ทั้งระดับประเทศและระดับองค์กรเพื่อใช้เป็นแนวทางในการส่งเสริม สนับสนุน รวมถึงกำกับดูแลเทคโนโลยีปัญญาประดิษฐ์ เพื่อทำให้ปัญญาประดิษฐ์มีความน่าเชื่อถือ มั่นคงปลอดภัย มีจริยธรรม ส่งผลให้ได้รับการพัฒนาและใช้งาน ก่อให้เกิดประโยชน์กับมนุษย์ สังคมและสิ่งแวดล้อม ด้วยความโปร่งใส ครอบคลุมและเป็นธรรม สอดคล้องกับกฎหมายจริยธรรมและสิทธิมนุษยชน รวมถึงสอดคล้องกับยุทธศาสตร์ชชาติ 20 ปี (2561 – 2580) ยุทธศาสตร์ที่ 2 ที่ได้กล่าวถึงการใช้เทคโนโลยีดิจิทัล ข้อมูลและปัญญาประดิษฐ์ในการเพิ่มศักยภาพและความสามารถในการแข่งขันของอุตสาหกรรมและบริการ และยุทธศาสตร์กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม พ.ศ. 2563 – 2567 ยุทธศาสตร์ที่ 1 เรื่องการพัฒนาโครงสร้างพื้นฐานดิจิทัลของประเทศ ซึ่งมีกลยุทธ์ที่จะพัฒนาและส่งเสริมการลงทุนและการใช้ประโยชน์จากอินเทอร์เน็ตในทุกสิ่ง (IOT) และปัญญาประดิษฐ์ (AI) โดยผู้ที่ปฏิบัติตามแนวทางจริยธรรมปัญญาประดิษฐ์จะได้มีบริการทางด้านเทคโนโลยีดิจิทัลที่สอดคล้องกับแนวทางการ

พัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคมของประเทศไทย ส่งผลให้มีบริการปัญญาประดิษฐ์ที่มีประสิทธิภาพในทุกมิติ และสามารถให้บริการได้อย่างยั่งยืนได้

นอกจากเอกสารแนวปฏิบัติจริยธรรมปัญญาประดิษฐ์ ซึ่งจัดทำโดยสำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ, สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ กระทรวงการอุดมศึกษา วิทยาศาสตร์ และนวัตกรรม ยังได้จัดทำ “แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ขึ้นเพื่อใช้เป็นหลักปฏิบัติและแนวทางการดำเนินงานด้านปัญญาประดิษฐ์ ผู้ที่ร่วมวิจัยหรือรับทุนวิจัยของ สวทช. ภาคเอกชนที่ใช้พื้นที่ภายในอุทยานวิทยาศาสตร์ประเทศไทย และเขตนวัตกรรมระเบียงเศรษฐกิจพิเศษภาคตะวันออก (EECi) รวมถึงผู้รับจ้างช่วง โดยมีหลักการสำคัญคือ การสร้างความสมดุลระหว่างการพัฒนาปัญญาประดิษฐ์ที่รวดเร็ว กับการรู้เท่าทันการนำปัญญาประดิษฐ์ไปใช้ในทางที่ผิดและเป็นอันตรายต่อมนุษย์ ทั้งนี้ ข้อปฏิบัติภายในแนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ประกอบด้วยข้อปฏิบัติซึ่งมีระดับการบังคับที่เข้มงวดและผ่อนคลายแตกต่างกันไป เพื่อให้สอดคล้องกับระยะการวิจัยและพัฒนาปัญญาประดิษฐ์ เพื่อส่งเสริมให้เกิดการพัฒนาปัญญาประดิษฐ์ที่รวดเร็ว และไม่เป็นอุปสรรคต่อนักวิจัย

ทั้งนี้ คณะทำงานได้นำหลักการด้านจริยธรรมที่กล่าวถึงในข้างต้นมาแสดงไว้ในที่นี้ด้วย ดังนี้

**แนวปฏิบัติจริยธรรมปัญญาประดิษฐ์ ของ สดช.**

สดช. ได้เสนอหลักการทางจริยธรรมปัญญาประดิษฐ์ 6 ข้อ ได้แก่

**1. ความสามารถในการแข่งขันและการพัฒนาอย่างยั่งยืน (Competitiveness and Sustainability Development)**

- ปัญญาประดิษฐ์ควรถูกสร้างและใช้งานเพื่อสร้างประโยชน์และความผาสุกให้แก่มนุษย์ สังคม เศรษฐกิจและสิ่งแวดล้อมอย่างยั่งยืน
- ปัญญาประดิษฐ์ควรถูกใช้งานเพื่อเพิ่มความสามารถในการแข่งขันและสร้างความเจริญให้กับมนุษย์ สังคม ประเทศ ภูมิภาค และโลกอย่างเป็นธรรม
- ปัญญาประดิษฐ์ควรได้รับการวิจัยและพัฒนาอย่างต่อเนื่อง เพื่อให้มนุษย์เกิดการสร้างสรรค์นวัตกรรมและอุตสาหกรรมใหม่

**2. ความสอดคล้องกับกฎหมาย จริยธรรม และมาตรฐานสากล (Laws Ethics and International Standards)**

- ปัญญาประดิษฐ์ควรได้รับการวิจัย ออกแบบ พัฒนา ให้บริการ และใช้งาน สอดคล้องกับกฎหมาย บรรทัดฐาน จริยธรรม คุณธรรมของมนุษย์ และมาตรฐานสากล โดยเคารพต่อความเป็นส่วนตัว เกียรติ สิทธิเสรีภาพ และสิทธิมนุษยชน
- ออกแบบปัญญาประดิษฐ์ควรใช้หลักการมนุษย์เป็นศูนย์กลางและเป็นผู้ตัดสินใจ
- ปัญญาประดิษฐ์ไม่ควรถูกใช้ในการกำหนดชะตาชีวิตของมนุษย์

**3. ความโปร่งใสและภาระความรับผิดชอบ (Transparency and Accountability)**

- ปัญญาประดิษฐ์ควรได้รับการวิจัย ออกแบบ พัฒนา ให้บริการและใช้งาน ด้วยความโปร่งใส สามารถอธิบายและคาดการณ์ได้ รวมถึงสามารถตรวจสอบกิจกรรมต่างๆ ที่เกิดขึ้นย้อนหลังได้
- ปัญญาประดิษฐ์ควรมีความสามารถในการสืบย้อนกลับ (Traceability) เฝ้าระวัง ตรวจสอบความผิดปรกติและวินิจฉัยปัญหาความล้มเหลวได้ (Diagnosability) ได้
- ผู้วิจัย ผู้ออกแบบ ผู้พัฒนา ผู้ให้บริการและผู้ใช้งานปัญญาประดิษฐ์ ควรมีภาระความรับผิดชอบ (Accountability) ต่อผลกระทบที่เกิดขึ้นจากปัญญาประดิษฐ์ตามภาระหน้าที่ของตน

**4. ความมั่นคงปลอดภัยและความเป็นส่วนตัว (Security and Privacy)**

- ปัญญาประดิษฐ์ควรถูกสร้างเพื่อบริการ แต่ไม่ควรถูกใช้เพื่อหลอกลวง ต่อต้าน และคุกคามมนุษย์
- ปัญญาประดิษฐ์ควรได้รับการออกแบบโดยใช้หลักการป้องกันความเสี่ยง เพื่อป้องกันการโจมตีจากภัยคุกคาม เพื่อรักษาไว้ซึ่งความมั่นคงปลอดภัยของข้อมูลและระบบ รวมถึงการคุ้มครองข้อมูลส่วนบุคคล จริยธรรม และความปลอดภัยของชีวิตและสิ่งแวดล้อมภายนอกตลอดวัฏจักรชีวิตของระบบ มีความสามารถในการตรวจสอบ รายงานและตอบสนองเพื่อหลีกเลี่ยงหรือลดผลกระทบ
- ปัญญาประดิษฐ์ควรมีกลไกให้มนุษย์แทรกแซงระบบเพื่อควบคุมความเสี่ยงที่อาจมีผลกระทบกับมนุษย์ได้
- หน่วยงานรัฐควรวางแผนกำกับดูแลการพัฒนาและให้ความร่วมมือกับนานาชาติในการหลีกเลี่ยงการแข่งขันสร้างอาวุธอัตโนมัติจากปัญญาประดิษฐ์ที่ร้ายแรง

**5. ความเท่าเทียม หลากหลาย ครอบคลุม และเป็นธรรม (Fairness)**

- การออกแบบและพัฒนาปัญญาประดิษฐ์ควรคำนึงถึงความหลากหลาย หลีกเลี่ยงการผูกขาด ลดการแบ่งแยกและเอนเอียง เพื่อก่อให้เกิดประโยชน์ต่อผู้คนจำนวนมากเท่าที่จะทำได้ โดยเฉพาะกลุ่มคนผู้ด้อยโอกาสในสังคม (Diversity)
- การตัดสินใจที่เกี่ยวข้องกับวิจัย ออกแบบ พัฒนา ให้บริการ และใช้งานปัญญาประดิษฐ์ที่สำคัญควรสามารถพิสูจน์ถึงความเป็นธรรมได้ (Fairness)

**6. ความน่าเชื่อถือ (Reliability)**

- ปัญญาประดิษฐ์ควรได้รับการสนับสนุนให้มีความน่าเชื่อถือและความมั่นใจในการใช้งานต่อสาธารณะ
- ปัญญาประดิษฐ์ควรสามารถคาดการณ์ ตัดสินใจ และให้คำแนะนำได้อย่างแม่นยำถูกต้อง (Accuracy) สร้างผลลัพธ์ที่สามารถเชื่อถือได้และสร้างใหม่ได้เมื่อต้องการ (Reliability and Reproducibility)
- ปัญญาประดิษฐ์ควรมีการควบคุมคุณภาพและตรวจสอบความครบถ้วนสมบูรณ์ของข้อมูล (Quality and integrity of data)
- ปัญญาประดิษฐ์ควรมีกระบวนการและช่องทางรับผลสะท้อนกลับ (Feedback) จากผู้ใช้งานเพื่อให้ผู้ใช้งานสามารถแจ้งความต้องการเพิ่มเติม รับเรื่องร้องเรียน แจ้งปัญหาของระบบที่ตรวจสอบพบ และให้ข้อเสนอแนะได้โดยง่ายและรวดเร็ว

แนวปฏิบัติจริยธรรมปัญญาประดิษฐ์ข้างต้น ที่นำเสนอโดย สดช. เป็นไปในทำนองเดียวกับ แนวปฏิบัติด้านจริยธรรมปัญญาประดิษฐ์ที่ สวทช. กำหนดขึ้น ดังแสดงในหัวข้อ 2.5.2 ข้างล่างนี้

**แนวปฏิบัติด้านจริยธรรมปัญญาประดิษฐ์ของ สวทช.**

แนวปฏิบัติด้านจริยธรรมปัญญาประดิษฐ์ของ สวทช. ประกอบด้วยหลักการ 7 ข้อ ดังต่อไปนี้

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)

**หลักการที่ 1 ความเป็นส่วนตัว (privacy)**

ปัญญาประดิษฐ์ควรถูกออกแบบให้สามารถปกป้องความเป็นส่วนตัว และเคารพต่อสิทธิเสรีภาพของบุคคล การนำข้อมูลส่วนบุคคลของผู้ใดไปใช้งาน รวมถึงการเผยแพร่และใช้ประโยชน์ผลลัพธ์จากการประมวลผลและ

การตัดสินใจของปัญญาประดิษฐ์ ต้องแจ้งให้ผู้ใช้งานปัญญาประดิษฐ์ทราบล่วงหน้าถึงข้อมูลที่จะถูกเก็บรวบรวมและลักษณะการนำข้อมูลดังกล่าวไปใช้ ซึ่งต้องได้รับการยินยอมจากเจ้าของข้อมูลก่อน การออกแบบและพัฒนาปัญญาประดิษฐ์ควรอยู่บนพื้นฐานของการปกป้องสิทธิเสรีภาพและศักดิ์ศรีของมนุษย์ให้คงอยู่ และต้องเคารพต่อกฎ ระเบียบ หรือกฎหมายภายในประเทศที่เกี่ยวข้องกับการคุ้มครองข้อมูลส่วนบุคคล เช่น พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 เป็นต้น นอกจากนี้ ยังรวมถึงการเคารพต่อกฎหมายระหว่างประเทศ เช่น General Data Protection Regulation (GDPR) ด้วย

**หลักการที่ 2 ความมั่นคงและปลอดภัย (security and safety)**

ปัญญาประดิษฐ์ควรถูกสร้างให้มีความมั่นคงและปลอดภัย รวมถึงป้องกันภัยอันตรายที่อาจเกิดขึ้นต่อมนุษย์ สิ่งแวดล้อม สังคม เศรษฐกิจ และประเทศ จากการใช้งานปัญญาประดิษฐ์ที่มากจนเกินไป (overused) และที่ไม่ถูกต้องเหมาะสม (misused) โดยการใช้งานและการตัดสินใจของปัญญาประดิษฐ์ควรเกิดจากความตั้งใจของมนุษย์ หรือมีกลไกให้มนุษย์สามารถแทรกแซงการดำเนินการต่าง ๆ เพื่อลดความเสี่ยงจากอันตรายที่อาจเกิดขึ้น เช่น การตัดสินใจที่ผิดพลาด การคุกคามจากผู้ไม่ประสงค์ดี และการนำไปใช้ในทางที่ผิด ซึ่งรวมถึงการใช้เสมือนเป็นอาวุธ (weaponization) และการทำให้เข้าใจผิด หรือให้ข้อมูลที่นำไปสู่การเข้าใจผิด (misinformation)

เนื่องจากในการวิจัย ออกแบบ และพัฒนาปัญญาประดิษฐ์ อาจใช้ชุดข้อมูลที่อ่อนไหว ชุดข้อมูลที่เป็นความลับ หรือชุดข้อมูลที่เป็นข้อมูลส่วนบุคคล ซึ่งหากถูกเข้าถึงโดยผู้ไม่ประสงค์ดีแล้วก็อาจนำมาซึ่งความเสียหายต่อผู้มีส่วนเกี่ยวข้องได้ จึงควรคำนึงถึงหลักการรักษาความมั่นคงปลอดภัยและการคุ้มครองข้อมูลส่วนบุคคลด้วยเช่นกัน

**หลักการที่ 3 ความไว้วางใจ (reliability)**

ผู้วิจัย ออกแบบ หรือพัฒนาปัญญาประดิษฐ์จะต้องสามารถสร้างความไว้ใจและความเชื่อมั่นในการใช้งานปัญญาประดิษฐ์ ซึ่งยังไม่ทราบผลกระทบจากการตัดสินใจของระบบที่จะเกิดขึ้นได้แน่ชัดให้แก่สาธารณชนได้ โดยการวิจัยและพัฒนาระบบการวิเคราะห์ ประมวลผล และตัดสินใจของปัญญาประดิษฐ์ให้แม่นยำถูกต้อง สร้างผลลัพธ์ที่เชื่อถือได้ และสร้างผลลัพธ์แบบเดียวกันใหม่ได้ (reproducible) รวมถึงมีการควบคุมคุณภาพของข้อมูลที่นำมาใช้งาน เพื่อป้องกันไม่ให้เกิดการตัดสินใจที่ผิดพลาด ซึ่งอาจส่งผลต่อความน่าเชื่อถือของระบบปัญญาประดิษฐ์ได้

**หลักการที่ 4. ความเป็นธรรม เท่าเทียม และไม่แบ่งแยก (fairness and non-discrimination)**

ปัญญาประดิษฐ์ควรถูกออกแบบ และนำไปใช้งานเพื่อส่งเสริมความเป็นธรรม ความเท่าเทียม ความหลากหลาย ความสามัคคี และความยุติธรรมของคนทุกกลุ่มในสังคม โดยที่ไม่เกิดความอคติ หรือความเอนเอียงใด ๆ รวมถึงการให้โอกาสประชาชนทุกคนในสังคม เช่น กลุ่มคนด้อยโอกาส ผู้พิการและผู้ทุพพลภาพ ให้ได้รับประโยชน์จากปัญญาประดิษฐ์ได้อย่างทั่วถึงและเท่าเทียม โดยไม่แบ่งแยกเชื้อชาติ สีผิวและไม่ก่อให้เกิดความเหลื่อมล้ำทางสังคม

**หลักการที่ 5 ความโปร่งใสและอธิบายได้ (transparency and explainability)**

ปัญญาประดิษฐ์ควรได้รับการออกแบบและนำไปใช้ โดยให้มนุษย์สามารถรู้ได้ว่ากำลังใช้ปัญญาประดิษฐ์อยู่ เข้าใจได้ว่าข้อมูลถูกนำไปใช้อย่างไร เข้าใจได้ถึงกระบวนการการตัดสินใจการคาดการณ์การกระทำต่าง ๆ ได้ และกำกับดูแลและตรวจสอบปัญญาประดิษฐ์ได้ โดยควรทำให้มีการแปลผลการดำเนินการของระบบให้เป็นข้อมูลที่สามารถอธิบายและเข้าใจได้โดยมนุษย์ สามารถตรวจสอบย้อนกลับไปยังแหล่งที่มาของชุดข้อมูลที่ได้รับ กระบวนการทำงาน และการตัดสินใจของปัญญาประดิษฐ์รวมถึงสถานที่ เวลา และวิธีการนำปัญญาประดิษฐ์ไปใช้ เพื่อใช้เฝ้าระวัง ตรวจสอบความผิดปกติ วินิจฉัยปัญหาและหาผู้รับผิดชอบได้อย่างมีประสิทธิภาพ อีกทั้งเพื่อช่วยสร้างความเชื่อมั่นให้แก่ผู้ใช้งานปัญญาประดิษฐ์

การสื่อสารคำอธิบายผลการดำเนินการ ความสามารถ และข้อจำกัดของปัญญาประดิษฐ์ ควรทำอย่างทันการณ์และเหมาะสมกับระดับความเชี่ยวชาญของผู้ที่ต้องการข้อมูลดังกล่าว

**หลักการที่ 6 ภาระความรับผิด (accountability)**

การวิจัย ออกแบบ หรือพัฒนาปัญญาประดิษฐ์ ต้องมีกลไกที่ทำให้เกิดความมั่นใจถึงการรับผิดชอบต่อผลกระทบที่เกิดจากปัญญาประดิษฐ์ของผู้ที่มีส่วนร่วมในการวิจัย ออกแบบ พัฒนา และนำไปใช้งานซึ่งต้องสามารถตรวจสอบย้อนกลับถึงผู้รับผิดชอบได้โดยชัดเจน รวมถึงมีกลไกแก้ไขปัญหา หรือรับผิดชอบต่อความเสียหายที่อาจเกิดขึ้น ตามภาระหน้าที่ของตนได้อย่างเพียงพออีกทั้งควรตระหนักถึงบทบาทสำคัญของผู้มีส่วนเกี่ยวข้องทุกคนที่มีส่วนร่วมในการออกแบบพัฒนา และนำไปใช้งาน ซึ่งผู้ที่มีส่วนได้ส่วนเสียเหล่านั้น จะต้องมีการปรึกษาร่วมกันเกี่ยวกับการทำงานของปัญญาประดิษฐ์อย่างเหมาะสม รวมถึงมีการวางแผนถึงการบริหารจัดการความเสี่ยง หรือผลกระทบในระยะยาวที่อาจเกิดขึ้น

**หลักการที่ 7 มนุษย์เป็นผู้ควบคุมปัญญาประดิษฐ์ เพื่อความยั่งยืนของมนุษยชาติ (human oversight and human agency)**

ในการออกแบบและการนำปัญญาประดิษฐ์ไปใช้งานนั้น ควรกำหนดให้คำนึงถึงมนุษย์เป็นหลัก (human centric) และคงไว้ซึ่งความสามารถในการควบคุมปัญญาประดิษฐ์ และสิทธิให้มนุษย์เป็นผู้ตัดสินใจในขั้นตอนการตัดสินใจที่เป็นกระบวนการสำคัญ

นอกจากนี้ ระบบปัญญาประดิษฐ์จะต้องถูกออกแบบและนำมาใช้เพื่อให้เกิดประโยชน์ต่อมนุษยชาติ ไม่ก่อให้เกิดอันตรายต่อมนุษย์ คำนึงถึงความเป็นอยู่ที่ดี และส่งเสริมคุณค่าของมนุษย์ รวมถึงเป็นการนำปัญญาประดิษฐ์มาใช้เพื่อให้เกิดการพัฒนาที่ยั่งยืน ทั้งต่อสังคมและสิ่งแวดล้อม ตลอดจนทำให้เกิดการพัฒนาในด้านอื่นๆ ต่อไป

**จ. เกณฑ์การประเมินและการทดสอบผลิตภัณฑ์**

การทดสอบระบบเทคโนโลยีปัญญาประดิษฐ์นอกจากจะดำเนินการทดสอบตามหน้าที่การทำงานของระบบ คุณลักษณะผลิตภัณฑ์แบบทั่วไปแล้ว ยังจำเป็นที่ต้องทดสอบคุณลักษณะเฉพาะของระบบดังนี้

**เกณฑ์การประเมินและทดสอบคุณลักษณะเฉพาะระบบเทคโนโลยีปัญญาประดิษฐ์**

*ตาราง 1 เกณฑ์การประเมินและทดสอบคุณลักษณะเฉพาะระบบเทคโนโลยีสารสนเทศ*

| | **หัวข้อ** | **เกณฑ์** | **เกณฑ์การตัดสิน** | **มาตรฐานอ้างอิง** | **คำอธิบายเพิ่มเติม** |
|---|---|---|---|---|---|
| C01 | User Controllability | ตรวจสอบว่าระบบสามารถควบคุมไม่ให้มนุษย์หรือระบบภายนอกอื่นๆสามารถเข้ามาแทรกแซงการทำงานได้ | -ดำเนินการทดสอบผ่านเงื่อนไข และกระบวนการตรวจสอบเอกสาร | ISO/IEC DIS 25059:2022<br>ISO/IEC 22989 | |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| C02 | *Functional Adaptability | - ตรวจสอบว่าระบบยังทำงานตาม functional requirements ได้อย่างถูกต้องและตรงตาม non-functional requirements เมื่อปรับให้เข้ากับสภาพแวดล้อมใหม่<br><br>- ตรวจสอบเวลาที่ระบบใช้ในการปรับตัวให้เข้ากับการเปลี่ยนแปลงในสภาพแวดล้อมใหม่<br><br>- ตรวจสอบทรัพยากรที่ใช้เมื่อปรับให้เข้ากับการเปลี่ยนแปลงในสภาพแวดล้อมใหม่ | เป็นไปตามที่ระบุไว้ใน requirements specification | ISO/IEC DIS 25059:2022<br><br>ISO 29119-11<br><br>ISTQB AI Testing | - Adaptability ถือเป็นคุณสมบัติที่ต้องดำเนินการได้โดยง่ายในการปรับตัวให้เข้ากับคุณสมบัติใหม่ เช่น การเปลี่ยนแปลง ฮาร์ดแวร์ หรือการเปลี่ยนสภาพแวดล้อมในการทำงาน<br><br>- การทดสอบอาจดำเนินการรูปแบบของ regression testing |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| C03 | *Flexibility | - พิจารณาว่าระบบจัดการอย่างไรในบริบทที่แตกต่างจากค่าเริ่มต้นที่กำหนดไว้ใน requirements specification<br><br>- ตรวจสอบเวลาที่ระบบใช้ และ/หรือ ทรัพยากรที่ใช้ในการเปลี่ยนแปลงตัวเองเพื่อจัดการบริบทใหม่ | เป็นไปตามที่ระบุไว้ใน requirements specification | ISO 29119-11<br><br>ISTQB AI Testing | - Flexibility คือความยืดหยุ่นของระบบในการใช้งานในบริบทที่แตกต่างจาก ต้นฉบับ Requirements specification |
| C04 | *Evolution | - ตรวจสอบว่าระบบเรียนรู้จากประสบการณ์ของตัวเองได้ดีเพียงใด<br><br>(Self-learning AI-based systems)<br><br>- ตรวจสอบว่าระบบจัดการได้ดีเพียงใดเมื่อมีการเปลี่ยนชุดข้อมูล | เป็นไปตามที่ระบุไว้ใน requirements specification | ISO 29119-11<br><br>ISTQB AI Testing | Evolution คือ ความสามารถของระบบในการปรับปรุงตัวเองเพื่อตอบสนองต่อการเปลี่ยนแปลงข้อจำกัดภายนอก เช่น ระบบ Self-learning และ Self-Learning AI-based |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| | | | | | system จำเป็นต้องระบุคุณสมบัตินี้ไว้ด้วย |
| C05 | *Autonomy | - ตรวจสอบว่าระบบตอบสนองอย่างไรต่อการถูกกระทำ (Force) จากสภาพแวดล้อมที่แตกต่าง ในขณะที่การดำเนินงานในส่วนนี้ต้องทำงานแบบอัตโนมัติ<br><br>- ตรวจสอบว่าระบบสามารถถูกชักชวนให้ร้องขอการแทรกแซงของมนุษย์ได้หรือไม่เมื่อระบบควรจะเป็นระบบอัตโนมัติ | เป็นไปตามที่ระบุไว้ใน Requirements specification | ISO 29119-11<br><br>ISTQB AI Testing | Autonomy คือ ความสามารถของระบบในการทำงานได้อย่างอิสระจากการควบคุมของมนุษย์ตามระยะเวลาที่กำหนด |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| CO6 | Functional correctness | -ทดสอบความสามารถของผลิตภัณฑ์ว่าสามารถทำงานได้อย่างถูกต้องในระดับความแม่นยำที่ต้องการ | เป็นไปตามที่ระบุไว้ใน Requirements specification | ISO/IEC DIS 25059:2022 | |
| C07 | Robustness | -ทดสอบว่าโมเดลยังทำงานได้ถูกต้องเมื่อมีการใส่ข้อมูลที่ไม่ถูกต้อง (ตัดออก)<br><br>- ทดสอบประสิทธิภาพของโมเดลด้วยวิธีการ Statistical approaches Mathematical metrics หรือ formal method หรือ empirical method<br><br>- ตรวจสอบกระบวนการ และหลักฐานการรีวิวเอกสารที่มีผลกระทบต่อประสิทธิภาพของโมเดล | ประเมินผลตามผลการทดสอบและตรวจสอบเอกสารหลักฐาน | ISO/IEC DIS 25059:2022<br><br>ISO/IEC TR 24029-1 | |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| C08 | Transparency | -ตรวจสอบความโปร่งใสโดยการดูความสะดวกในการเข้าถึงอัลกอริทึมและชุดข้อมูล<br><br>- ประเมินผ่านกระบวนการตรวจสอบเอกสารหลักฐาน ในการให้ข้อมูลที่เหมาะสมแก่ผู้เกี่ยวข้องที่อาจได้รับผลกระทบจากระบบ เช่นข้อมูลด้านความปลอดภัย การใช้เทคโนโลยีปัญญาประดิษฐ์ในระบบ เป้าประสงค์ในการใช้งาน ข้อจำกัด การประเมินความเสี่ยง | - ดำเนินการทดสอบผ่านเงื่อนไข และกระบวนการตรวจสอบเอกสาร | ISO/IEC DIS 25059:2022<br><br>ISO 29119-11<br><br>ISTQB AI Testing | -ความโปร่งใสพิจารณาจากความสะดวกในการเข้าถึงอัลกอริทึมและชุดข้อมูล<br><br>-การให้ข้อมูลที่จำเป็นต่อผู้ใช้งาน |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| C09 | Explainability | - ระบุปัจจัยที่มีผลกระทบกับผลลัพธ์ของเทคโนโลยีปัญญาประดิษฐ์<br><br>- ตรวจสอบเอกสารหลักฐาน ที่อธิบายระบบ โดยให้เหตุผลในการเลือกใช้โมเดล การประเมินความเสี่ยง ทดสอบ<br><br>โดยตั้งคำถามกับผู้ใช้ระบบ หรือผู้ที่มีพื้นฐานใกล้เคียงกัน<br><br>- continuity<br><br>- consistency<br><br>- selectivity | ประเมินผ่านการทดสอบและการตรวจสอบเอกสารหลักฐาน | ISO/IEC DIS 25059:2022<br><br>ISO 29119-11<br><br>ISTQB AI Testing<br><br>ISO/IEC TR 24028: 2020<br><br>ข้อ 9.3.7 | การทำให้ผู้ใช้สามารถเข้าใจได้ว่าระบบให้ผลลัพธ์อย่างไรต่อผู้ใช้งาน |
| C10 | Bias/Fairness | -ทดสอบว่าโมเดล AI ไม่มีคุณสมบัติที่ก่อให้เกิดการอคติ โดยการทดสอบ | กำหนดเกณฑ์การผ่าน ไม่ผ่าน โดยดูเปอร์เซ็นต์ | ISO/IEC DIS 25059:2022 | Bias ในความหมายของระบบที่ใช้เทคโนโลยี |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| | | ผลลัพธ์ของโมเดลกับข้อมูลทดสอบ (Ground truth)<br><br>- เปรียบเทียบผลการทดสอบโดยใช้ข้อมูลภายนอก เพื่อตรวจสอบอคติที่ไม่ต้องการในตัวแปรอนุมาน (external validity testing)<br><br>- ใช้ Statistical bias metrics, fairness metrics, algorithmic fairness metrics ฯ. ในการทดสอบ | ความถูกต้องตาม metrics ที่นำมาใช้ | ISO 29119-11<br><br>ISTQB AI Testing<br><br>ISO/IEC TR 24027: 2021 | ปัญญาประดิษฐ์หมายถึงการวัดค่าทางสถิติระหว่างลัพธ์ที่ได้จากระบบกับผลลัพธ์ที่เป็นค่าที่ควรจะเป็นหรือมีเกณฑ์ค่าที่ควรจะเป็น ซึ่งไม่ลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง Bias สามารถเป็นได้ในเชิงคุณลักษณะต่างๆ เช่น เพศ เชื้อชาติ ชาติพันธ์ รสนิยมทางเพศ ระดับรายได้ อายุ เป็นต้น |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| C11 | Intervenability<br><br>(Safety and AI) | -ตรวจสอบว่าระบบจะไม่ก่อให้เกิดอันตรายต่อผู้ใช้งาน ทรัพย์สิน หรือสภาพแวดล้อม<br><br>-ตรวจสอบว่าระบบพยายามทำให้เกิดความเสียหายต่อตัวระบบเองหรือไม่ | ประเมินผลผ่านการทดสอบและตรวจสอบเอกสารหลักฐาน | ISO/IEC DIS 25059:2022<br><br>ISTQB AI Testing | |
| C12 | Societal and ethical risk mitigation | ตรวจสอบหลักฐานการประเมินความเสี่ยงและบรรเทาความเสี่ยงในหัวข้อดังนี้ accountability, fairness and non-discrimination (Bias), transparency and explain ability, professional responsibility, promotion of human values, privacy, human control of technology, community involvement and development, human centered design, respect for the rule of law, | ตรวจสอบหลักฐานการประเมินความเสี่ยงและบรรเทา | ISO/IEC DIS 25059:2022<br><br>ISO/IEC TR 24368: 2022 | |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| | | respect for international norms of behavior, labour practices | | | |
| C12 | Side-effects | ระบุผลข้างเคียงที่อาจเป็นอันตรายและพยายามสร้างการทดสอบ<br><br>ที่ทำให้ระบบแสดงผลข้างเคียงเหล่านี้ | ดำเนินการผ่านการทดสอบ | ISTQB AI Testing | |
| C13 | Reward Hacking | Independent tests can identify reward hacking when these tests use a different means of measuring success compared to the intelligent agent<br><br>being tested. | | ISTQB AI Testing | |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| C14 | Repeatability/Reproducibility | ประเมินเอกสารหลักฐานที่มาของโมเดล แหล่งที่มาของข้อมูล การใช้เครื่องมือในการควบคุมเวอร์ชั่น (Version Control) | ประเมินผ่านการตรวจสอบกระบวนการ | | |
| C15 | Human Agency and Oversight | ได้รับการออกแบบในลักษณะที่จะไม่ลดความสามารถของมนุษย์ในการตัดสินใจหรือควบคุมระบบ ซึ่งรวมถึงการกำหนดบทบาทของมนุษย์ในการกำกับดูแลและควบคุมระบบ AI เช่น human-in-the-loop, human-over-the-loop หรือ human-out-of-the-loop | ประเมินผลผ่านประเมินผ่านกระบวนการตรวจสอบเอกสารหลักฐาน | | |

**หมายเหตุ**

*** Adaptability, Flexibility, Evaluation, Autonomy คุณสมบัติของผลิตภัณฑ์ที่ควรต้องระบุไว้ใน Requirements Specification**

**1.1 เกณฑ์การประเมินและทดสอบจริยธรรมการวิจัย (Checklist & Criteria)**

หลักการจริยธรรมงานวิจัยที่ประมวลจาก สวทช. และ สดช. ช้เป็นเกณฑ์ประเมินได้ดังต่อไปนี้

*ตาราง 2 แสดงเกณฑ์ประเมินและการทดสอบจริยธรรมการวิจัย*

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| 1 | ความเป็นส่วนตัว (Privacy) | - คุ้มครองข้อมูลส่วนบุคคล<br>- เคารพสิทธิเสรีภาพ<br>-ผู้ใช้งานปัญญาประดิษฐ์ทราบล่วงหน้าถึงข้อมูลที่จะถูกเก็บรวบรวมและลักษณะการนำข้อมูลไปใช้ | -ตรวจสอบจากเอกสารที่เกี่ยวข้อง<br>-พิจารณาจาก data และ algorithms | ISO/IEC 25059 | (ถ้ามี) |
| 2 | ความมั่นคงและปลอดภัย (Security & Safety) | -ไม่หลอกลวงและคุกคามมนุษย์<br>-มีกลไกเพื่อควบคุมความเสี่ยง<br>-ไม่สร้างความเสียหายแก่บุคคลและหน่วยงานที่เกี่ยวข้อง | -ตรวจสอบจากเอกสารที่เกี่ยวข้อง<br>-พิจารณาจาก data และ algorithms | ISO/IEC 25059 | (ถ้ามี) |
| 3 | ความไว้วางใจ (Reliability) | -มีการควบคุมคุณภาพของข้อมูลที่ใช้<br>-สร้างผลลัพธ์ที่เชื่อถือได้<br>-สร้างผลลัพธ์แบบเดียวกันใหม่ได้ (reproducible | -ตรวจสอบจากเอกสารที่เกี่ยวข้อง<br>-พิจารณาจาก data และ algorithms | ISO/IEC 25059 | (ถ้ามี) |

| | หัวข้อ | เกณฑ์ | เกณฑ์การตัดสิน | มาตรฐานอ้างอิง | คำอธิบายเพิ่มเติม |
|---|---|---|---|---|---|
| | | -มีกระบวนการและช่องทางรับ feedback | | | |
| 4 | ความเป็นธรรมและเท่าเทียม (Fairness & Nondiscrimination) | -ไม่อคติหรือเอนเอียงไปทางใดทางหนึ่ง | -ตรวจสอบจากเอกสารที่เกี่ยวข้อง<br>-พิจารณาจาก data และ algorithms | ISO/IEC 25059 | (ถ้ามี) |
| 5 | ความโปร่งใส (Transparency) | -ให้ผู้ที่เกี่ยวข้องทราบว่ากำลังใช้ปัญญาประดิษฐ์<br>-สามารถตรวจสอบย้อนกลับได้<br>-ตรวจสอบความผิดปกติ และวินิจฉัยปัญหาความล้มเหลวได้ | -ตรวจสอบจากเอกสารที่เกี่ยวข้อง<br>-พิจารณาจาก data และ algorithms | ISO/IEC 25059 | (ถ้ามี) |
| 6 | สามารถอธิบายได้ (Explanabilty) | -สื่อสารการอธิบายผลการดำเนินการ ความสามารถและข้อจำกัดได้ | -ตรวจสอบจากเอกสารที่เกี่ยวข้อง<br>-พิจารณาจาก data และ algorithms | ISO/IEC 25059 | (ถ้ามี) |

ทั้งนี้เกณฑ์เกณฑ์การประเมินและทดสอบคุณลักษณะเฉพาะระบบเทคโนโลยีปัญญาประดิษฐ์ และเกณฑ์การประเมินและทดสอบจริยธรรมการวิจัยจะถูกนำไปใช้ในกระบวนการทดสอบของ AI testing Center ต่อไป

---

**บรรณานุกรม**

**ภาษาไทย**


คณะกรรมการจริยธรรมปัญญาประดิษฐ์ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ. (2565) แนวปฏิบัติจริยธรรมด้านปัญญาประดิษฐ์ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ.

สุภาภรณ์ เกียรติสิน, มนัสศิริ จันสุทธิวรางกูร, ปรัชญ์ สง่างาม และ ยุทธพงศ์ อุณหทวีทรัพย์. (2564) เอกสารแนวปฏิบัติจริยธรรมปัญญาประดิษฐ์ (Thailand AI Ethics Guideline). สำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ, กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม


**ภาษาอังกฤษ**


European Commission. Ethics Guidelines for Trustworthy AI. Apr 8, 2019. https://www.aepd.es/sites/default/files/2019-12/ai-ethics-

International Software Testing Qualifications Board. Certified Tester AI Testing (CT-AI) Syllabus. https://isqi.org/en/88-istqb-certified-tester-ai-testing-ct-ai.html

ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML). https://www.iso.org/standard/74438.html

ISO/IEC 25010:2011(en) Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en

ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations. https://www.iso.org/standard/56641.html

ISO/IEC DIS 25059 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems. https://www.iso.org/standard/80655.html

ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. https://www.iso.org/standard/77608.html

ISO/IEC TR 24030:2021 Information technology — Artificial intelligence (AI) — Use cases. https://www.iso.org/standard/77610.html

ISO/IEC TR 24372:2021 Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems. https://www.iso.org/standard/78508.html

ISO/IEC TR 29119-11:2020 Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems. https://www.iso.org/standard/79016.html

The Info-communication Media Development Authority. Invitation to Pilot AI Verify: AI Governance Testing Framework & Toolkit. May 25, 2022, https://file.go.gov.sg/aiverify.pdf

---