SCBX AI Outlook 2025

Beaconing the Future of Artificial Intelligence.

SCBX

SCBX AI Outlook 2025:
Beaconing the Future of Artificial Intelligence

ปลายปี 2022 หลังจากที่ OpenAI เปิดตัว ChatGPT ทำให้ทั้งโลกได้รู้จักกับ Generative AI หรือ Gen AI โลกทั้งโลกก็ไม่เหมือนเดิมอีกต่อไป

AI เป็นเรื่องที่ใกล้ตัวผู้คนมากขึ้น ในฐานะ 'ตัวเปลี่ยนเกม' ทั้งในระดับบุคคลและองค์กร ที่ทำให้ชีวิตง่ายขึ้นไปจนถึงยกระดับคุณภาพงานในเชิงธุรกิจจนหลาย ๆ องค์กรต่างหารูปแบบการใช้และการพัฒนา AI เพื่อสร้างความได้เปรียบในการแข่งขันให้ได้มากที่สุด

SCBX AI Outlook 2025: Beaconing the Future of Artificial Intelligence จึงเป็นรายงานที่จะพาผู้อ่านไปสำรวจเทรนด์ AI ที่จะเกิดขึ้นในปี 2025 อันประกอบไปด้วย 4 เทรนด์สำคัญ ได้แก่

ACT I: Two Philosophies, One Future.
The Battle Between Open-Source and Closed-Source AI Intensifies

ACT II: Tiny Titans - Small, but Mighty.
More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

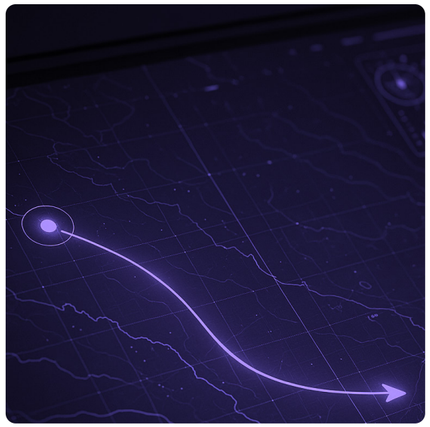
ACT III: AI at Your Fingertips.
Agentic AI: Rise of the Agents

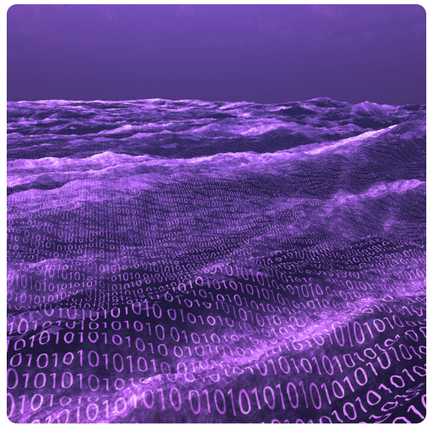
ACT IV: Not Quite Human, But Almost There.
Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

และส่งท้ายด้วย EPILOGUE: The AI storm - Infinite impact. Case Studies from Inside the Eye of the Typhoon กรณีศึกษาของ SCBX ในการใช้เครื่องมือ AI 'Typhoon' ในธุรกิจต่าง ๆ

เทรนด์และกรณีศึกษาในรายงานฉบับนี้จะช่วยให้เห็นภาพในอนาคตในการวางกลยุทธ์ต่าง ๆ เพื่อโอกาสทางธุรกิจในโลกที่โอบรับ AI มากขึ้นเรื่อย ๆ

Contents





ACT I:
Two Philosophies, One Future.

The Battle Between Open-Source and Closed-Source AI Intensifies

การพัฒนาโมเดลปัญญาประดิษฐ์ (AI) ระดับสูงมีอยู่สองแนวทางหลักคือ 'แบบเปิด' (Open Source) และ 'แบบปิด' (Closed Source)

โมเดลแบบเปิดคือโมเดลที่เปิดเผยโค้ดและองค์ความรู้ให้สาธารณชนทุกคนสามารถนำไปใช้ ปรับปรุง หรือดัดแปลงได้อย่างเสรี ในขณะที่โมเดลแบบปิดเป็นโมเดลที่พัฒนาและควบคุมโดยองค์กรใดองค์กรหนึ่ง โค้ดและข้อมูลถูกเก็บเป็นความลับไม่เปิดเผยให้คนนอกใช้โดยตรง

โมเดลปิดที่มีชื่อเสียงเช่น ChatGPT ของ OpenAI หรือ Claude ของ Anthropic มักมีสมรรถนะสูงแต่ก็เป็น "กล่องดำ" ที่ผู้ใช้ไม่ทราบรายละเอียดภายใน ในทางกลับกัน โมเดลเปิดชื่อดังอย่าง DeepSeek R1 จาก DeepSeek หรือ Llama จาก Meta เปิดโอกาสให้ชุมชนนักพัฒนามีส่วนร่วม จนเกิดการพัฒนาต่อยอดอย่างรวดเร็ว

"Open VS. Closed-Source AI Models" เป็นหัวข้อถกเถียงสำคัญที่ส่งผลต่อทิศทางวงการ AI แต่ละแบบก็มีข้อดีข้อเสียที่แตกต่างกัน ในขณะที่ศักยภาพของโมเดลแบบเปิดที่เริ่มทัดเทียมแบบปิดได้มากขึ้นเรื่อย ๆ ส่งผลกระทบต่ออุตสาหกรรมและเศรษฐกิจ รวมถึงการนำไปปรับใช้ให้เหมาะสมในภาคธุรกิจก็เป็นเรื่องสำคัญไม่แพ้กัน [1]

ผลกระทบและการเปลี่ยนแปลงจากโมเดลแบบเปิดและแบบปิด

โมเดลแบบปิดซึ่งพัฒนาโดยบริษัทยักษ์ใหญ่ เช่น OpenAI, Google และ Anthropic ได้ขับเคลื่อนความก้าวหน้าของ AI ในระยะแรก (ตั้งแต่ช่วงปี 2018 ที่ GPT-1 ถูกปล่อยออกมา) และสร้างความตื่นตัวในวงกว้าง ตัวอย่างเช่น ChatGPT ของ OpenAI (ขับเคลื่อนด้วยโมเดลปิดอย่าง GPT-3.5 และ GPT-4) ได้แสดงให้โลกเห็นถึงศักยภาพอันน่าทึ่งของ AI ในการสนทนาและสร้างเนื้อหา ซึ่งกระตุ้นให้หลายองค์กรเริ่มนำ AI มาใช้งานจริงจังมากขึ้น [2]

อย่างไรก็ตามการที่โมเดลเหล่านี้เป็นระบบปิด "กล่องดำ" ก็ก่อให้เกิดข้อกังวลด้านความโปร่งใสและความปลอดภัยของข้อมูล หลายองค์กรลังเลที่จะส่งข้อมูลสำคัญเข้าสู่ระบบที่ตนไม่สามารถควบคุมได้ ก็สะท้อนผลกระทบด้านลบที่ตามมาจากการพึ่งพา 'โมเดลปิด' หากโมเดลนั้นไม่มีมาตรการควบคุมภายในที่ดีมากพอ

ในทางกลับกัน โมเดล 'แบบเปิด' ได้เข้ามาเปลี่ยนสมการของวงการ AI อย่างมีนัยสำคัญ เมื่อเปิดให้สาธารณะเข้าถึงและพัฒนาต่อยอดได้ ศักยภาพจากการร่วมมือของชุมชนทำให้โมเดล AI ใหม่ ๆ ถือกำเนิดและพัฒนาอย่างรวดเร็ว [3] [4]

ปรากฏการณ์นี้เห็นชัดเจนมากขึ้นเรื่อย ๆ จากรายงานของ Stanford Institute for Human-Centered Artificial Intelligence (HAI) บอกว่า ในช่วงปี 2023 มีการเปิดตัวโมเดลพื้นฐาน (foundation model) จำนวนมากเป็นประวัติการณ์ ถึง 149 โมเดล ภายในปีเดียว ซึ่งในจำนวนนี้กว่า

	จากสัดส่วน	และเพียง
65.7%	**44.4%**	**33.3%**
เป็นโมเดลแบบเปิดเพิ่มขึ้น	ในปี 2022	ในปี 2021 [5]

สถิตินี้ชี้ให้เห็นถึงการเคลื่อนตัวของอุตสาหกรรมไปสู่วิถีแบบเปิดมากขึ้นอย่างรวดเร็ว จากเดิมที่เคยถูกครอบงำด้วยโมเดลปิดของบริษัทใหญ่ไม่กี่แห่ง ปัจจุบันองค์กรต่าง ๆ รวมถึงสถาบันวิจัย เริ่มปล่อยโมเดลออกมาในรูปแบบเปิดมากขึ้น ส่งผลให้ความรู้และเทคโนโลยีกระจายตัวสู่วงกว้าง เกิดการลองผิดลองถูกและนวัตกรรมใหม่ ๆ โดยไม่จำกัดอยู่แค่เบื้องหลังประตูบ้านของบริษัทยักษ์ใหญ่เพียงไม่กี่ราย

การแข่งขันที่เพิ่มขึ้นระหว่างโมเดลแบบเปิดและแบบปิดได้กลายเป็นประเด็นสำคัญ โดยธุรกิจต่างๆ เริ่มใช้โมเดลแบบเปิดสำหรับแช็ตบ็อตและแอปพลิเคชัน AI ของตนเอง ซึ่งเป็นการท้าทายโดยตรงต่อบริการ AI แบบชำระเงินที่ใช้โมเดลแบบปิด

นอกจากนี้ โมเดลแบบเปิดยังช่วยส่งเสริมนวัตกรรม เพราะนักพัฒนาและผู้ใช้งานทั่วไปสามารถนำเครื่องมือเหล่านี้ไปประยุกต์สร้างสรรค์สิ่งใหม่หรือแก้ปัญหาเฉพาะทางได้โดยไม่ต้องลงทุนสร้างโมเดลขึ้นมาใหม่ตั้งแต่ต้น

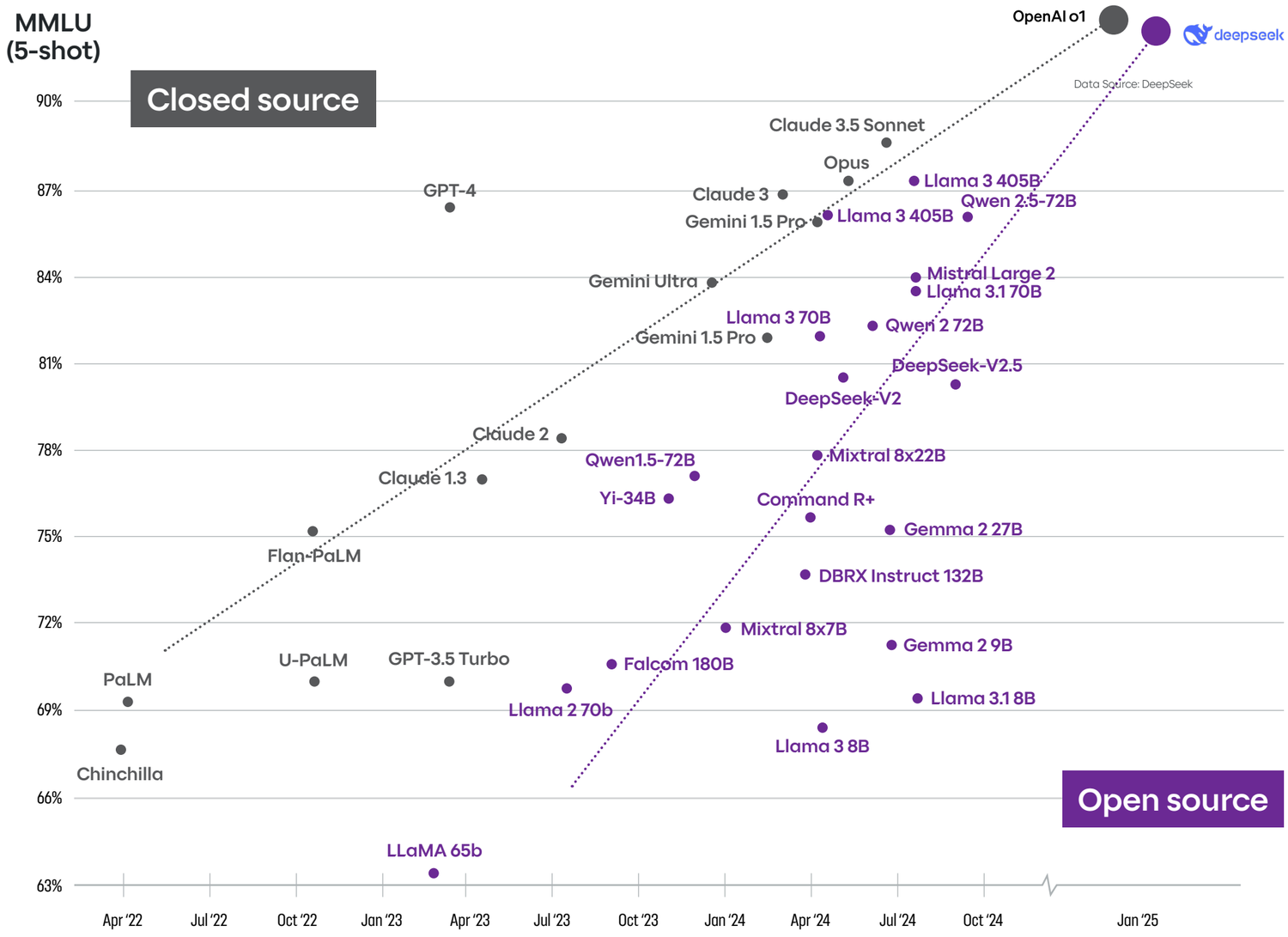
DeepSeek และการปิดช่องว่างของโมเดลเปิดกับปิด

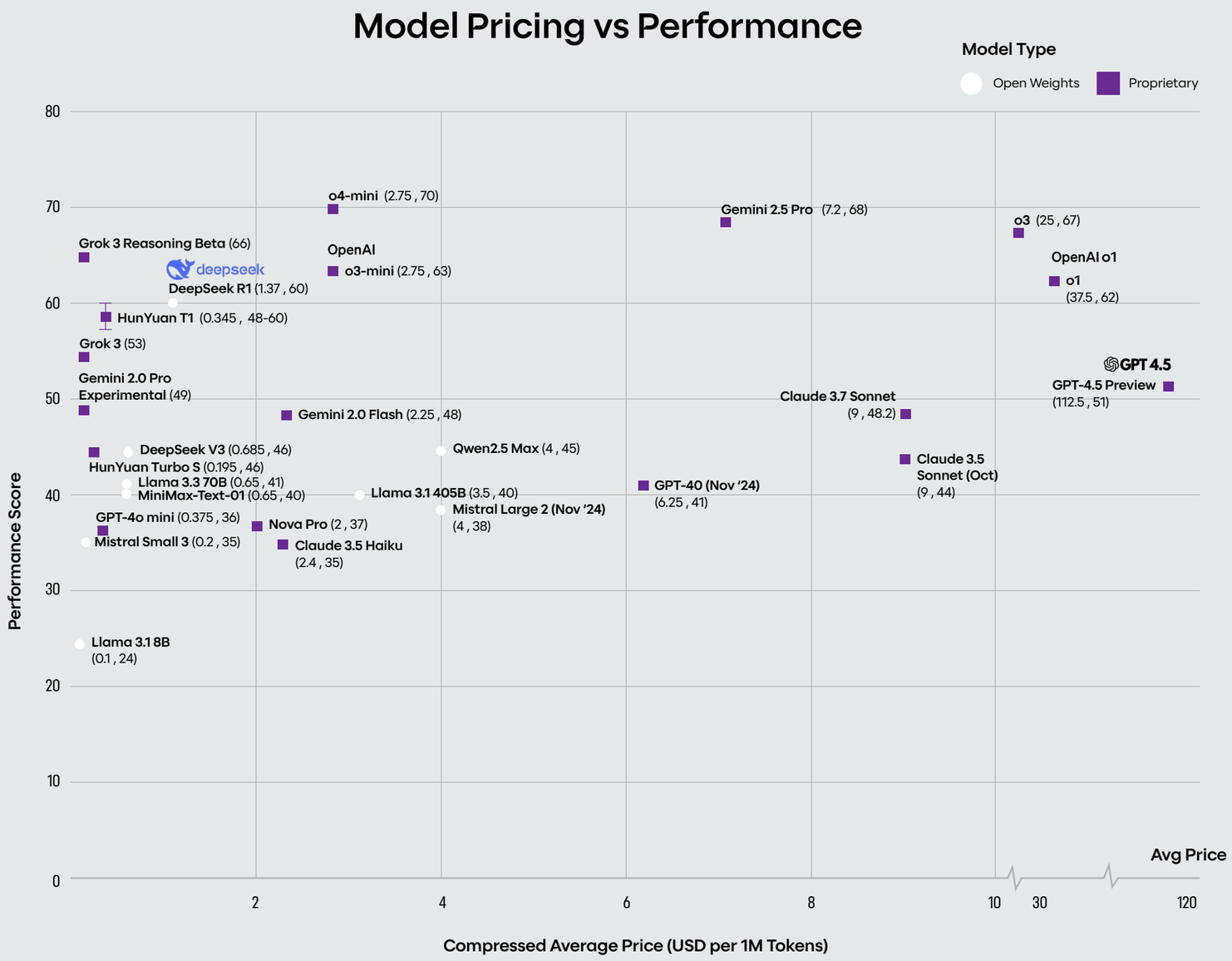
แม้ที่ผ่านมาโมเดลปิดระดับแนวหน้าจะครองความได้เปรียบด้านประสิทธิภาพอยู่บ้าง แต่ช่องว่างระหว่างโมเดลแบบปิดกับแบบเปิดกำลังลดลงอย่างรวดเร็ว [6]

DeepSeek เป็นตัวอย่างของโมเดลแบบเปิดที่เติบโตอย่างรวดเร็ว ภายในเวลาเพียงสองปี ได้เปิดตัวหลายเวอร์ชัน ตั้งแต่ DeepSeek-Coder, DeepSeek-V2, ไปจนถึง DeepSeek R1 ซึ่งแสดงให้เห็นถึงความสามารถในการพัฒนาอย่างต่อเนื่อง จุดเด่นสำคัญคือการที่ DeepSeek R1 สามารถปิดช่องว่างด้านประสิทธิภาพกับโมเดลปิดที่ใช้ทรัพยากรและต้นทุนสูงในการพัฒนาได้เร็วกว่าที่คาดไว้ โดยเฉพาะผลการทดสอบใน MMLU (Massive Multitask Language Understanding) [7] ซึ่งเป็นเกณฑ์มาตรฐานสำคัญที่ใช้ประเมินความสามารถของโมเดล AI ในด้านความเข้าใจภาษา ความรู้ทั่วไป และการแก้ปัญหาขั้นสูง

DeepSeek R1 ทำคะแนนได้ใกล้เคียงกับ OpenAI o1 ของ OpenAI ซึ่งเป็นโมเดลปิดชั้นนำ แสดงให้เห็นว่าประสิทธิภาพของโมเดลแบบปิดกำลังเข้าใกล้โมเดลเปิดมากขึ้นเรื่อย ๆ ทั้งที่ DeepSeek ใช้ทุนสร้างเพียง 5.6 ล้านดอลลาร์สหรัฐ หรือ ประมาณ 190 ล้านบาท (แต่ตัวเลขนี้นักวิเคราะห์หลายคนก็มองว่าอาจจะต่ำไปกว่าความเป็นจริงเพราะไม่รวมค่าใช้จ่ายเพิ่มเติม เช่น การฝึกหลายรอบ การรวบรวมข้อมูล โครงสร้างพื้นฐาน และค่าใช้จ่ายในการดำเนินงาน เมื่อพิจารณาปัจจัยเหล่านี้แล้ว การลงทุนทั้งหมดอาจสูงกว่าที่ประมาณการไว้อย่างมาก) [8]

ผลกระทบของ DeepSeek กระเพื่อมไปสู่ตลาดหุ้นทางฝั่งเทคโนโลยีอย่างเห็นได้ชัดหลังจากมีการเปิดตัวของ DeepSeek ไปในวันที่ 27 มกราคม 2025 อุตสาหกรรม Datacenter Energy เช่น Vertiv หรือ GE ติดลบไปมากถึง 30% ส่วนหุ้น Semiconductor ที่ถือว่าเป็น Growth Stocks ที่มีส่วนเกี่ยวข้องกับ AI อย่าง Nvidia และ TSMC ลงไปกว่า 20% แต่ผู้ให้บริการคลาวด์ (Cloud Service Provide) อย่าง Google, Microsoft, Amazon หรือ Meta ไม่ได้รับผลกระทบรุนแรงเท่าไหร่นัก และสุดท้ายหุ้นที่ได้รับผลบวก คือหุ้นที่เกี่ยวข้องกับ application layer อย่างเช่นหุ้น Adobe, Salesforce หรือ SAP





ภายหลังระหว่าง 28-30 มกราคม 2025 หุ้นในส่วนที่ติดลบแรง ๆ ดีดตัวกลับขึ้นมาบ้าง แต่ก็ไม่ยังไม่ฟื้นไปยังระดับก่อนการเปิดตัวของ DeepSeek ในส่วนของผู้ให้บริการคลาวด์ก็ทรงตัว มีบางตัวอย่าง Google/Meta ที่ปรับขึ้นมาบวกด้วย

การที่ DeepSeek สามารถสร้างโมเดลที่มีประสิทธิภาพเทียบเคียงโมเดลปิดได้ภายในเวลาเพียงสองปี สะท้อนให้เห็นถึงแนวโน้มที่โมเดลแบบเปิดอาจมีศักยภาพเทียบเท่าโมเดลปิดในอนาคตอันใกล้ สิ่งนี้อาจส่งผลให้ AI พัฒนาเร็วขึ้น ลดต้นทุน และกระจายการเข้าถึง เทคโนโลยี AI ในอุตสาหกรรมต่าง ๆ การแข่งขันระหว่างทั้งสองโมเดลมีแนวโน้มจะทำให้ตลาด AI เปลี่ยนแปลงไปอย่างรวดเร็ว ทั้งในแง่ของต้นทุน การนำไปใช้ และการพัฒนาอย่างต่อเนื่องในอนาคต

ยกตัวอย่างในภาพ เราจะเห็นได้เลยว่า GPT4.5 ของ OpenAI นั้นถึงแม้จะควบคุมปัญหาเรื่องการสร้างข้อมูลขึ้นมาเอง (Hallucination) ได้เป็นอย่างดี แต่ค่าใช้จ่ายในการประมวลผล 1M Tokens จะอยู่ที่ราว ๆ 112.5 ดอลลาร์สหรัฐ และมีผลคะแนนประสิทธิภาพการทำงานที่ 51 (การประเมินประสิทธิภาพเป็นการวัดความสามารถของโมเดลภาษาในด้านการให้เหตุผล ความรู้คณิตศาสตร์ และการเขียนโปรแกรม) ซึ่งค่าใช้จ่ายตรงนี้สูงมากหากเทียบกับ DeepSeek R1 ที่ใช้เพียง 1.37 ดอลลาร์สหรัฐต่อการประมวลผล 1M Tokens และมีผลคะแนนประสิทธิภาพการทำงานที่ 60[9]

นอกจากเรื่องของประสิทธิภาพและค่าใช้จ่ายแล้ว อีกเทรนด์หนึ่งที่ตอนนี้ทั้งสองฝั่งกำลังทำเหมือนกันคือการผลักดันในเรื่องของ 'ความเร็ว' (Speed) ในเรื่องการตอบอย่างมีเหตุผลรองรับ (Reasoning) เพราะในโมเดลรุ่นก่อน ๆ การประมวลผลเพื่อให้ได้คำตอบที่มีเหตุผลต้องใช้พลังค่อนข้างมากในขั้นตอนการเรียกใช้งานโมเดล (inference) เพราะฉะนั้นเรื่องของความเร็วในการตอบสนองอย่างมีเหตุผลก็จะกลายเป็นอีกตัววัดสำคัญสำหรับโมเดล AI ในรุ่นต่อ ๆ ไปด้วย

ถ้าจะบอกว่า "AI โมเดลแบบเปิดกำลังไล่หลังมาติด ๆ" และมีศักยภาพที่จะทัดเทียมหรืออาจแซงหน้าโมเดลปิดในบางด้านได้ในที่สุดก็คงไม่ใช่คำกล่าวที่เกินจริงนัก

อย่างไรก็ดี โมเดลปิดชั้นนำยังคงครองความได้เปรียบในงานบางประเภท (เช่น ความเข้าใจภาษาขั้นสูง หรือการจัดการบริบทยาว ๆ ที่ GPT-4 ทำได้ดี) เนื่องจากได้รับการฝึกด้วยทรัพยากรมหาศาลและข้อมูลขนาดใหญ่เฉพาะทางที่บริษัทเจ้าของโมเดลมี แต่ด้วยอัตราการพัฒนาของโมเดลแบบเปิดในปัจจุบันที่ก้าวกระโดด - ทั้งจำนวนโมเดลใหม่ที่ออกสู่สาธารณะเพิ่มขึ้นเท่าตัวทุกปี และการที่โมเดลแบบเปิดได้รับการ "ปั้น" ต่อโดยนักพัฒนาทั่วโลก - ทำให้ช่องว่างประสิทธิภาพดังกล่าวแคบลงเรื่อย ๆ ยิ่งไปกว่านั้นโมเดลเปิดยังมีข้อได้เปรียบเรื่องความยืดหยุ่นและการปรับแต่งที่บางครั้งทำให้มันตอบโจทย์งานเฉพาะได้ดีกว่าโมเดลปิดทั่วไป ตัวอย่างเช่น บริษัทที่ต้องการ AI วิเคราะห์เอกสารเฉพาะทาง อาจเลือกนำโมเดลแบบเปิดมาฝึกต่อ (fine-tune) กับชุดข้อมูลของบริษัท เพื่อให้ได้ผู้ช่วย AI ที่เข้าใจบริบทงานของตนอย่างลึกซึ้ง ซึ่งโมเดลปิดสำเร็จรูปอาจไม่สามารถให้ผลลัพธ์เฉพาะทางได้เทียบเท่า

บทสรุป : โมเดล AI แบบเปิดกำลังเปลี่ยนแปลงเศรษฐศาสตร์ของ AI

การแข่งขันระหว่าง Open-Source และ Closed-Source AI ไม่เพียงแต่เปลี่ยนแปลงด้านเทคโนโลยี แต่ยังพลิกโฉมเศรษฐศาสตร์ของ AI ด้วยหลัก Economies of Scale และ Economies of Speed

1. Economies of Scale:

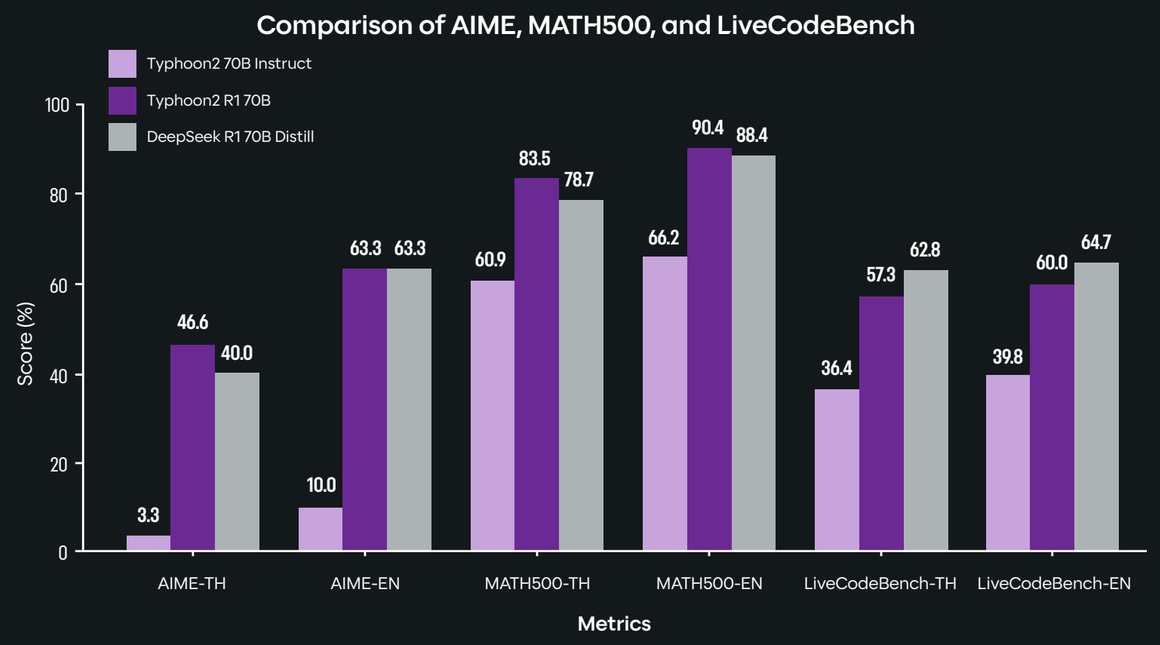
โมเดลเปิดช่วยลดต้นทุนการพัฒนา AI โดยกระจายภาระงานให้กับชุมชนและนักพัฒนาทั่วโลก ซึ่งต่างจากโมเดลปิดที่ต้องลงทุนมหาศาลเพียงผู้เดียว ธุรกิจสามารถนำโมเดลแบบเปิดมาใช้งานโดยปรับแต่งเพิ่มเติมได้โดยไม่ต้องเริ่มต้นจากศูนย์ ลดค่าใช้จ่ายและเร่งการใช้งาน AI ในเชิงพาณิชย์

2. Economies of Speed:

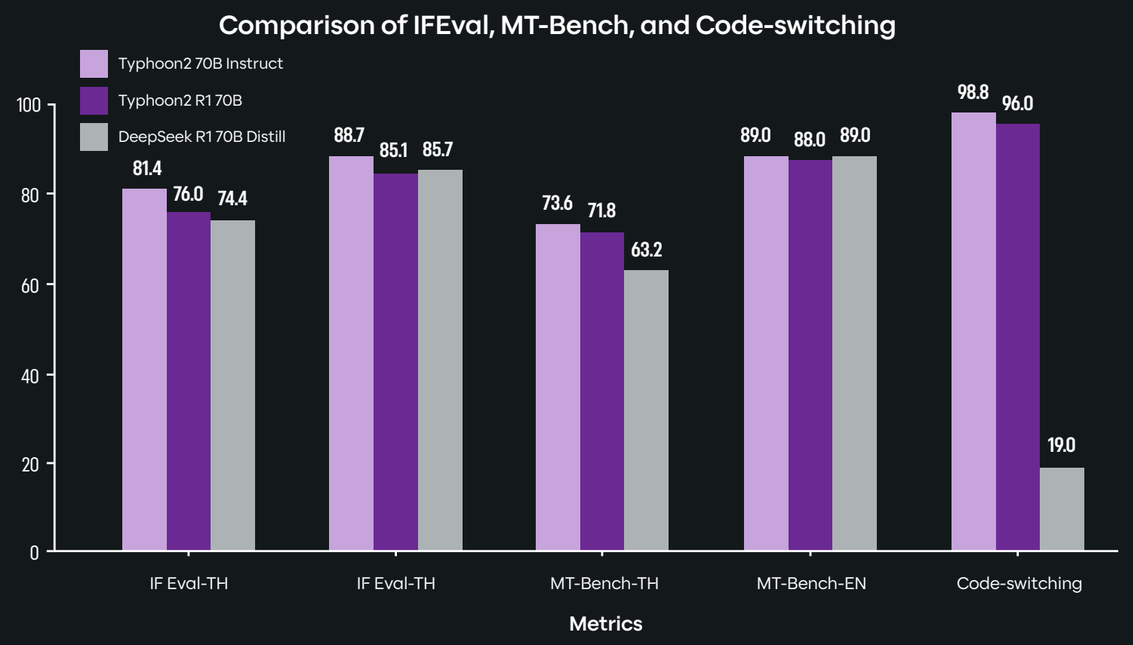
การเปิดซอร์สโค้ดช่วยให้การพัฒนา AI เร็วยิ่งขึ้น การที่นักพัฒนาจำนวนมากสามารถเข้ามาปรับปรุงโมเดล ทำให้การแก้ปัญหาและพัฒนานวัตกรรมใหม่เกิดขึ้นในอัตราที่เร็วกว่าการพัฒนาแบบปิดที่จำกัดอยู่เฉพาะภายในบริษัท การปรับตัวและการเรียนรู้ร่วมกันจากทั้งโลก ทำให้โมเดลเปิดสามารถพัฒนาให้เทียบเท่าโมเดลปิดได้ในเวลาที่สั้นลงกว่าที่เคย

ยกตัวอย่างที่ชัดเจนมาก ๆ คือ Typhoon2 R1 70B โมเดลแบบเปิดของ SCB10X ที่ใช้งบประมาณเพียง 120 ดอลลาร์สหรัฐ (ประมาณ 4,000 บาท) และระยะเวลาสั้นมากแค่ 15 ชั่วโมงในการพัฒนา โดยอาศัยการรวม (merging) โมเดลสองตัว อย่าง Typhoon2 70B (โมเดลที่เชี่ยวชาญด้านภาษาไทย) กับ DeepSeek R1 70B Distill (โมเดลที่มีความสามารถ reasoning สูง) จนกลายเป็นโมเดลที่มีความสามารถในเรื่อง Reasoning และ ภาษาไทยได้ดีขึ้นกว่า DeepSeek R1 อย่างมีนัยสำคัญ [10] [11]

Reasoning Benchmarks



Instruction Following/Code-Switching Benchmarks



'โมเดลปิด' มอบประสิทธิภาพสูงสุดและความเชื่อมั่นในการสนับสนุน แต่ก็แลกมาด้วยความโปร่งใสที่น้อยและการกระจุกตัวของอำนาจในมือผู้พัฒนาไม่กี่ราย

ส่วน 'โมเดลเปิด' ให้คำมั่นสัญญาแห่งความร่วมมือและการเข้าถึงที่ทั่วถึง ซึ่งกำลังพิสูจน์ตัวเองว่ามีศักยภาพสูงขึ้นจนเกือบเทียบเคียงโมเดลปิดได้ในหลายด้าน การผสานสองแนวทางนี้เข้าด้วยกันอย่างสมดุลจะเป็นกุญแจสำคัญที่ทำให้วงการ AI เติบโตอย่างต่อเนื่องและนำประโยชน์มาสู่เศรษฐกิจและสังคมโดยรวม

ในท้ายที่สุด ไม่ว่าจะเป็น Open หรือ Closed-Source AI Models เป้าหมายปลายทางคือการใช้ปัญญาประดิษฐ์ให้เกิดประโยชน์สูงสุดต่อมนุษยชาติ การมีทั้งสองแนวทางอยู่ร่วมกันจะช่วยให้เป้าหมายนี้สัมฤทธิ์ผลได้รวดเร็วและรอบด้านมากยิ่งขึ้น พร้อมกับรักษาสมดุลระหว่างนวัตกรรม ความโปร่งใส และความรับผิดชอบต่อไปในอนาคต

ACT II:
Tiny Titans - Small, but Mighty.

More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

ปัญญาประดิษฐ์ (AI) กำลังพัฒนาอย่างรวดเร็ว โดยมี 3 เทรนด์หลักที่โดดเด่นในช่วงนี้ ได้แก่ (1) โมเดลที่มีความสามารถรับและประมวลผลข้อมูลหลายรูปแบบ, (2) โมเดลขนาดเล็กสำหรับงานเฉพาะทาง, และ (3) การสเกลความสามารถของโมเดลในช่วง inference เพื่อการให้เหตุผลที่ดีขึ้น แต่ละเทรนด์มีความสำคัญและส่งผลต่อธุรกิจและเศรษฐกิจในรูปแบบที่แตกต่างกัน เป็นความพยายามขยายขีดความสามารถของ AI ไปสู่น่านน้ำใหม่ที่ตอบโจทย์การใช้งานและข้ามข้อจำกัดทางเทคนิคต่าง ๆ ที่กำลังเผชิญอยู่

1 โมเดลที่มีความสามารถรับและประมวลผลข้อมูลหลายรูปแบบ (Versatile Multimodal Models)

โมเดล 'AI แบบมัลติโหมด' คือโมเดลที่สามารถประมวลผลข้อมูลได้หลายรูปแบบพร้อมกัน เช่น ข้อความ ภาพ เสียง วิดีโอ เป็นต้น ทำให้ระบบสามารถเข้าใจบริบทได้ลึกและตอบสนองงานที่ซับซ้อนได้หลากหลายยิ่งขึ้น โมเดลภาษาขนาดใหญ่แบบมัลติโหมด (Multimodal LLM) กำลังเป็นแนวโน้มที่มาแรงเพราะสามารถมอบบริบทที่ลึกซึ้งและความยืดหยุ่นสูงในการใช้งานกับงานจริงหลากหลายประเภท

เทรนด์นี้นำไปสู่การพัฒนาโมเดลใหม่ ๆ ที่เข้าใจข้อมูลหลากรูปแบบ และเริ่มมีการผนวกรวมโมเดลเหล่านี้เข้ากับระบบ AI agent เพื่อให้ทำงานแบบอัตโนมัติได้อย่างชาญฉลาด Demis Hassabis CEO ของ Google DeepMind กล่าวไว้ว่า "ความสามารถแบบมัลติโหมดและการผนวก AI Agent เข้าด้วยกันถือเป็นหนึ่งในก้าวที่สำคัญในการไปสู่ AGI (Artificial General Intelligence)"[1]

2021 & 2022

1

Image-text models boom

2021

- OpenAI CLIP
- Google ALIGN
- Salesforce ALBEF

ALEPH ALPHA

Introduced multimodal capabilities in '21

2022

- Google LiT
- Google COCA
- DeepMind FLAMINGO
- Salesforce BLIP

2023 & 2024

2

LLMs understand Images

2023

- Salesforce BLIP 2
- Microsoft KOSMOS
- OpenAI GPT-4
- LLaVA

OpenAI

Mar' 23

Natively Multimodal model emerge

- Google Gemini
- Adept FUYU-8B

Gemini

Google introduces first commercial model taking diverse inputs in Dec' 23

Rising of more multimodal and its applications

2024

- Advancements in Multimodal Models

OpenAI GPT-4o Mar | Gemini Sep | NVIDIA Sep | LLaMA by Meta Oct

- Versatility Across Applications

Typeface Text Image

- Human-AI interaction

2025

3

Multimodal reasoning accelerates progress toward AGI

deepseek Janus Pro | OpenAI o3-mini

Qwen2.5 Max

Jan

Grok

Gemini 2.0 Flash

Feb

Gemma 3

Qwen2.5-Omni

Gemini 2.0 Pro

Baidu 百度 ERNIE 4.5

Mar

อย่างที่เห็นในภาพ แนวโน้มนี้เติบโตมาตั้งแต่ช่วงปี 2021 - 2022 ซึ่งถือเป็นยุคการเติบโตของโมเดลที่รองรับหลายโหมดข้อมูล หลังจากนั้นก็เติบโตมาเรื่อย ๆ คาดว่าจะเติบโตต่อเนื่องไปในปี 2025 (และอนาคต) การพัฒนาโมเดลมัลติโหมดมากขึ้นและประยุกต์ใช้ในงานต่าง ๆ อย่างแพร่หลาย รวมถึงความก้าวหน้าในการให้เหตุผลจากหลายข้อมูล (multimodal reasoning) ที่ถูกมองว่าจะเร่งให้เราเข้าใกล้ AGI เร็วยิ่งขึ้น[2] [3] [4]

ผลกระทบต่อเศรษฐกิจและอุตสาหกรรม

โมเดลมัลติโหมดมีศักยภาพในการสร้างมูลค่าทางเศรษฐกิจสูง เพราะสามารถนำไปใช้แก้ปัญหาในหลากหลายอุตสาหกรรมได้ในโมเดลเดียว ตัวอย่างเช่น ในภาคการฝึกอบรมและการศึกษา สามารถใช้สร้างผู้ช่วยสอนอัจฉริยะที่เข้าใจทั้งข้อความและภาพประกอบ ช่วยอธิบายเนื้อหายาก ๆ ให้ผู้เรียนได้ดีขึ้น

ยกตัวอย่าง บริษัท SCB X Public Company Limited ได้นำมาใช้ในระบบฝึกอบรมพนักงานผู้จัดการความสัมพันธ์ลูกค้า (Smart RM Training) ที่เป็นการจำลอง (Simulate) ลูกค้าเพื่อใช้เป็นคู่สนทนาฝึกการขายให้กับฝ่ายขาย โดยใช้โมเดล Voice-to-Voice Realtime ซึ่งถือว่าเป็นการพัฒนาไปอีกขั้นหนึ่งของ Multimodal LLM ผลลัพธ์คือบุคลากรได้รับการพัฒนาทักษะอย่างทั่วถึงและรวดเร็ว ในขณะที่บริษัทก็ประหยัดค่าใช้จ่ายลง

หรืออย่างในวงการแพทย์ สามารถช่วยแพทย์วินิจฉัยโรคโดยอ่านผลเอ็กซเรย์หรือ MRI ควบคู่กับการอ่านรายงานข้อความ และในภาคบริการลูกค้า แชทบ็อตอัจฉริยะสามารถเข้าใจทั้งคำพูดและรูปภาพจากลูกค้า ทำให้ตอบสนองได้ตรงจุดมากขึ้น เทรนด์นี้ช่วยเพิ่มประสิทธิภาพการทำงานในหลายภาคส่วนและยกระดับประสบการณ์ผู้ใช้ เนื่องจาก AI เข้าใจบริบทจากหลายแหล่งข้อมูล

แนวโน้มอนาคต

คาดว่าเทรนด์นี้จะยิ่งมีบทบาทสำคัญในอนาคต โมเดลอาจสามารถรับข้อมูลจากประสาทสัมผัสต่าง ๆ (เสียง, ภาพ, ข้อความ, เซ็นเซอร์ IoT) แล้วประมวลผลร่วมกันเพื่อให้ผลลัพธ์ที่สอดคล้องและมีเหตุผลลึกซึ้งขึ้น เช่น หุ่นยนต์ผู้ช่วยในโรงงานที่มองเห็นสถานการณ์และฟังคำสั่งมนุษย์ไปพร้อมกัน หรือนักวิเคราะห์ข้อมูลที่ให้อธิบายข้อมูลตัวเลขพร้อมกับกราฟและรูปภาพประกอบได้อย่างเป็นธรรมชาติ

นอกจากนี้ การผสานโมเดลมัลติโหมดเข้ากับ Agentic AI (Full-Duplex in Agentic AI) ที่เป็นความสามารถของ Agentic AI ในการโต้ตอบแบบสองทางพร้อมกัน (Simultaneous Two-Way Communication) ฟังก์ชันนี้ไม่เพียงแต่จะช่วยเพิ่มประสิทธิภาพและความเป็นธรรมชาติของการสื่อสารระหว่างมนุษย์กับระบบ AI แต่ยังช่วยให้ AI สามารถวิเคราะห์และประมวลผลข้อมูลได้อย่างรวดเร็ว นำไปสู่การตัดสินใจที่แม่นยำและทันเวลา ซึ่งเป็นประโยชน์อย่างยิ่งในงานที่ต้องการการตอบสนองทันที เช่น การบริการลูกค้า การสนับสนุนทางเทคนิค หรือการช่วยเหลือในสถานการณ์ฉุกเฉิน[5]

แนวทางนี้จะเพิ่มขีดความสามารถในการใช้งาน AI โดยไม่ต้องมีมนุษย์กำกับทุกขั้นตอน ซึ่งจะเปลี่ยนรูปแบบการทำงานของหลายธุรกิจ อย่างไรก็ตาม ความท้าทายในอนาคตคือการบริหารจัดการต้นทุนการประมวลผลและพลังงานของโมเดลขนาดใหญ่เหล่านี้ และการจัดการข้อมูลหลายประเภทอย่างมีประสิทธิภาพ

2 โมเดลขนาดเล็กสำหรับงานเฉพาะทาง (Small Models for Specific Tasks)

เทรนด์ที่สองคือการเน้นพัฒนา โมเดล AI ขนาดเล็ก (Small Language Model: SLM) ที่ได้รับการปรับแต่งให้ทำงานเฉพาะทางได้อย่างมีประสิทธิภาพ โมเดลเหล่านี้มีจำนวนพารามิเตอร์น้อยกว่าโมเดลขนาดใหญ่ (เช่น GPT-3/GPT-4) หลายเท่า แต่ถูกฝึกมาให้เชี่ยวชาญในงานใดงานหนึ่ง ทำให้สามารถประมวลผลได้รวดเร็ว ใช้ทรัพยากรต่ำ และปรับแต่งเพิ่มเติม (fine-tune) ได้ง่ายและประหยัดกว่า

จากเหตุผลเหล่านั้นทำให้โมเดลขนาดเล็กกำลังได้รับความนิยมมากขึ้นเรื่อย ๆ อีกทั้งยังสามารถปรับใช้ (deploy) บนอุปกรณ์ปลายทาง (edge devices) อย่างโทรศัพท์มือถือหรืออุปกรณ์ IoT ได้ง่าย ช่วยเพิ่มการเข้าถึงเทคโนโลยีและรักษาความเป็นส่วนตัวของข้อมูลเพราะประมวลผลบนเครื่องผู้ใช้โดยตรง

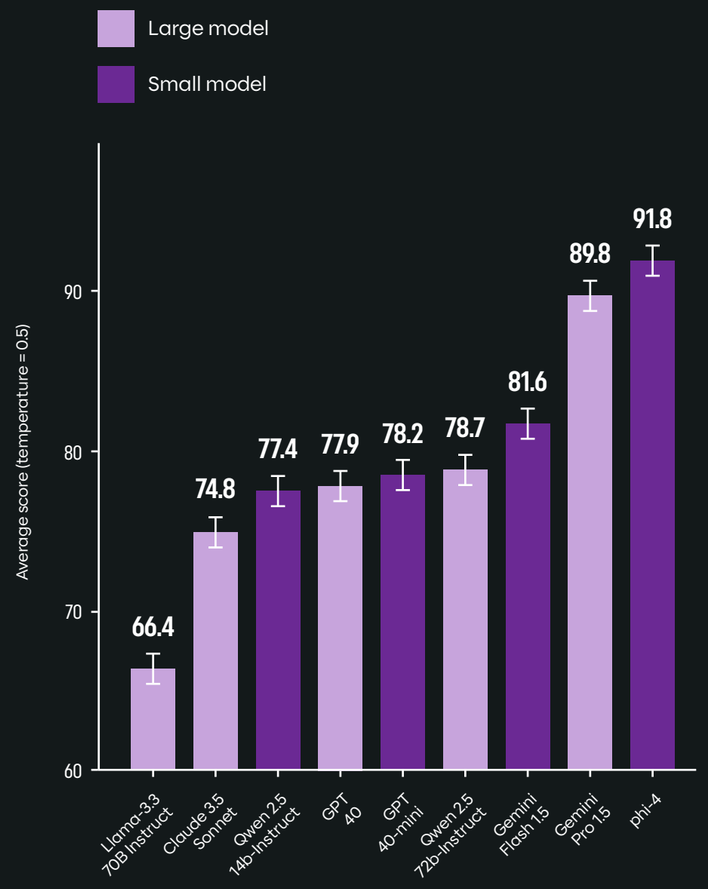
ผลกระทบต่อเศรษฐกิจและอุตสาหกรรม

โมเดลเล็กเฉพาะทางเปิดโอกาสให้ธุรกิจขนาดกลางและเล็กสามารถนำ AI ไปใช้ได้อย่างกว้างขวางยิ่งขึ้น โดยไม่ต้องลงทุนโครงสร้างพื้นฐานขนาดใหญ่ ลดความเหลื่อมล้ำทางเทคโนโลยีระหว่างองค์กรขนาดใหญ่กับรายย่อย

ยกตัวอย่างเช่น ธุรกิจสมาร์ทโฟน เช่น ผู้ช่วยเสียงอัจฉริยะ (voice assistant) บนมือถือรุ่นใหม่ ก็เริ่มประมวลผลคำสั่งเสียงบนเครื่องโดยตรง แทนที่จะส่งข้อมูลไปยังเซิร์ฟเวอร์กลาง ส่งผลให้ตอบสนองได้เร็วขึ้น และข้อมูลส่วนตัวเสียงของผู้ใช้ไม่หลุดออกสู่ภายนอก ภาคการแพทย์ เช่น แอปฯ วินิจฉัยโรคทางผิวหนังที่ทำงานบนแท็บเล็ตหรือโทรศัพท์ในพื้นที่ห่างไกล สามารถช่วยคัดกรองผู้ป่วยเบื้องต้นได้โดยไม่ต้องเชื่อมต่ออินเทอร์เน็ต นอกจากนี้ โมเดลเล็กที่ทำงานเฉพาะทางยังมีแนวโน้มที่จะมีความปลอดภัยและโปร่งใสสูงกว่าในบางด้าน เพราะโค้ดและพารามิเตอร์มีขนาดเล็กกว่าจึงตรวจสอบและควบคุมได้ง่ายกว่า

แม้จะมีขนาดเล็ก แต่ประสิทธิภาพของโมเดลเฉพาะทางหลายตัวน่าประทับใจ จนสามารถเทียบชั้นหรือเหนือกว่าโมเดลใหญ่ทั่วไปในงานบางประเภทได้ งานวิจัยล่าสุดของ Microsoft นำเสนอโมเดลชื่อ Phi-4 ที่มีขนาดเพียง 14,000 ล้านพารามิเตอร์ แต่ให้ผลลัพธ์การให้เหตุผลเชิงซับซ้อนสูง สามารถแก้ปัญหาคณิตศาสตร์ได้ดีกว่าโมเดลขนาดใหญ่อย่าง Gemini Pro 1.5 อย่างที่เราเห็นในภาพ พร้อมทั้งถูกออกแบบให้คำนึงถึงความรับผิดชอบของ AI (Responsible AI) ตั้งแต่ต้น อีกตัวอย่างหนึ่งคือบริษัทสตาร์ทอัพ

Average performance on November 2024 AMC-10/12 tests



Mistral AI ที่เปิดตัวโมเดลเล็ก Mistral 3B และ 8B (มีขนาด 3,000 ล้าน และ 8,000 ล้านพารามิเตอร์ตามลำดับ) ซึ่งกลายเป็นโมเดลระดับแนวหน้าที่ทำคะแนนทดสอบมาตรฐานได้เหนือกว่าโมเดลอื่น ๆ ในระดับขนาดใกล้เคียงกัน[6] [7]

นวัตกรรมเช่นนี้ชี้ให้เห็นว่า "โมเดล AI ขนาดเล็กสามารถแข่งขันกับโมเดลใหญ่ได้ในงานเฉพาะทาง ช่วยประหยัดทรัพยากร และทำงานได้บนเครื่องลูกข่าย" (ซึ่งเป็นแนวคิดที่กำลังเปลี่ยนมุมมองของอุตสาหกรรมว่าไม่จำเป็นต้องมีขนาดใหญ่เสมอไปจึงจะเก่ง) นอกจากนี้ บทความของ MIT Technology Review ยังยกให้ "โมเดลภาษาขนาดเล็ก" เป็นหนึ่งใน 10 เทคโนโลยีพลิกวงการปี 2025 โดยชี้ว่าหลังจากที่โมเดลภาษาขนาดใหญ่ได้สร้างความก้าวหน้าอย่างมหาศาลในช่วงก่อนหน้า ถึงเวลาที่จะหันมาให้ความสำคัญกับโมเดลที่เล็กลงและฉลาดขึ้น เพื่อให้ AI กระจายไปสู่การใช้งานจริงได้มากขึ้น

แนวโน้มอนาคต

เราจะได้เห็นโมเดลเฉพาะทางขนาดเล็กจำนวนมากขึ้นสำหรับงานอันหลากหลาย ไม่ว่าจะเป็นผู้ช่วยเขียนโค้ดเฉพาะด้าน, โมเดลวิเคราะห์ข้อมูลการแพทย์เฉพาะโรค, หรือโมเดลแนะนำสินค้าเฉพาะกลุ่มลูกค้า เป็นต้น

การแข่งขันระหว่างบริษัทเทคโนโลยีจะมุ่งไปที่ใครสามารถพัฒนาโมเดลขนาดเล็กที่เก่งพอ ๆ กับโมเดลใหญ่ในงานที่เจาะจงแต่ใช้ทรัพยากรน้อยกว่าอย่างมีนัยสำคัญได้ก่อน ซึ่งหากทำได้จะเป็นข้อได้เปรียบทางธุรกิจมหาศาล

Arvind Krishna CEO ของ IBM ชี้ว่าหากเราสามารถฝึกโมเดลให้มีประสิทธิภาพใกล้เคียงกันโดยใช้ต้นทุนเพียงหนึ่งในร้อย และทำงานบนโครงสร้างพื้นฐานที่เล็กลงได้ ก็ย่อมทำให้ทุกองค์กรต้องเร่งปรับตัวเพื่อแข่งขันในจุดนี้[8]

เทรนด์การเปลี่ยนแปลงนี้จะช่วยส่งเสริมวงการ AI แบบเปิด (Open Source AI) มากขึ้น เพราะโมเดลเล็กพัฒนาต่อยอดได้ง่าย ชุมชนนักพัฒนาสามารถปรับปรุงร่วมกันและแชร์ได้สะดวก เมื่อโมเดลเล็กฉลาดขึ้นเรื่อย ๆ ผู้ใช้งานทั่วไปอาจเข้าถึง AI คุณภาพสูงได้โดยไม่ต้องพึ่งพาบริษัทใหญ่หรือฮาร์ดแวร์ราคาแพง ซึ่งจะเป็นการกระจายโอกาสและประโยชน์ของ AI สู่วงกว้างในสังคม

3 การสเกลการประมวลผลช่วง inference เพื่อการให้เหตุผลที่ดีขึ้น (Scaling Intelligence for Improved Reasoning)

เทรนด์ที่สามเป็นการตั้งคำถามกับแนวคิดการเพิ่มขนาดโมเดลเพื่อให้ AI เก่งขึ้น (Scaling Up) ที่ครองวงการมาตลอดทศวรรษที่ผ่านมา [9]

โมเดลภาษาขนาดใหญ่ (LLMs) รุ่นใหม่ ๆ อย่าง GPT-3, GPT-4 มีจำนวนพารามิเตอร์มหาศาลและฝึกบนชุดข้อมูลขนาดใหญ่ ส่งผลให้ความสามารถในการเข้าใจและให้เหตุผลดีขึ้นตามการเพิ่มขึ้นของขนาดโมเดล ซึ่งก็สอดคล้องกับกฎการปรับขนาด (Scaling Laws) ที่นักวิจัยพบว่าโดยทั่วไปการเพิ่มขนาดโมเดล ปริมาณข้อมูล และพลังประมวลผล จะส่งผลให้ประสิทธิภาพของโมเดลเพิ่มขึ้นแบบก้าวกระโดด

อย่างไรก็ตาม ในปัจจุบันเมื่อขนาดโมเดลใหญ่ขึ้นมาก สิ่งที่เริ่มปรากฏขึ้นคือสัญญาณของสิ่งที่เรียกว่ากฎผลตอบแทนลดน้อยถอยลง (law of diminishing returns) กล่าวคือ อัตราการพัฒนาความสามารถของ AI จะชะลอลงแม้จะเพิ่มพารามิเตอร์หรือข้อมูลมากขึ้น ซึ่งทำให้การขยายขนาดต่อไปมีต้นทุนสูงมากแต่ได้ประโยชน์น้อยลง

นอกจากนี้ยังมีข้อจำกัดด้านข้อมูลและต้นทุนการฝึกที่เป็นอุปสรรคใหญ่เหมือนกำแพงอยู่ด้วย นักวิจัยคาดการณ์ว่าภายในเวลาไม่กี่ปีเราจะเผชิญภาวะที่ "ข้อมูลคุณภาพสูงสำหรับฝึกโมเดลมีไม่เพียงพอ" มีการประมาณว่าข้อมูลข้อความคุณภาพสูงบนโลกอาจถูกใช้จนหมดภายในปี 2026 ซึ่งแนวโน้มนี้จะเป็นปัจจัยฉุดรั้งความก้าวหน้าของ AI อย่างมากหากไม่มีการปรับเปลี่ยนแนวทาง

เพราะฉะนั้นการสเกลโมเดลในแบบเดิม (แค่เพิ่มขนาดและหวังผลลัพธ์ดีขึ้นเรื่อย ๆ) อาจเดินทางมาถึงจุดอิ่มตัว และวงการจำเป็นต้องหาแนวทางใหม่ในการพัฒนา AI ที่ชาญฉลาดขึ้นโดยไม่พึ่งการขยายขนาดเพียงอย่างเดียว

ผลกระทบต่อเศรษฐกิจและอุตสาหกรรม

ความท้าทายด้านการสเกลนี้ส่งผลสองด้าน ด้านหนึ่ง การสร้างโมเดลระดับแนวหน้าที่มีความสามารถในการให้เหตุผลสูง (เช่น GPT-4) ต้องการการลงทุนมหาศาล ทั้งในแง่ค่าใช้จ่ายการประมวลผล (เช่น การ์ดประมวลผล GPU จำนวนมากและไฟฟ้าจำนวนมหาศาล) และข้อมูลปริมาณมหาศาล บริษัทขนาดใหญ่ไม่กี่รายเท่านั้นที่มีทรัพยากรจะทำได้ ซึ่งอาจนำไปสู่การผูกขาดทางเทคโนโลยีและความเหลื่อมล้ำ (บริษัทเล็กหรือแม้แต่ประเทศกำลังพัฒนาอาจไม่สามารถตามทัน)

อีกด้านหนึ่ง เมื่อแนวทาง Scaling Up เริ่มถึงขีดจำกัดทำให้ทั้งภาคอุตสาหกรรมและนักวิจัยเริ่มมองหาวิธีใหม่ที่มีประสิทธิภาพกว่า ซึ่งหากพบแนวทางที่ดีกว่า ก็จะลดต้นทุนการพัฒนา AI ลงอย่างมากและเปิดโอกาสให้ผู้เล่นรายใหม่ ๆ เข้ามาแข่งขันได้ ส่งผลดีต่อเศรษฐกิจภาพรวมในระยะยาว

แนวทางใหม่ในการพัฒนา AI

นอกจากการเพิ่มขนาดโมเดลดิบ ๆ หนึ่งในนั้นคือการให้โมเดลได้ใช้พลังประมวลผลมากขึ้นในขั้นตอนการเรียกใช้งาน (Inference) แทนที่จะทุ่มไปในการฝึกโมเดลให้ใหญ่ขึ้นเพียงอย่างเดียว

กล่าวคือ ให้ AI "ใช้เวลาคิดนานขึ้น" เมื่อเจอปัญหาซับซ้อน แทนที่จะตอบทันทีแบบโมเดลปัจจุบัน วิธีนี้บางครั้งเรียกว่า "Test-time Compute" หรือ การ 'คำนวณเพิ่ม' ระหว่างการใช้งานจริง

ยกตัวอย่างเช่น แทนที่โมเดลจะสร้างคำตอบมาเพียงคำตอบเดียวทันทีที่รับคำถาม โมเดลอาจลองสร้างวิธีแก้ปัญหาหลาย ๆ แบบภายใน (หลายคำตอบหรือหลายขั้นตอนการคิด) แล้วนำมาประเมิน จากนั้นเลือกคำตอบที่เหมาะสมที่สุดเป็นคำตอบสุดท้ายเพื่อมอบให้ผู้ใช้

แนวทางนี้คล้ายกับการที่มนุษย์ใช้เวลาขบคิดปัญหา และอาจทำให้โมเดลสามารถแก้โจทย์ที่ต้องใช้เหตุผลหลายขั้นตอนได้ดีขึ้น โดยไม่จำเป็นต้องเพิ่มขนาดโมเดลเลย นักวิจัยหลายสถาบันเริ่มทดลองใช้วิธีนี้ในการพัฒนาโมเดลรุ่นใหม่ เช่น OpenAI ได้พัฒนาโมเดลชุดหนึ่งชื่อรหัสว่า "o1"

ซึ่งออกแบบมาให้สามารถคิดแบบหลายขั้นตอนในระหว่างตอบคำถาม (multi-step reasoning) และใช้ข้อมูลที่คัดสรรมาจากผู้เชี่ยวชาญมาช่วยสอนเพิ่มเติมเพื่อเพิ่มความสามารถในการให้เหตุผลเชิงลึก

ผลลัพธ์คือโมเดลขนาดไม่ใหญ่มากแต่มีความสามารถแก้ปัญหายาก ๆ ได้ดีขึ้น

แนวทางดังกล่าวถูกมองว่าจะเปลี่ยนสมการการแข่งขันของวงการ AI เนื่องจากไม่จำเป็นต้องพึ่งพาชิปประมวลผลหรือพลังงานมหาศาลแบบเดิมอีกต่อไป หากโมเดลสามารถฉลาดขึ้นด้วยวิธีอื่น บริษัทต่าง ๆ ก็อาจหันมาแข่งกันที่คุณภาพของอัลกอริทึมและเทคนิคการฝึกมากกว่าแค่ขนาดของทรัพยากรเท่านั้น

นอกจากเรื่องการใช้เวลา inference แล้ว ยังมีแนวทางอื่น ๆ เช่น การใช้ข้อมูลสังเคราะห์เพิ่มเติมในการฝึก (เพื่อแก้ปัญหาขาดแคลนข้อมูลจริง) หรือ การปรับปรุงสถาปัตยกรรมโมเดลให้มีประสิทธิภาพการเรียนรู้สูงขึ้นโดยไม่ต้องเพิ่มพารามิเตอร์ (เช่นโมเดลที่จำลองการคิดเป็นขั้นตอน)

ตัวอย่างและแนวโน้มอนาคต

Ilya Sutskever ผู้ร่วมก่อตั้ง OpenAI และ Safe SuperIntelligence ซึ่งเคยเป็นผู้เสนอแนวคิดการ “เพิ่มขนาดแล้วจะสำเร็จ” ตั้งแต่ยุคแรก ๆ ก็ออกมายอมรับกับ Reuters เมื่อไม่นานมานี้ว่าผลลัพธ์จากการเพิ่มขนาดโมเดลด้วยวิธีการเดิมเริ่มถึงจุดอิ่มตัวแล้ว และเราจำเป็นต้องกลับมาสู่ “ยุคแห่งความมหัศจรรย์และการค้นพบ” อีกครั้ง (age of wonder and discovery once again) ซึ่งก็คือการค้นหาไอเดียใหม่ ๆ ในการสร้าง AI ที่ชาญฉลาดยิ่งขึ้น คำกล่าวอดีตของเขาที่ว่า “ถ้าคุณมีชุดข้อมูลขนาดใหญ่และคุณฝึกเครือข่ายประสาทที่ใหญ่มาก ๆ ความสำเร็จก็รับประกัน” อาจไม่จริงอีกต่อไปในทศวรรษนี้[10]

การเปลี่ยนแปลงนี้หมายถึงอนาคตของ AI จะไม่ได้ขับเคลื่อนด้วยขนาดเพียงอย่างเดียว แต่จะเน้นที่ประสิทธิภาพเชิงคุณภาพ - ในอนาคตอันใกล้ เราน่าจะเห็นการผสมผสานแนวทางต่าง ๆ เช่น โมเดลขนาดกลางที่ผ่านการสอนให้คิดแบบหลายขั้นตอน ใช้ข้อมูลจริงผสมกับข้อมูลสังเคราะห์ และอาจมีการเลือกปรับใช้โมเดลย่อยที่เชี่ยวชาญหลายตัวร่วมกันแทนการใช้โมเดลยักษ์ตัวเดียว (mixture of experts)[11]

สิ่งเหล่านี้จะช่วยลดภาระด้านทรัพยากรและพลังงานอย่างมาก หากประสบความสำเร็จย่อมส่งผลดีทั้งด้านเศรษฐกิจและสิ่งแวดล้อม อีกทั้งเปิดทางให้ผู้เล่นหน้าใหม่ ๆ สามารถพัฒนาโมเดลอัจฉริยะได้โดยไม่ถูกกีดกันด้วยกำแพงค่าใช้จ่ายมหาศาลเหมือนที่ผ่านมา

นี่คือสามเทรนด์หลักของ AI ในยุคที่กำลังจะมาถึง เทรนด์เหล่านี้สะท้อนให้เห็นทิศทางของวงการ AI ที่พยายามขยายขีดความสามารถของเทคโนโลยีให้รับและประมวลผลข้อมูลหลายรูปมากขึ้น ขนาดเล็กลง ฉลาดขึ้น และเข้าถึงได้มากขึ้น โมเดลมัลติโหมดมุ่งให้ AI เข้าใจโลกแบบองค์รวมผ่านข้อมูลหลากหลายประเภท เสริมศักยภาพการใช้งานในสถานการณ์จริงที่ซับซ้อน โมเดลขนาดเล็กเฉพาะทางมุ่งให้การนำ AI ไปใช้แพร่หลายขึ้น ลดข้อจำกัดด้านทรัพยากรและเปิดโอกาสให้ทุกภาคส่วนเข้าถึง AI ที่เหมาะสมกับงานของตนได้ง่าย ส่วนการสเกลการประมวลของโมเดลในช่วง inference เพื่อการให้เหตุผลที่ดีขึ้นก็สะท้อนความพยายามที่จะผลักเพดานความสามารถของ AI ให้สูงขึ้นอีกขั้น โดยไม่ถูกจำกัดด้วยวิธีการแบบเดิม ๆ

ทั้งสามแนวโน้มนี้ไม่ได้พัฒนาแยกจากกัน แต่เกื้อหนุนกันในการสร้าง AI แห่งอนาคตที่ทั้งเก่งรอบด้าน, ประหยัดทรัพยากร, และสามารถให้เหตุผลแก้ปัญหายาก ๆ ได้ดีขึ้น หากเราติดตามและปรับตัวให้ทันต่อเทรนด์เหล่านี้ ธุรกิจและเศรษฐกิจโดยรวมจะสามารถเก็บเกี่ยวประโยชน์จาก AI ได้อย่างเต็มที่ ขณะเดียวกันก็เตรียมพร้อมรับมือกับความท้าทายใหม่ ๆ ที่จะตามมาไม่ว่าจะเป็น

1. การใช้งาน AI ในประเด็นด้านการกำกับดูแลและจริยธรรม
2. การพัฒนา AI ภายใต้ข้อจำกัดด้านข้อมูลและทรัพยากรที่ใช้ในการพัฒนา
3. ผลกระทบทางเศรษฐกิจและสังคมที่เกิดจาก AI

ACT III:
AI at Your Fingertips.

Agentic AI: Rise of the Agents

ช่วงสองปีที่ผ่านมา เราได้เห็นความก้าวหน้าอย่างรวดเร็วของเทคโนโลยี AI โดยเฉพาะ Large Language Models หรือ LLMs ที่มาปฏิวัติวิธีการที่มนุษย์ปฏิสัมพันธ์กับคอมพิวเตอร์ แต่ AI ที่เราคุ้นเคยกัน ส่วนใหญ่ยังอยู่ในรูปแบบของ Generative AI ซึ่งมีความสามารถหลักในการสร้างเนื้อหาตามคำสั่งที่ได้รับจากมนุษย์

แต่ทุกอย่างกำลังทรงพลังขึ้นกว่าเดิม หลังการมาถึงของ Agentic AI ซึ่งสามารถรับรู้ ให้เหตุผล วางแผน และดำเนินการตามแผนโดยใช้เครื่องมือต่าง ๆ ได้อย่างเป็นอิสระ เสมือนความต่างระหว่างพนักงานที่คอยรับคำสั่งกับผู้ช่วยที่เข้าใจเป้าหมายและสามารถตัดสินใจได้เองในระดับหนึ่ง

องค์กรระดับโลกอย่าง World Economic Forum และสถาบันชั้นนำต่าง ๆ ทั้ง Stanford HAI[1], MIT Sloan[2] และผู้เชี่ยวชาญในอุตสาหกรรมต่างให้ความเห็นตรงกันว่า Agentic AI จะเป็นเทรนด์สำคัญที่จะครองตลาด AI ในปี 2025 โดยได้รับการสนับสนุนจากผู้นำในอุตสาหกรรม สถาบันการศึกษา บริษัทเทคโนโลยียักษ์ใหญ่ และสตาร์ทอัพ

ในงาน World Economic Forum 2025 ที่ดาวอส มีการพูดคุยกันอย่างกว้างขวางเกี่ยวกับทิศทางของ AI โดยตัวแทนจากหลากหลายภูมิภาคได้ร่วมหารือกับผู้นำโลกและผู้กำหนดนโยบายเกี่ยวกับแนวโน้ม AI

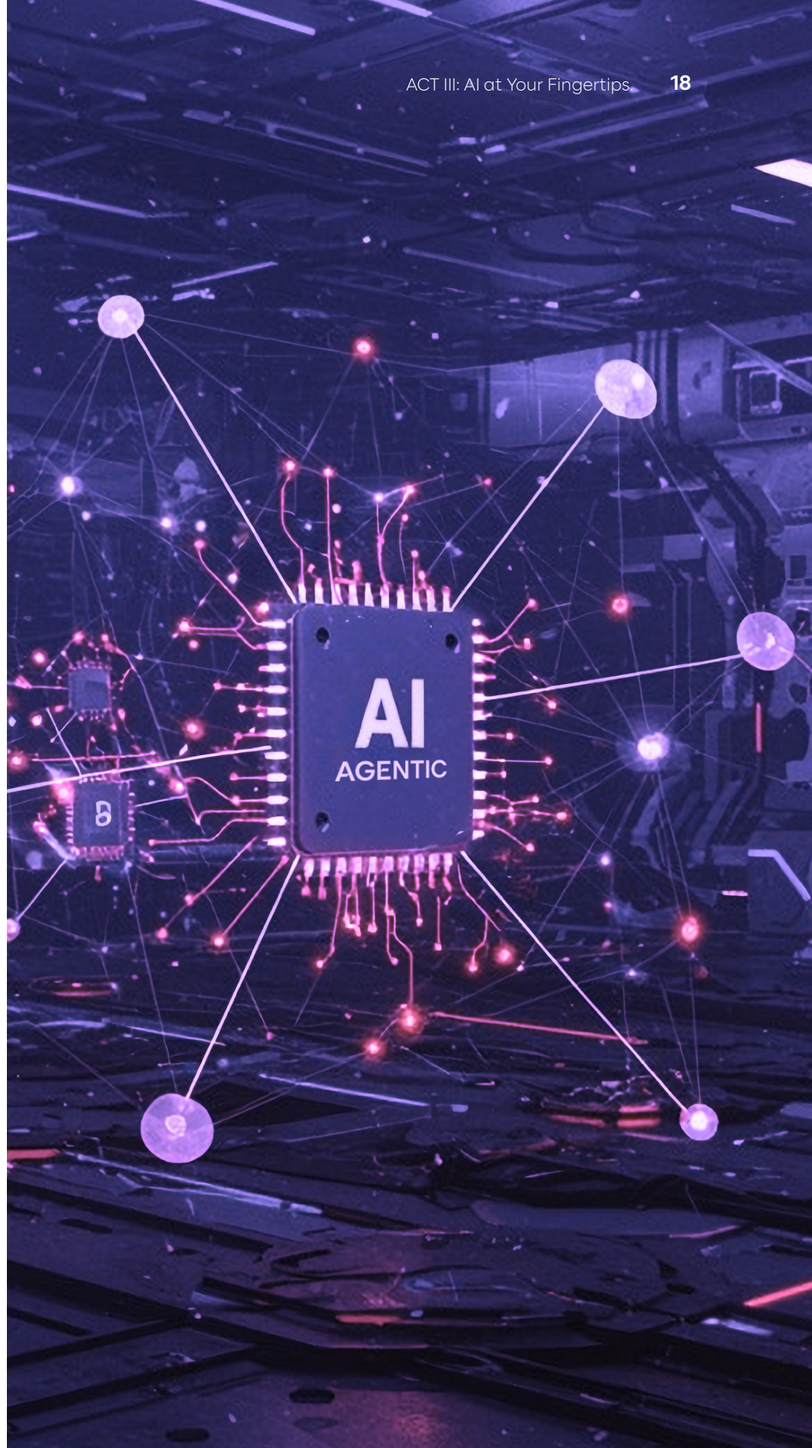
เป็นการสนทนาที่มุ่งเน้นไปยังการประยุกต์ใช้งานในโลกความเป็นจริงมากกว่าเพียงแค่ศักยภาพของมัน

ในขณะเดียวกัน MIT Sloan Management Review ก็ได้นำเสนอบทความ “Five Trends in AI and Data Science for 2025” โดย Thomas H. Davenport และ Randy Bean ซึ่งมีใจความว่า ‘แนวโน้ม Agentic AI ในปี 2025 สมควรได้รับความสนใจอย่างใกล้ชิดจากผู้นำ’ เน้นย้ำว่า Agentic AI เป็นหนึ่งในเทรนด์สำคัญที่สุดที่ผู้นำองค์กรควรให้ความสนใจในปี 2025

ความแตกต่างสำคัญอีกประการหนึ่งคือวิธีการตอบสนองต่อความไม่แน่นอนหรือข้อมูลใหม่ Gen AI มักต้องการให้มนุษย์ป้อนข้อมูลใหม่หรือปรับคำสั่ง ในขณะที่ Agentic AI สามารถปรับตัวและหาวิธีแก้ปัญหาใหม่ ๆ ได้เองในระดับหนึ่ง

ตัวอย่างเช่น ในการวางแผนการลงทุน Gen AI อาจสามารถให้คำแนะนำทั่วไปเกี่ยวกับกลยุทธ์การลงทุนตามข้อมูลที่ได้รับ แต่หากตลาดมีความผันผวนหรือมีข่าวสำคัญที่ส่งผลกระทบ มนุษย์จะต้องป้อนข้อมูลใหม่เพื่อให้ได้คำแนะนำที่ปรับปรุงแล้ว ในขณะที่ Agentic AI สามารถติดตามข้อมูลตลาดแบบเรียลไทม์ วิเคราะห์ผลกระทบของข่าวสาร และปรับเปลี่ยนคำแนะนำโดยอัตโนมัติ

ความแตกต่างนี้สำคัญอย่างยิ่งในบริบทของภาคการเงินการธนาคาร ซึ่งต้องการการตัดสินใจที่รวดเร็ว แม่นยำ และสามารถปรับเปลี่ยนได้ตามสถานการณ์ที่เปลี่ยนแปลงอย่างรวดเร็ว



Gen AI (Generative AI) ทำงานในลักษณะนี้

รับคำสั่ง
รับภารกิจหรือวัตถุประสงค์
จากมนุษย์

→

รับรู้คำสั่ง
ประมวลผลข้อมูลนำเข้าเพื่อเข้าใจบริบท
และรวบรวมข้อมูลที่เกี่ยวข้อง (หากจำเป็น)

→

สร้างคำตอบ
สร้างการตอบสนองที่เกี่ยวข้อง
โดยใช้โมเดลที่ผ่านการฝึกฝนมาแล้ว

รอการป้อนคำสั่งเพิ่มเติมจากมนุษย์
มนุษย์ตีความผลลัพธ์แล้วสร้างคำสั่งใหม่เพื่อให้ทำงานต่อ

กระบวนการทำงานของ Gen AI จึงเป็นวงจรที่ต้องพึ่งพามนุษย์อย่างมาก มนุษย์ต้องเป็นผู้ตีความผลลัพธ์ที่ได้และจัดทำคำสั่งใหม่หากต้องการดำเนินการต่อ ทำให้มีข้อจำกัดในการทำงานที่ต้องการการตัดสินใจหลายขั้นตอนหรือต้องดำเนินการอย่างต่อเนื่อง เช่น การทำธุรกรรมทางการเงินที่ซับซ้อน หรือการจัดการข้อมูลจากหลายแหล่ง

ในขณะที่ Agentic AI มีกระบวนการทำงานที่ซับซ้อนและเป็นอิสระมากกว่า

รับคำสั่ง
รับภารกิจหรือวัตถุประสงค์จากมนุษย์
(เหมือนกับ Gen AI)

→

รับรู้และให้เหตุผล
ประมวลผลข้อมูลนำเข้าเพื่อเข้าใจบริบท
และรวบรวมข้อมูลที่เกี่ยวข้องจากแหล่งต่าง ๆ โดยอัตโนมัติ

→

วางแผนและประสานงาน
เข้าใจ ประสานงาน และวางแผนงาน
เพื่อสร้างผลลัพธ์ที่เป็นประโยชน์

→

ลงมือทำ
ดำเนินการตามแผนเพื่อบรรลุเป้าหมาย
โดยใช้เครื่องมือต่าง ๆ เช่น APIs

เรียนรู้อย่างต่อเนื่อง
ปรับตัวอย่างต่อเนื่องโดยอาศัยการตอบกลับจากสภาพแวดล้อม มนุษย์ และกระบวนการทำงานแบบ agent เพื่อปรับปรุงการตอบสนองในอนาคต

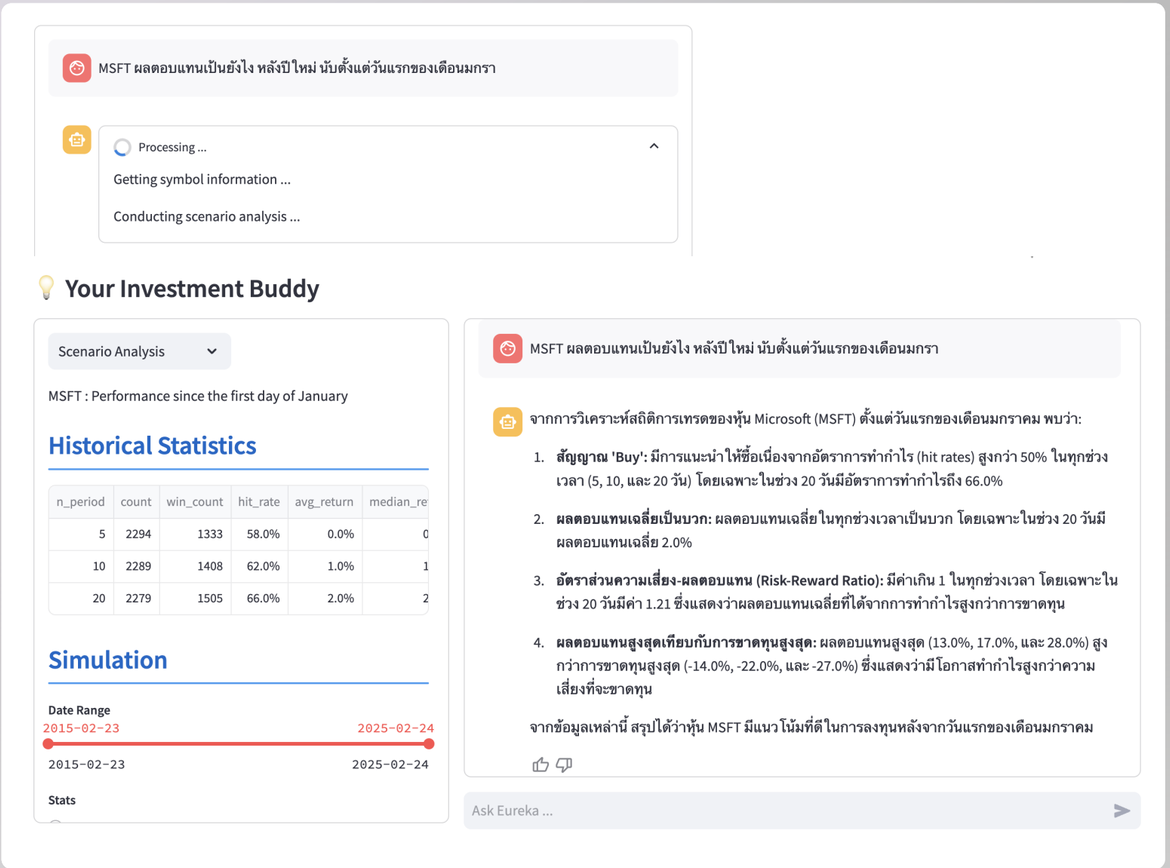
กระบวนการของ Agentic AI จึงมีความ 'คล้ายมนุษย์' มากกว่า เนื่องจากสามารถรับข้อมูล คิดวิเคราะห์ วางแผน และดำเนินการได้เองโดยไม่ต้องรอคำสั่งในทุกขั้นตอน นอกจากนี้ ยังมีวงจรการเรียนรู้ที่ต่อเนื่องที่ช่วยให้ระบบปรับปรุงตัวเองได้จากประสบการณ์

SCBX Agentic AI Financial Advisory Platform

SCBX เห็นศักยภาพของ Agentic AI และนำมาพัฒนาเป็นแพลตฟอร์มให้คำปรึกษาทางการลงทุน ที่สามารถแสดงให้เห็นถึงข้อได้เปรียบอย่างชัดเจน

A Gen AI Approach

SCBX Agentic AI Approach



แพลตฟอร์มดังกล่าวเป็นระบบให้คำปรึกษาด้านการลงทุนที่ขับเคลื่อนด้วย Agentic AI โดยเมื่อเปรียบเทียบกับการใช้ Gen AI อย่าง ChatGPT หรือ Claude 3.7 Sonnet ในการวิเคราะห์หุ้น จะเห็นความแตกต่างอย่างชัดเจน

SCBX ได้ทดสอบประสิทธิภาพของทั้งสองระบบในการวิเคราะห์หุ้น Microsoft (MSFT) โดยให้ทั้ง Gen AI ทั่วไปและ Agentic AI ของ SCBX ตอบคำถามเดียวกันเกี่ยวกับการลงทุนในหุ้น MSFT ผลลัพธ์แสดงให้เห็นความแตกต่างอย่างมีนัยสำคัญในคุณภาพของการวิเคราะห์และความสามารถในการให้คำแนะนำ

- Gen AI แบบทั่วไป เมื่อถูกขอให้วิเคราะห์หุ้น MSFT (Microsoft) จะเพียงแค่
  - ค้นหาข้อมูลบนเว็บแต่ได้ข้อมูลไม่ครบถ้วน
  - ไม่สามารถให้คำแนะนำที่เป็นรูปธรรมได้
  - มีความรู้จำกัดและไม่อัปเดต
  - ไม่สามารถทำการวิเคราะห์เชิงลึกได้

ผลลัพธ์ที่ได้จาก Gen AI จึงมักจะเป็นการให้ข้อมูลทั่วไปที่ไม่ได้มีการวิเคราะห์เชิงลึก เช่น “หุ้น Microsoft มีผลประกอบการที่ดีในไตรมาสล่าสุด” หรือ “Microsoft เป็นบริษัทที่มีพื้นฐานแข็งแกร่ง” ซึ่งเป็นข้อมูลที่ไม่เพียงพอต่อการตัดสินใจลงทุน นอกจากนี้ ข้อมูลที่ได้อาจไม่เป็นปัจจุบันและขาดบริบทที่สำคัญสำหรับการตัดสินใจลงทุน

ในทางกลับกัน Agentic AI Approach ของ SCBX มีกระบวนการทำงานที่ซับซ้อนและให้ผลลัพธ์ที่มีคุณค่ามากกว่า

รับและตีความข้อความ

เข้าใจคำถามหรือข้อมูลที่ต้องการอย่างถูกต้อง เช่น การวิเคราะห์หุ้น MSFT

ให้เหตุผลและวางแผน

ออกแบบการวิเคราะห์ที่เหมาะสม โดยพิจารณาว่าต้องดึงข้อมูลอะไรบ้าง เช่น ข้อมูลราคาย้อนหลัง อัตราส่วนทางการเงิน การวิเคราะห์แนวโน้มตลาด และข่าวล่าสุดที่อาจส่งผลกระทบ

เรียกใช้เครื่องมือเพื่อการวิเคราะห์

ใช้ API เครื่องมือทางการเงินต่าง ๆ เช่น การวิเคราะห์งบการเงิน การวิเคราะห์สัญญาณทางเทคนิค เครื่องมือสแกนหุ้น และเครื่องมืออื่น ๆ ที่เกี่ยวข้อง เพื่อรวบรวมข้อมูลสำหรับการวิเคราะห์และการตัดสินใจเชิงกลยุทธ์ โดยจัดทำบทวิเคราะห์สรุปที่กระชับและนำเสนอมุมมองของ InnovestX (บริษัทหลักทรัพย์ภายใต้กลุ่ม SCBX) อย่างชัดเจน

แสดงผลข้อมูลที่ได้รับการอัปเดตล่าสุด

นำเสนอข้อมูลในรูปแบบที่เข้าใจง่ายและมีประโยชน์ เช่น

- กราฟแสดงผลการดำเนินงานตั้งแต่ต้นปี
- ตารางสถิติแสดงอัตราชนะในช่วงเวลาต่าง ๆ (5, 10, 20 วัน)
- ผลตอบแทนเฉลี่ยและมัธยฐาน
- อัตราส่วนความเสี่ยงต่อผลตอบแทน
- การจำลองผลตอบแทนที่เป็นไปได้ในอนาคต

สรุปเชิงลึกพร้อมคำแนะนำ

ให้มุมมองของผู้เชี่ยวชาญจาก InnovestX พร้อมข้อเสนอแนะเชิงปฏิบัติ โดยพิจารณาจากข้อมูลที่วิเคราะห์ได้ทั้งหมด

ตัวอย่างเช่น เมื่อถามเกี่ยวกับหุ้น MSFT แทนที่จะได้คำตอบทั่วไปว่า “Microsoft เป็นบริษัทที่น่าลงทุน” ลูกค้าจะได้รับข้อมูลเชิงลึกที่มีรายละเอียดเพื่อให้สามารถนำไปวิเคราะห์ต่อได้ เพื่อให้การตัดสินใจลงทุนมีประสิทธิภาพมากยิ่งขึ้น เช่น “จากการวิเคราะห์พบว่า MSFT มีอัตราชนะ 58% ในระยะเวลา 5 วัน และ 60% ในระยะเวลา 20 วัน โดยมีอัตราส่วนความเสี่ยงต่อผลตอบแทนที่ 1.21 ซึ่งแสดงว่าผลตอบแทนที่คาดหวังสูงกว่าความเสี่ยง”

แสดงให้เห็นถึงความแตกต่างอย่างมีนัยสำคัญในคุณภาพของคำแนะนำและความสามารถในการจัดการกับงานที่ซับซ้อนระหว่าง Gen AI แบบดั้งเดิมกับ Agentic AI แบบล้ำสมัย โดยลูกค้าของ SCBX อาทิ ลูกค้าจากบริษัทหลักทรัพย์ อินโนเวสท์ เอกซ์ จำกัด (InnovestX) จะได้รับประโยชน์จากการวิเคราะห์ที่ครอบคลุม แม่นยำ และเป็นปัจจุบัน ซึ่งช่วยให้สามารถตัดสินใจลงทุนได้อย่างมั่นใจมากขึ้น

อย่างไรก็ตาม การลงทุนมีความเสี่ยง ผู้ลงทุนควรศึกษาข้อมูลก่อนตัดสินใจลงทุน

ACT IV:
Not Quite Human, But Almost There.

Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

ในยุคที่ปัญญาประดิษฐ์ (AI) กำลังเติบโตและพัฒนาอย่างก้าวกระโดด คำว่า 'AGI' หรือ Artificial General Intelligence เริ่มถูกกล่าวถึงมากขึ้นในวงการเทคโนโลยี สิ่งนี้เสมือนฝันสูงสุดของนักพัฒนา AI แต่ในตอนนี้ความเข้าใจเกี่ยวกับ AGI ยังคงมีความหลากหลายและยังไม่มีข้อสรุปที่ชัดเจน

AGI คืออะไร?

ARTIFICIAL NARROW INTELLIGENCE

IDEA

Machine's ability to perform a single task extremely well, even better than humans.

VS

ARTIFICIAL GENERAL INTELLIGENCE

IDEA

Machines can be made to think and function as the human mind.

Artificial General Intelligence (AGI) หมายถึงปัญญาประดิษฐ์ที่คิดและเรียนรู้ และปฏิบัติงานใด ๆ ก็ตามในระดับที่เทียบเท่ามนุษย์ นิยามนี้ดูเรียบง่าย แต่ซ่อนความซับซ้อนไว้มาก เพราะการจะระบุว่าอะไรคือ "ความสามารถทางปัญญาระดับมนุษย์" นั้นเป็นคำถามที่นักวิทยาศาสตร์ นักปรัชญา และผู้เชี่ยวชาญด้าน AI ยังคงถกเถียงกันอยู่

ความแตกต่างที่สำคัญระหว่าง AI ที่มีอยู่ในปัจจุบันกับ AGI คือขอบเขตและความยืดหยุ่นของความสามารถ

AI ในปัจจุบัน แม้จะมีความซับซ้อนและทรงพลัง แต่ส่วนใหญ่ยังเป็นระบบ Narrow AI ที่ถูกออกแบบมาเพื่อทำงานเฉพาะด้านได้ดี เช่น การรู้จักและจดจำภาพ การแปลภาษา การให้ข้อมูล หรือการวิเคราะห์ข้อมูลทางการเงิน

ในขณะที่ AGI จะต้องมีความสามารถที่หลากหลายมาผสมกัน และสามารถผสานความรู้จากด้านหนึ่งไปใช้กับอีกด้านหนึ่งได้ และที่สำคัญคือการปรับตัวเข้ากับสถานการณ์ใหม่ ๆ ได้โดยไม่ต้องผ่านการฝึกฝนเพิ่มเติม คล้ายกับวิธีที่มนุษย์เรียนรู้และปรับตัว

การเปรียบเทียบความสามารถของมนุษย์กับเครื่องจักรนั้นเป็นเรื่องที่ซับซ้อน เพราะสมองมนุษย์ทำงานในลักษณะที่แตกต่างจากคอมพิวเตอร์อย่างสิ้นเชิง มนุษย์มีความเข้าใจในบริบทของเหตุการณ์ รวมถึง ความคิดสร้างสรรค์และการตัดสินใจบนพื้นฐานของค่านิยมและอารมณ์ความรู้สึก ซึ่งยากที่จะจำลองสิ่งนี้ระบบคอมพิวเตอร์

เสียงที่แตกต่างจากผู้นำในอุตสาหกรรม AI

ในวงการ AI ปัจจุบัน ผู้นำและผู้เชี่ยวชาญ มีมุมมองที่แตกต่างกันอย่างมาก เกี่ยวกับกรอบเวลาและความเป็นไปได้ในการพัฒนา AGI สิ่งนี้สะท้อนให้เห็นถึงความไม่แน่นอนและความท้าทายในการคาดเดาถึงเทคโนโลยีที่มีความซับซ้อนและอาจเปลี่ยนแปลงโลกเช่นนี้

"

กล่าวว่า เขาน่าจะก้าวถึง AGI เร็วกว่าที่คนส่วนใหญ่ในโลกคิด ซึ่งความเชื่อมั่นนี้ สอดคล้องกับหลักของ OpenAI ที่ตั้งเป้าหมายในการพัฒนา AGI ที่ปลอดภัยและเป็นประโยชน์[1]

Sam Altman
CEO of OpenAI

เชื่อว่า การพัฒนา AGI อาจใช้เวลานานกว่าที่คาดการณ์ไว้ - อาจถึง 10 ปี[2]

Mustafa Suleyman
CEO of Microsoft AI

แสดงความเชื่อมั่นใน AGI ที่ฉลาดกว่ามนุษย์จะเกิดขึ้นใน 5-10 ปีข้างหน้า ซึ่งสิ่งนี้จะมีศักยภาพในการยกระดับไปจนถึงปฏิวัติวงการต่าง ๆ เช่น การรักษาโรค แก้วิกฤติโลกร้อน[3]

Sir Demis Hassabis
หัวหน้าของ DeepMind
(บริษัทในเครือ Google)

กรอบการประเมินระดับขั้นของ AGI

การพัฒนา AGI ไม่ใช่การก้าวกระโดดจากระบบปัจจุบัน ไปสู่ระบบที่มีความสามารถเทียบเท่ามนุษย์ในทันที แต่เป็นการพัฒนาแบบต่อเนื่อง ผ่านระดับขั้นต่าง ๆ นักวิจัยทั้งจากภาควิชาการและอุตสาหกรรมจึงได้ร่วมมือกันพัฒนากรอบการประเมินระดับขั้นของ AGI เพื่อให้สามารถติดตามความก้าวหน้า และกำหนดเป้าหมายในการพัฒนา ได้ชัดเจนยิ่งขึ้น กรอบการประเมินที่ได้รับการพัฒนาขึ้นแบ่งระดับขั้นของ AI ออกเป็นหลายระดับ ตั้งแต่ระดับ 0 (No AI) ไปจนถึงระดับ 5 (Superhuman) โดยแต่ละระดับมีลักษณะและความสามารถที่แตกต่างกัน[4]

ระดับ 0

No AI (ไม่มี AI)

เป็นระบบซอฟต์แวร์คำนวณทั่วไป ไม่มีคุณสมบัติของ AI เช่น เครื่องคำนวณ, คอมไพเลอร์ (Compiler-โปรแกรมที่ทำหน้าที่แปลภาษาระดับสูง) หรือในกรณีทั่วไป ก็คือระบบที่มนุษย์ทำงานเบื้องหลัง เช่น Amazon Mechanical Turk ซึ่งเป็น human-in-the-loop computing

ระดับ 1

Emerging (กำลังเกิดขึ้น)

ระบบที่เทียบเท่าหรือดีกว่ามนุษย์ที่ไม่มีทักษะเล็กน้อย ในกรณี Narrow AI คือระบบที่ใช้กฎอย่างง่าย ส่วนในกรณี General AI ได้แก่ ChatGPT (OpenAI), Llama 2 (Meta) และ Gemini (Google) ซึ่งถือเป็น Emerging AGI

ระดับ 2

Competent (มีความสามารถ/ทักษะ)

ระบบที่มีประสิทธิภาพเทียบเท่าผู้ใหญ่ ที่มีทักษะในระดับ 50 เปอร์เซ็นต์ไทล์ ในกรณี Narrow AI เช่น ผู้ช่วยเสียงอัจฉริยะ เช่น Siri (Apple), Alexa (Amazon), Google Assistant (Google) และ SOTA LLMs สำหรับงานเฉพาะทาง เช่น การเขียนเรียงความสั้นๆ หรือการเขียนโค้ดอย่างง่าย ส่วนในกรณี Competent AGI ยังไม่มีระบบใดบรรลุถึงระดับนี้

ระดับ 3

Expert (ผู้เชี่ยวชาญ)

ระบบที่มีประสิทธิภาพเทียบเท่าผู้ใหญ่ที่มีทักษะระดับ 90 เปอร์เซ็นต์ไทล์ ในกรณี Narrow AI ได้แก่ ระบบตรวจสอบการสะกดและไวยากรณ์ เช่น Grammarly, โมเดลสร้างภาพ เช่น Imagen (Google Brain) หรือ Dall-E 2 (OpenAI) ส่วนในกรณี Expert AGI ยังไม่มีระบบใดบรรลุถึง

ระดับ 4

Virtuoso (ระดับเลิศ)

ระบบที่มีประสิทธิภาพเทียบเท่าผู้ใหญ่ที่มีทักษะระดับ 99 เปอร์เซ็นต์ไทล์ ในกรณี Narrow AI ได้แก่ Deep Blue (Campbell et al., 2002) และ AlphaGo (Silver et al., 2016; 2017) ส่วนในกรณี Virtuoso AGI ยังไม่มีระบบใดบรรลุถึง

ระดับ 5

Superhuman (เหนือมนุษย์)

ระบบที่มีประสิทธิภาพเหนือกว่ามนุษย์ 100% ในกรณี Narrow AI ได้แก่ AlphaFold และ AlphaZero (Google DeepMind) ส่วนในกรณี ASI (Artificial Superintelligence) ยังไม่มีระบบใดบรรลุถึง

หมายเหตุ : Narrow AI = AI ทำงานเฉพาะด้าน เช่น เล่นหมากรุกหรือแนะนำสินค้า ทำงานได้ดีเฉพาะในขอบเขตที่กำหนดเท่านั้น
General AI = AI ระดับมนุษย์ที่เรียนรู้และแก้ปัญหาได้หลากหลาย

เส้นทางสู่ AGI ที่แท้จริง

อียาน อ็องเดร เลอเกิง (Yann LeCun) VP & Chief AI Scientist at Meta - Facebook ได้แบ่งปันคำแนะนำ AGI ที่แท้จริง ในงาน AI Action Summit 2025[5] ไว้ว่า

การพัฒนาไปสู่ AGI ที่แท้จริง ไม่ใช่เพียงแค่การปรับปรุงโมเดลที่มีอยู่ในปัจจุบัน ให้มีประสิทธิภาพมากขึ้นเท่านั้น แต่อาจต้องการการเปลี่ยนแปลงกระบวนความคิด (Paradigm) ครั้งใหญ่ในการออกแบบระบบ AI โดยเขามีคำแนะนำดังนี้

1 Abandon Generative Models

ทิ้งการใช้ AI ที่สร้างข้อความ ภาพ หรือเสียงจากข้อมูลที่เรียนรู้มา (เช่น GPT, Stable Diffusion) แทนที่จะมุ่งพัฒนาโมเดลประเภทนี้ ผู้เชี่ยวชาญแนะนำให้หันไปสู่ joint-embedding architectures ซึ่งเน้นการเข้าใจความสัมพันธ์ระหว่างข้อมูลในรูปแบบที่ต่างกันไปในระดับลงลึกถึงความหมายของข้อมูลนั้น ๆ แทนการสร้างเนื้อหาเพียงอย่างเดียว

2 Abandon Probabilistic Model

โมเดลที่เน้นความน่าจะเป็น ทำงานโดยคำนวณโอกาสที่เหตุการณ์จะเกิดขึ้น เช่น โอกาสที่คำใดจะตามหลังประโยคที่กำหนด (เช่นระบบเดาคำ) โดยเสนอให้ใช้โมเดลที่ใช้พลังงานเป็นฐาน (energy-based models) ซึ่งประเมินสถานะของการทำงานในรูปแบบของพลังงาน สิ่งนี้ช่วยให้เข้าใจโครงสร้างข้อมูลที่ซับซ้อนได้ดีกว่า และอาจเลียนแบบการทำงานของสมองมนุษย์ได้ใกล้เคียงกว่า

3 Abandon Contrastive Methods

วิธีการเปรียบเทียบความแตกต่าง เป็นเทคนิคที่สอนให้ AI แยกแยะข้อมูลโดยเปรียบเทียบตัว ข้อเสนอคือให้ใช้วิธีการที่มีการควบคุม (regularized methods) ซึ่งเน้นการปรับปรุงโมเดลให้ทั่วถึงและมีเสถียรภาพมากขึ้น เพื่อป้องกันการเรียนรู้ที่เอนเอียง หรือขาดความสมบูรณ์

4 Abandon Reinforcement Learning

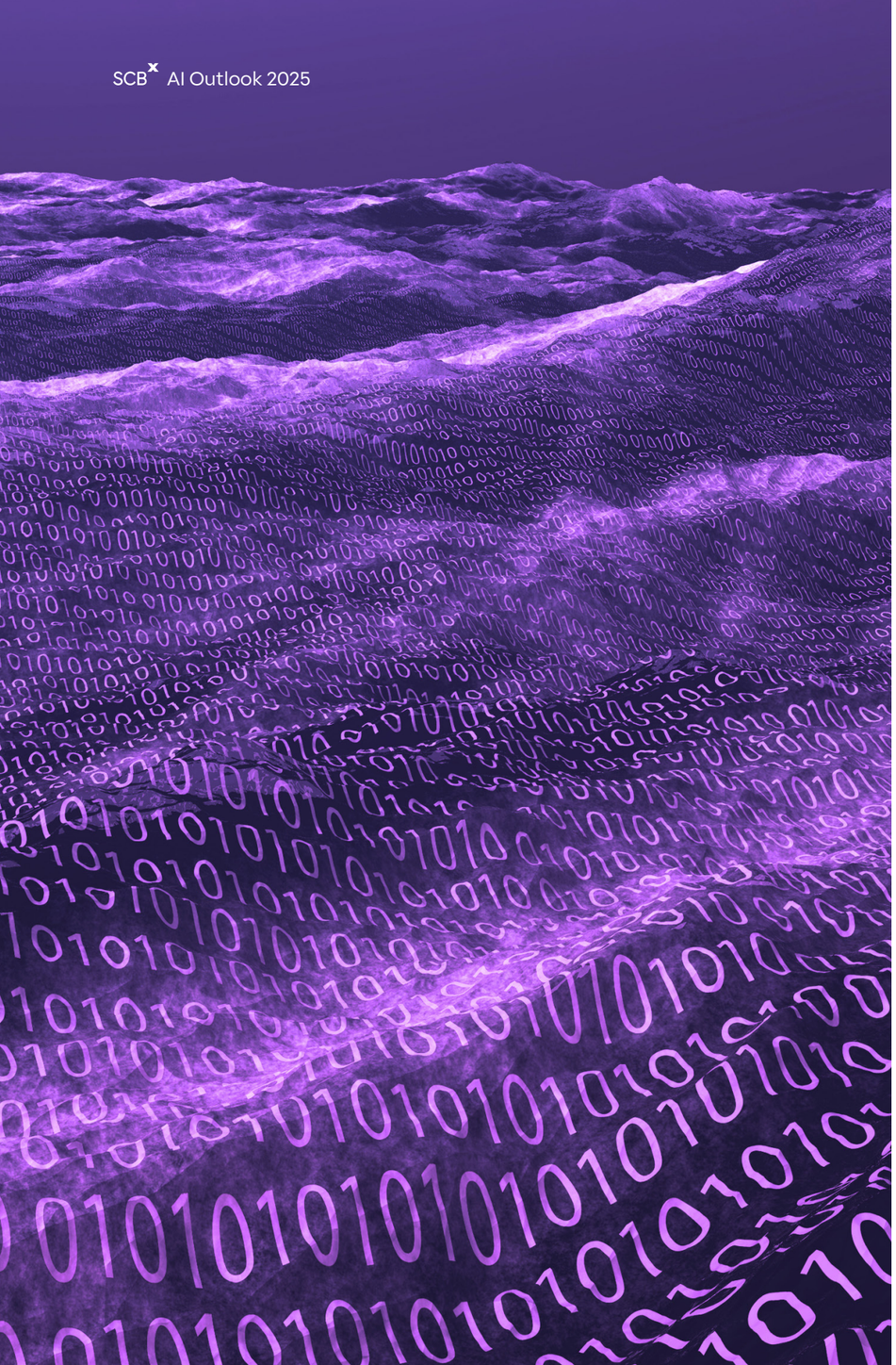
การเรียนรู้แบบเสริมกำลังเป็นวิธีที่ AI เรียนรู้จากการได้รับรางวัลเมื่อทำถูกและถูกลงโทษเมื่อทำผิด คำแนะนำคือให้เน้นการใช้ model-predictive control ซึ่งเป็นเทคนิคควบคุมที่ใช้โมเดลเพื่อทำนายผลลัพธ์ของการกระทำในอนาคต แล้วค้นหาลำดับของการกระทำที่ดีที่สุดในเวลานั้นเป็นหลัก และใช้ Reinforcement Learning เฉพาะเมื่อการวางแผนไม่ได้ผลลัพธ์ตามที่คาดหวัง เพื่อปรับโมเดลโลกหรือตัววิจารณ์ให้แม่นยำยิ่งขึ้น

5 คำแนะนำสุดท้าย หากสนใจ AI ระดับมนุษย์ อย่าทำงานกับ LLMs

(IF YOU ARE INTERESTED IN HUMAN-LEVEL AI, DON'T WORK ON LLMs)

LLMs แม้จะมีความสามารถที่น่าประทับใจ แต่อาจไม่ใช่เส้นทางที่จะนำไปสู่ AGI ได้ในที่สุด เพราะ LLMs มีข้อจำกัดในการเข้าใจโลกแบบองค์รวม, การเชื่อมโยงเหตุผล และการแก้ปัญหาที่ซับซ้อน

การเดินทางสู่ AGI ยังคงเต็มไปด้วยความท้าทายและข้อถกเถียง แม้เทคโนโลยี AI จะก้าวหน้าอย่างรวดเร็ว แต่การบรรลุถึงปัญญาประดิษฐ์ที่มีความสามารถระดับมนุษย์อย่างแท้จริง ยังคงเป็นเป้าหมายที่ยังไม่มีข้อสรุปชัดเจน



EPILOGUE: The AI storm - Infinite impact. Case Studies from Inside the Eye of the Typhoon กรณีศึกษาของ SCBX

TYPHOON

กลุ่มบริษัท SCBX ได้นำ AI มาใช้งานจริงในกรณีศึกษาของ Smart RM Training เป็นระบบฝึกอบรมพนักงานผู้จัดการความสัมพันธ์ลูกค้า (บทที่ 2) และ Agentic AI Financial Advisory Platform ซึ่งเป็นแพลตฟอร์มให้การศึกษาทางด้านการลงทุน (บทที่ 3)

นอกจากนี้ SCBX นำโดย SCB 10X ได้พัฒนา 'Typhoon' ซึ่งเป็นโมเดลภาษาไทยขนาดใหญ่ (Large Language Model - LLM) ที่ออกแบบมาเพื่อเสริมศักยภาพในการประมวลผลภาษาไทยโดยเฉพาะ โดยถูกนำไปประยุกต์ใช้ในหลายภาคส่วนผ่านความร่วมมือกับองค์กรชั้นนำของประเทศ ได้แก่ โรงพยาบาลศิริราช, VISAI, และสถาบันวิจัยเพื่อการพัฒนาประเทศไทย (TDRI) บทความนี้จะนำเสนอกรณีศึกษาของแต่ละองค์กรในการนำ 'Typhoon' มาใช้เพื่อเพิ่มประสิทธิภาพและนวัตกรรมในการดำเนินงานในด้านต่างๆ ดังกรณีศึกษาดังต่อไปนี้

ด้านการจัดการสาธารณสุข :
กรณีศึกษาโรงพยาบาลศิริราชกับการยกระดับระบบสารสนเทศทางการแพทย์

โรงพยาบาลศิริราชเป็นสถาบันการแพทย์ชั้นนำที่ให้บริการทางการแพทย์ การศึกษา และการวิจัย ด้วยปริมาณข้อมูลและเอกสารทางการแพทย์ที่มหาศาล การจัดการองค์ความรู้ภายในองค์กรจึงเป็นความท้าทายสำคัญ

ด้วยเหตุนี้เอง ทางโรงพยาบาลศิริราชจึงนำ Typhoon มาใช้เพื่อลดความซับซ้อนของการบริหารเอกสารที่สำคัญ เช่น ข้อมูลเวชระเบียนและเอกสารภายใน ซึ่งต้องอาศัยการค้นหาและเรียกใช้ข้อมูลจำนวนมากแบบเรียลไทม์ โดย Typhoon ใช้ Retrieval Augmented Generation (RAG) ในการดึงข้อมูลที่เกี่ยวข้องได้อย่างแม่นยำ

ประโยชน์ที่ได้รับจากกรณีศึกษานี้ก็คือ ลดภาระของบุคลากรทางการแพทย์ในการอ่านและจดจำข้อมูลจากเอกสารจำนวนมาก, เพิ่มความรวดเร็วในการเข้าถึงข้อมูลสำคัญ, ลดเวลาในการตัดสินใจ และช่วยให้กระบวนการดูแลผู้ป่วยมีประสิทธิภาพและแม่นยำมากขึ้น

ในอนาคต ทางโรงพยาบาลมีความตั้งใจที่จะขยายการใช้งาน Typhoon เพื่อช่วยสนับสนุนการดูแลผู้ป่วยโดยยังคงรักษาความเป็นส่วนตัวของข้อมูล และใช้เทคโนโลยีของ Typhoon ในการจัดการเอกสารที่ถูกสแกนเพื่อขยายการใช้งานไปยังแผนกต่าง ๆ

ด้านกฎหมาย :
VISAI กับ AI ด้านการศึกษากฎหมาย

VISAI เป็นองค์กรที่มุ่งเน้นการพัฒนาเทคโนโลยี AI เพื่อสนับสนุนงานด้านกฎหมายประเทศไทย

องค์กรได้พัฒนาโครงการ ‘สมหมาย’ ซึ่งเป็นแช็ตบ็อตด้านกฎหมายที่ใช้ Typhoon ในการตอบคำถามเกี่ยวกับกฎหมายธุรกิจและพาณิชย์ รวมถึงการวิเคราะห์ช่องว่างของงานวิจัยด้านกฎหมาย โดยระบบสามารถรองรับเอกสารขนาดใหญ่ซึ่งให้ผลลัพธ์ที่แม่นยำเทียบเคียงระดับสากล

‘สมหมาย’ มีความสามารถในการทำความเข้าใจเนื้อหากฎหมายที่ซับซ้อน, รองรับการประมวลผลเอกสารขนาดใหญ่และให้ผลลัพธ์ที่แม่นยำ

ความสามารถของ ‘สมหมาย’ คือการค้นหาข้อมูลด้านกฎหมายธุรกิจและพาณิชย์ (Corporate and Commercial Laws-CCL) ครอบคลุม 35 สาขากฎหมาย และสนับสนุนงานวิจัยด้านกฎหมายและช่วยให้การเปรียบเทียบระบบกฎหมาย, ศึกษาช่องว่างและความแตกต่างของระบบกฎหมายไทยกับต่างประเทศได้สะดวกขึ้น ทำให้ช่วยเพิ่มประสิทธิภาพและความแม่นยำของ AI ทางกฎหมายให้สามารถใช้งานได้ในบริบทของประเทศไทยซึ่งเป็นความก้าวหน้าครั้งสำคัญของวงการกฎหมายไทย

ด้านการวางแผนแรงงาน :
TDRI กับการวิเคราะห์ตลาดแรงงาน

Thailand Development Research Institute (TDRI) หรือ สถาบันวิจัยเพื่อการพัฒนาประเทศไทย เป็นสถาบันวิจัยอิสระที่มุ่งเน้นการศึกษานโยบายสาธารณะและการพัฒนาประเทศในด้านต่าง ๆ

TDRI ใช้ Typhoon ในการประมวลผลข้อมูลจากประกาศรับสมัครงานออนไลน์เพื่อวิเคราะห์แนวโน้มของตลาดแรงงานในประเทศไทย ระบบสามารถระบุข้อมูลเกี่ยวกับตำแหน่งงาน สถานที่ทำงาน และทักษะที่เป็นที่ต้องการได้อย่างแม่นยำ

การใช้ Typhoon ในการทำงานให้ผลลัพธ์ที่รวดเร็วและแม่นยำกว่าระบบที่เคยใช้ และช่วยให้สามารถพัฒนานโยบายได้อย่างทันท่วงทีโดยอ้างอิงจากข้อมูลเชิงลึกของตลาด ผลที่ได้นี้ช่วยให้ผู้กำหนดนโยบายเข้าใจแนวโน้มของตลาดแรงงานและทักษะที่จำเป็นในอนาคตในเชิงลึกได้ดียิ่งขึ้น

เราขอเชิญทุกท่านที่สนใจมุ่งมั่นผลักดันการเติบโตของนวัตกรรม AI มาร่วมเป็นพันธมิตรกับเรา:

ทีม Typhoon โดย SCB 10X
contact@opentyphoon.ai
ฝ่ายวิจัยและพัฒนา SCBX
rd@scbx.com

อ้างอิง :

ACT I: Two Philosophies, One Future.
The Battle Between Open-Source and Closed-Source AI Intensifies

[1] ข้อมูลเปรียบเทียบข้อดี-ข้อเสียของ AI แบบเปิดและปิดในบริบทธุรกิจ : https://www.multimodal.dev/post/open-source-ai-vs-closed-source-ai

[2] Closed Models Outperform Open Models, at Staggering Cost : https://virtualizationreview.com/Articles/2024/04/16/ai-index.aspx

[3] บทความ Why Open-Source LLMs Are Reshaping the Economics of AI : https://aithority.com/machine-learning/why-open-source-llms-are-reshaping-the-economics-of-ai/

[4] อัปเดตอนาคตของ "Open Source AI" จาก SCB 10X : https://www.scb10x.com/blog/open-source-ai-trends-opportunities

[5] 'Zhang, D., Sahai, T., Jarrett, D., Zhao, S., Tamkin, A., Bai, Y., ... & Etchemendy, J. (2024). The AI Index 2024 Annual Report. Stanford Institute for Human-Centered Artificial Intelligence. https://aiindex.stanford.edu/report/

[6] กรณีศึกษาจาก World Economic Forum เกี่ยวกับโมเดล DeepSeek : https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek

[7] Epoch. (2024, November 28). The state of open foundation models. Epoch. From https://epoch.ai/blog/open-models-report

[8] "Should Enterprises Adopt Closed-Source or Open-Source AI Models?" CB Insights, Mar. 2025, www.cbinsights.com/research/enterprise-adoption-closed-source-open-source-ai-models

[9] 'Artificial Analysis. (2025, February). Artificial Analysis Intelligence Benchmarking Methodology. From https://artificialanalysis.ai/methodology/intelligence-benchmarking

[10] Typhoon. (2025, February 7). Introducing Typhoon2-R1–70B: Enhanced Reasoning with DeepSeek R1. Typhoon Blog. From https://opentyphoon.ai/blog/en/introducing-typhoon2-r1-70b-enhanced-reasoning-with-deepseek-r1-abcedaa9b7ad

[11] SCB 10X. "SCB 10X Hugging Face Profile." Hugging Face, https://huggingface.co/scb10x

ACT II: Tiny Titans - Small, but Mighty.
More Versatile, Smaller, and Smarter: 3 Trends of the Next AI Evolution

[1] Google Reveals Gemini 2, AI Agents, and a Prototype Personal Assistant https://www.wired.com/story/google-gemini-2-ai-assistant-release/

[2] '"Introducing GPT-4o and More Tools to ChatGPT Free Users." OpenAI, 13 May 2024, https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/

[3] "Gemini 2.5: Our Most Intelligent AI Model." The Keyword, 25 Mar. 2025, https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/#gemini-2-5-thinking

[4] "Alibaba Cloud Releases Qwen2.5-Omni-7B: An End-to-end Multimodal AI Model." Alizila, 27 Mar. 2025, www.alizila.com/alibaba-cloud-releases-qwen2-5-omni-7b-an-end-to-end-multimodal-ai-model/

[5] Veluri, B., Peloquin, B. N., Yu, B., Gong, H., & Gollakota, S. (2024). Beyond Turn-Based Interfaces: Synchronous LLMs as Full-Duplex Dialogue Agents. arXiv preprint arXiv:2409.15594

[6] Phi-4 ทำผลงานได้ดีกว่า LLM อย่าง Gemini Pro1.5: https://www.techradar.com/computing/artificial-intelligence/microsoft-announced-phi-4-a-new-ai-thats-better-at-math-and-language-processing

[7] Abdin, M., et al. (2024). Phi-4 Technical Report. Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2024/12/P4TechReport.pdf

[8] "How IBM CEO Arvind Krishna Is Thinking About AI and Quantum Computing." TIME, 11 Mar. 2025, https://time.com/7263217/ibm-ceo-arvind-krishna-interview/

[9] "How Scaling Laws Drive Smarter, More Powerful AI." NVIDIA Blog, 12 Feb. 2025, https://blogs.nvidia.com/blog/ai-scaling-laws/

[10] อธิบายแนวคิดการให้โมเดลลองสร้างและประเมินหลายคำตอบระหว่างการใช้งานจริง เพื่อเลือกคำตอบที่ดีที่สุดในท้ายที่สุด และ บทสัมภาษณ์ Ilya Sutskever ที่กล่าวถึงการเริ่มเห็นขีดจำกัดของการเพิ่มขนาดโมเดล และความจำเป็นต้องหาแนวทางใหม่ ๆ ในการพัฒนา AI : https://www.reuters.com/technology/artificial-intelligence/openai-rivals-seek-new-path-smarter-ai-current-methods-hit-limitations-2024-11-11

[11] Mixture of Experts แทนที่จะเป็นแค่ LLM แค่หนึ่ง : https://www.datacamp.com/blog/mixture-of-experts-moe

ACT III: AI at Your Fingertips.
Agentic AI: Rise of the Agents

[1] Stanford Institute for Human-Centered Artificial Intelligence. (2024). HAI at Davos: Key Insights on AI From the World Economic Forum. Stanford HAI. https://hai.stanford.edu/news/hai-at-davos-key-insights-on-ai-from-the-world-economic-forum

[2] Davenport, T. H., & Bean, R. (2025). Five trends in AI and data science for 2025. MIT Sloan Management Review. https://sloanreview.mit.edu/article/five-trends-in-ai-and-data-science-for-2025/

ACT IV: Not Quite Human, But Almost There.
Artificial General Intelligence (AGI) and the Unresolved Path to Human-Level AI

[1] Health, A. (2024). Sam Altman lowers the bar for AGI. The Verge. https://www.theverge.com/2024/12/4/24313130/sam-altman-openai-agi-lower-the-bar

[2] Davis, W. (2024). Microsoft's AI boss and Sam Altman disagree on what it takes to get to AGI. The Verge. https://www.theverge.com/2024/12/9/24316969/mustafa-suleyman-sam-altman-microsoft-openai-agi

[3] Taylor, J. (2024). AI has the potential to 'cure all diseases', says DeepMind chief. The Times. https://www.thetimes.com/business-money/technology/article/ai-has-the-potential-to-cure-all-diseases-says-deepmind-chief-9hbdp5kpm

[4] Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., & You, J. (2024). How Far Are We From AGI: Are LLMs All We Need? [Preprint]. arXiv. https://arxiv.org/abs/2403.10932

[5] DSAI by Dr. Osbert Tay. (2025). The Shape of AI to Come! Yann LeCun at AI Action Summit 2025 [Video]. YouTube. https://www.youtube.com/watch?v=xnFmnU0Pp-8

About SCBX

SCBX เป็นยานแม่ของกลุ่มธุรกิจเทคโนโลยีทางการเงิน ครอบคลุม 3 กลุ่มธุรกิจสำคัญ ได้แก่ ธุรกิจธนาคาร (Banking Business) ธุรกิจบริการทางการเงินดิจิทัลและสินเชื่อเพื่อรายย่อย (Consumer and Digital Finance Business) ธุรกิจแพลตฟอร์มและเทคโนโลยี (Platform and Technology Business) รวมถึง Climate Technology

มุ่งขับเคลื่อนยุทธศาสตร์สู่การเป็นกลุ่มเทคโนโลยีทางการเงินระดับภูมิภาคที่น่าชื่นชม ดำเนินธุรกิจด้วยความคล่องตัว รอบคอบในการกำกับดูแลและบริหารความเสี่ยง ตลอดจนมีขีดความสามารถและศักยภาพในการแข่งขันกับคู่แข่งระดับโลกได้อย่างทัดเทียม

Authors

Tutanon Sinthuprasith, Ph.D.
R&D and Innovation Lab LEAD, SCBX

Kaweewut Temphuwapat
Chief Innovation Officer, SCBX

รายงานนี้เป็นความร่วมมือกันระหว่าง

SCBX SCB 10X dataX

SCBX

www.scbx.com