**Sustainable AI’ พัฒนา AI อย่างไรให้เป็นมิตรกับสิ่งแวดล้อม**

**การใช้พลังงานและการปล่อยคาร์บอน**

การพัฒนาเทคโนโลยี AI ที่สร้างสรรค์เนื้อหาใหม่ๆ จากข้อมูลที่มีอยู่เดิม (Generative AI หรือ GenAI) ด้วยโมเดลภาษาขนาดใหญ่ (Large Language Model: LLM) จำเป็นต้องใช้พลังงานอย่างเข้มข้น จึงส่งผลให้มีคาร์บอนฟุตพริ้นท์สูงด้วย ตัวอย่างเช่น **การฝึกฝนโมเดล Generative Pre-trained Transformer 3 (GPT-3) ซึ่งเป็นโมเดลเบื้องหลัง ChatGPT ด้วยฐานข้อมูลขนาดใหญ่จำนวน 5 แสนล้านคำ ต้องใช้พลังงานไฟฟ้าราว 1,300 เมกะวัตต์ชั่วโมง ซึ่งเทียบเท่ากับการใช้ไฟฟ้าตลอดทั้งปีของบ้านเรือนกว่า 120 หลังในสหรัฐฯ** อีกทั้งการฝึกฝนโมเดลดังกล่าวยังปล่อยคาร์บอนกว่า 500 ตัน ซึ่งเทียบเท่ากับการที่ผู้โดยสาร 1 คนเดินทางด้วยเครื่องบินจากนครนิวยอร์กไปยังซานฟรานซิสโกจำนวนกว่า 500 เที่ยว หรือพอๆ กับการใช้รถยนต์ 8 คัน (รวมน้ำมัน) ตลอดอายุการใช้งาน (ภาพที่ 1) อย่างไรก็ดี การฝึกฝนโมเดล GPT-4 ที่ซับซ้อนและมีพารามิเตอร์2/ มากขึ้น อาจใช้พลังงานไฟฟ้ามากกว่า GPT-3 ถึง 50 เท่า3/

ยิ่งไปกว่านั้น งานวิจัยหลายชิ้นสรุปว่าการใช้พลังงานส่วนใหญ่ของเทคโนโลยี AI เกิดจาก**การใช้งานหรือที่เรียกว่า 'การอนุมาน (inferrence) ’4/** เนื่องจาก AI ต้องประมวลผลโมเดลที่ซับซ้อนและข้อมูลจำนวนมากแบบเรียลไทม์ **โดยการถาม-ตอบกับ ChatGPT 1 ครั้ง ใช้พลังงานเทียบเท่ากับการเปิดหลอดไฟ LED ขนาด 5 วัตต์เป็นเวลา 1 ชั่วโมง 20 นาที5/** อย่างไรก็ดี ความเข้มข้นของพลังงานในขั้นตอนการใช้ AI ขึ้นอยู่กับ 1) ประเภทของงาน โดยการใช้ AI สร้างภาพ (Image Generation) สิ้นเปลืองพลังงานมากกว่าการใช้ AI เพื่อจำแนกประเภทข้อความ (Text Classification)กว่า 1,000 เท่า (ภาพที่ 2) และ 2) ประเภทของโมเดล โดยโมเดล AI ที่มีพารามิเตอร์มากกว่าจะใช้พลังงานในการประมวลผลมากกว่า6/

**คาร์บอนฟุตพริ้นท์ของ AI ส่วนใหญ่เกิดใน Data Center** ซึ่งเป็นศูนย์กลางการประมวลผลและจัดเก็บข้อมูลขนาดใหญ่สำหรับการฝึกฝนและใช้งานโมเดล AI **ทั้งนี้ Data Center เป็นโครงสร้างพื้นฐานที่ใช้พลังงานเข้มข้นกว่าอาคารสำนักงานทั่วไปถึง 10-50 เท่า7/**โดยใช้ไฟฟ้าประมาณ 40%-50% เพื่อจ่ายพลังงานให้อุปกรณ์คอมพิวเตอร์ และใช้อีก 30%-40% เพื่อระบายความร้อนให้กับอุปกรณ์ดังกล่าว8/

ท่ามกลางกระแสการเติบโตของ AI ทำให้องค์การพลังงานระหว่างประเทศ (International Energy Agency: IEA)9/ คาดการณ์ไว้ว่า**ภายในปี 2569 ปริมาณการใช้ไฟฟ้าของ Data Center ทั่วโลกอาจอยู่ที่ระดับ 1,000 เทราวัตต์ชั่วโมง (TWh) ซึ่งเทียบเท่ากับการใช้ไฟฟ้าของประเทศญี่ปุ่นทั้งประเทศ หรือเพิ่มขึ้นเป็นสองเท่าเมื่อเทียบกับปี 2565**ทั้งนี้ IEA คาดว่า Data Center จะเป็นภาคส่วนสำคัญที่ทำให้การใช้ไฟฟ้าในสหรัฐฯ เพิ่มสูงขึ้น โดยคิดเป็น 1 ใน 3 ของปริมาณไฟฟ้าที่ต้องการเพิ่มขึ้นในช่วงปี 2568-2569 เช่นเดียวกับในสหภาพยุโรปและจีน (ภาพที่ 3)

**การใช้น้ำ**

**การดำเนินงานของ Data Center จำเป็นต้องอาศัยน้ำจืดที่สะอาดเพื่อใช้ในระบบทำความเย็นในปริมาณมาก**เพื่อป้องกันไม่ให้อุปกรณ์อิเล็กทรอนิกส์เผชิญความร้อนสูงเกินไป โดยการฝึกฝนโมเดล GPT-3 ใน Data Center ในสหรัฐฯ ใช้น้ำไปประมาณ 5.4 ล้านลิตร **ในขณะที่การถาม-ตอบกับ ChatGPT (GPT-3) จำนวน 10-50 คำถาม ทำให้สูญเสียน้ำในปริมาณเทียบเท่ากับการดื่มน้ำครึ่งลิตร**และคาดว่า GPT-4 ซึ่งเป็นโมเดลที่ซับซ้อนกว่า จะใช้น้ำในปริมาณมากยิ่งขึ้นอีก10/

นอกจากนี้ การพัฒนา AI ต้องอาศัยฮาร์ดแวร์สมรรถนะสูง เช่น เซิร์ฟเวอร์ และอุปกรณ์จัดเก็บข้อมูล ซึ่งล้วนมีชิปเป็นส่วนประกอบสำคัญ ทั้งนี้ น้ำถือเป็นปัจจัยที่ขาดไม่ได้ในกระบวนการผลิตชิป เนื่องจากโรงงานผลิตชิปต้องใช้น้ำเพื่อรักษาความเย็นให้กับเครื่องจักร และทำความสะอาดแผ่นเวเฟอร์ **โดยในปี 2567 S&P Global ประมาณการว่าผู้ผลิตชิปทั่วโลกใช้น้ำรวมกันพอๆ กับปริมาณการใช้น้ำทั้งหมดในฮ่องกง และโดยเฉลี่ยแล้วโรงงานผลิตชิป 1 แห่ง ใช้น้ำบริสุทธิ์ถึง 10 ล้านแกลลอนต่อวัน ซึ่งเทียบเท่ากับการใช้น้ำในแต่ละวันของครัวเรือนในสหรัฐฯ 33,000 ครัวเรือน11/** ยิ่งไปกว่านั้น การผลิตชิปต้องใช้โลหะ เช่น ทองแดง อะลูมิเนียม และลิเธียม ซึ่งกระบวนการทำเหมืองเพื่อขุดและแปรรูปแร่เหล่านี้ อาจสร้างมลพิษต่อแหล่งน้ำและสิ่งแวดล้อมอื่นๆอีกด้วย

**ขยะอิเล็กทรอนิกส์**

พัฒนาการของ AI ก้าวหน้าไปอย่างรวดเร็ว ส่งผลให้ต้องมีการปรับปรุงอุปกรณ์อิเล็กทรอนิกส์สมรรถนะสูงให้ทันสมัยอยู่เสมอ ฮาร์ดแวร์เหล่านี้จึงมักมีอายุการใช้งานสั้นเพียง 2-5 ปีก่อนจะกลายเป็นขยะอิเล็กทรอนิกส์ ซึ่งมีสารเคมีอันตรายอย่างตะกั่ว ปรอท และแคดเมียม ที่สามารถปนเปื้อนดินและแหล่งน้ำหากไม่ได้รับการจัดการอย่างเหมาะสม จึงเป็นภัยต่อสุขภาพมนุษย์และสิ่งแวดล้อมได้ งานวิจัยที่เผยแพร่ในวารสาร Nature Computational Science เมื่อปลายเดือนตุลาคม 2567 คาดการณ์ไว้ว่า **GenAI จะก่อให้เกิดขยะอิเล็กทรอนิกส์เพิ่มขึ้นถึง 1.2 – 5 ล้านตันในช่วงปี 2566-2573** ซึ่งแม้จะมีสัดส่วนไม่มากเมื่อเทียบกับปริมาณขยะอิเล็กทรอนิกส์ทั่วโลกในปี 2565 จำนวน 62 ล้านตัน แต่**ขยะอิเล็กทรอนิกส์จาก GenAI มีแนวโน้มเพิ่มขึ้นอย่างรวดเร็ว โดยปริมาณขยะที่เกิดขึ้นใหม่ในปี 2573 อาจสูงกว่าในปี 2566 เกือบ 1,000 เท่า หรือใกล้เคียงกับปริมาณขยะจาก iPhone 15 Pro กว่าหมื่นล้านเครื่อง12/** **จะเห็นว่าความก้าวหน้าอย่างรวดเร็วของเทคโนโลยี AI ไม่ได้นำมาซึ่งผลเชิงบวกต่อโลกเท่านั้น แต่ยังมาพร้อมกับผลกระทบด้านพลังงาน ทรัพยากร และสิ่งแวดล้อม** ดังนั้น โลกจึงพยายามมุ่งไปสู่ AI ที่ยั่งยืน ซึ่งให้ความสำคัญกับการพัฒนาและใช้เทคโนโลยี AI ให้เกิดประโยชน์สูงสุด ในขณะเดียวกันก็ไม่ส่งผลกระทบต่อสิ่งแวดล้อมและสังคมทั่วทุกมุมโลก

**AI ที่ยั่งยืนคืออะไร และทำไมต้องให้ความสำคัญ**

**AI ที่ยั่งยืนเป็นแนวคิดการพัฒนาเทคโนโลยี AI โดยคำนึงถึงผลกระทบด้านความยั่งยืนตลอดวงจรชีวิตของ AI ตั้งแต่การจัดหาวัสดุอุปกรณ์ทางเทคโนโลยี การพัฒนาโครงสร้างพื้นฐาน (Data Center) การจัดเก็บข้อมูล การฝึกฝนและทดสอบโมเดล การใช้งาน ไปจนถึงการกำจัดขยะอิเล็กทรอนิกส์ที่เกิดจาก AI**(ภาพที่ 4) ทั้งนี้ ปัจจุบัน Sustainable AI เน้นลดผลกระทบด้านสิ่งแวดล้อมจากการใช้พลังงานและทรัพยากรมหาศาล ซึ่งสอดรับกับกระแสการเปลี่ยนผ่านไปสู่การปล่อยก๊าซเรือนกระจกสุทธิเป็นศูนย์ (Net-Zero Emission) ที่ทุกอุตสาหกรรมทั่วโลกให้ความสำคัญ ดังนั้น เราจึงอาจพบเจอคำว่า **AI ที่เป็นมิตรกับสิ่งแวดล้อม หรือ AI สีเขียว** ซึ่งล้วนแล้วแต่มุ่งเน้นไปในทิศทางเดียวกัน

การพัฒนาเทคโนโลยี AI อย่างยั่งยืนสามารถดำเนินการผ่านแนวทาง 3 ด้าน ได้แก่ 1) การใช้อุปกรณ์อิเล็กทรอนิกส์ที่เป็นมิตรกับสิ่งแวดล้อมมากขึ้น (Hardware Optimization) 2) การพัฒนาและฝึกฝนโมเดล AI ที่ใช้พลังงานลดลง และ 3) การเพิ่มประสิทธิภาพพลังงานและสัดส่วนพลังงานหมุนเวียนใน Data Center ซึ่งมีรายละเอียดดังนี้

**Hardware Optimization: ใช้ฮาร์ดแวร์อย่างมีประสิทธิภาพและเป็นมิตรกับสิ่งแวดล้อม**

วงจรชีวิตของการพัฒนา AI เกิดขึ้นตั้งแต่ต้นน้ำของห่วงโซ่อุปทาน (Supply Chain) กล่าวคือ การทำเหมืองแร่เพื่อให้ได้วัตถุดิบในการผลิตชิ้นส่วนและอุปกรณ์อิเล็กทรอกนิกส์ที่เป็นฮาร์ดแวร์สำคัญของ AI เช่น หน่วยประมวลผลกลาง (Central Processing Unit: CPU) 13/ และหน่วยประมวลผลกราฟิก (Graphics Processing Unit: GPU) 14/ ซึ่งส่งผลกระทบต่อสิ่งแวดล้อมตั้งแต่ต้นน้ำ สะท้อนจากคาร์บอนฟุตพริ้นท์ของกระบวนการผลิต GPU ที่คิดเป็น 22% ของการปล่อยคาร์บอนทั้งหมดจากการฝึกฝนโมเดล BLOOM ซึ่งเป็นโมเดลภาษาขนาดใหญ่ (Large Language Model: LLM)ที่พัฒนาโดย BigScience15/ **ดังนั้น แนวคิด AI ที่ยั่งยืนจึงสนับสนุนการใช้วัสดุที่เป็นมิตรกับสิ่งแวดล้อมในการผลิตฮาร์ดแวร์**เช่น เหล็กและอะลูมิเนียมรีไซเคิล พลาสติกชีวภาพ วัสดุทดแทนธาตุหายาก (Rare Earth Alternatives) **อีกทั้งยังสนับสนุนการยืดอายุและปรับปรุงการใช้ฮาร์ดแวร์ที่มีอยู่เดิมให้มีประสิทธิภาพสูงสุด** เพื่อลดความต้องการการผลิตฮาร์ดแวร์ใหม่ๆ รวมถึงลดขยะอิเล็กทรอนิกส์จาก AI

นอกจากนี้ ปัจจุบันมีการพัฒนา**ฮาร์ดแวร์ที่ออกแบบมาโดยเฉพาะเพื่อเพิ่มประสิทธิภาพการประมวลผลและลดการใช้พลังงาน เช่น หน่วยประมวลผลเทนเซอร์ (Tensor Processing Unit: TPU)** ซึ่งเป็นวงจรรวมแบบเฉพาะทาง (Application-Specific Integrated Circuit: ASIC)ที่พัฒนาโดย Google ทั้งนี้ TPU สามารถประมวลผลได้เร็วกว่าและใช้พลังงานน้อยกว่าเมื่อเทียบกับ CPU และ GPU แบบทั่วไป ในขณะเดียวกันก็ยังคงมีความแม่นยำสูง16/

**Model Optimization: พัฒนาโมเดล AI ให้ใช้พลังงานน้อยลง**

การออกแบบ ฝึกฝน และใช้งาน AI จะปล่อยคาร์บอนด้วยความเข้มข้นมากหรือน้อยนั้น ขึ้นอยู่กับขนาดและความซับซ้อนของโมเดล โดยการประมวลผลโมเดลขนาดใหญ่และใช้ข้อมูลมหาศาลต้องการฮาร์ดแวร์ที่มีประสิทธิภาพสูงและการระบายความร้อนที่ดี จึงส่งผลให้ใช้พลังงานเพิ่มขึ้นด้วย ดังนั้น แนวทางหนึ่งที่จะลดคาร์บอนฟุตพริ้นท์จาก AI ได้คือ **การปรับปรุงการออกแบบโมเดล AI เพื่อเพิ่มประสิทธิภาพและลดการใช้พลังงานในการประมวลผล (Algorithm optimization)** ซึ่งในปัจจุบันผู้พัฒนา AI มุ่งเน้นไปที่การลดขนาดหน่วยความจำ (Memory Footprint) และความซับซ้อนในการประมวลผล (Computational Complexity) ด้วยวิธีต่างๆ อาทิ 1) **การตัดส่วนที่ไม่จำเป็น (Pruning)** ในโครงข่ายประสาทเทียม (Artificial Neural Networks: ANNs) เพื่อลดขนาดโมเดลโดยไม่กระทบกับความสามารถของโมเดล ซึ่งมักทำหลังขั้นตอนการฝึกฝนโมเดล 2) **การลดระดับความแม่นยำ (Quantization)** เช่น เปลี่ยนรูปแบบการเก็บตัวเลขจากที่มีจุดทศนิยมเป็นจำนวนเต็ม เพื่อลดการใช้หน่วยความจำ แต่ยังคงความแม่นยำในระดับที่ยอมรับได้ 3) **การถ่ายทอดความรู้จากโมเดลขนาดใหญ่ไปยังโมเดลขนาดเล็กที่มีประสิทธิภาพใกล้เคียงกัน (Distillation)** และ 4) **การปรับปรุงประสิทธิภาพการคำนวณของกลไกแอตเทนชัน (Flash Attention)** ที่เป็นส่วนสำคัญของโมเดลภาษาขนาดใหญ่ (Large Language Model: LLM) เช่น GPT จึงลดการใช้หน่วยความจำและพลังงานทั้งในขั้นตอนการฝึกฝนและใช้งานโมเดล17/

นอกจากนี้ **การพัฒนาโมเดลขนาดเล็ก (Tiny Machine Learning: TinyML)**ได้รับความนิยมมากขึ้น เนื่องจากเป็นโมเดลที่สามารถประมวลผลบนอุปกรณ์ที่มีความจำไม่มาก และลดการใช้พลังงานได้ โดย TinyML มักใช้เทคนิคการลดขนาดและความซับซ้อนของโมเดล (Pruning and Quantization)18/ จึงเหมาะกับการใช้งานที่เน้นการประหยัดพลังงาน

**Data Center Optimization: เพิ่มประสิทธิภาพและสัดส่วนพลังงานหมุนเวียนใน Data Center**

Data Center ใช้พลังงานไฟฟ้าในการทำงานของฮาร์ดแวร์สมรรถนะสูง การจัดการข้อมูลขนาดใหญ่ และระบบทำความเย็น ซึ่งล้วนต้องอาศัยพลังงานไฟฟ้าทั้งในขณะที่มีการฝึกฝนและใช้งานโมเดล AI (Dynamic Energy Consumption) และในขณะที่ไม่มีการทำงานของอุปกรณ์ใดๆ (Idle Energy Consumption) **ดังนั้น การลดคาร์บอนฟุตพริ้นท์จาก Data Center สามารถทำได้โดยการปรับปรุงประสิทธิภาพการใช้พลังงาน**ไม่ว่าจะเป็นการปรับปรุงระบบทำความเย็นด้วยเทคโนโลยีที่ใช้อากาศเย็นภายนอกอาคาร (Free Cooling) การใช้เทคโนโลยีเซิร์ฟเวอร์เสมือน (Server Virtualization) เพื่อลดจำนวนเครื่องเซิร์ฟเวอร์ที่ต้องเปิดใช้งาน หรือการติดตั้งระบบบริหารจัดการศูนย์ข้อมูล (Data Center Infrastructure Management: DCIM) เพื่อติดตามการใช้พลังงาน โดยมีเป้าหมายหลักคือการทำให้ค่าประสิทธิภาพการใช้พลังงาน (Power Usage Effectiveness: PUE) 19/ เข้าใกล้ 1 มากที่สุด