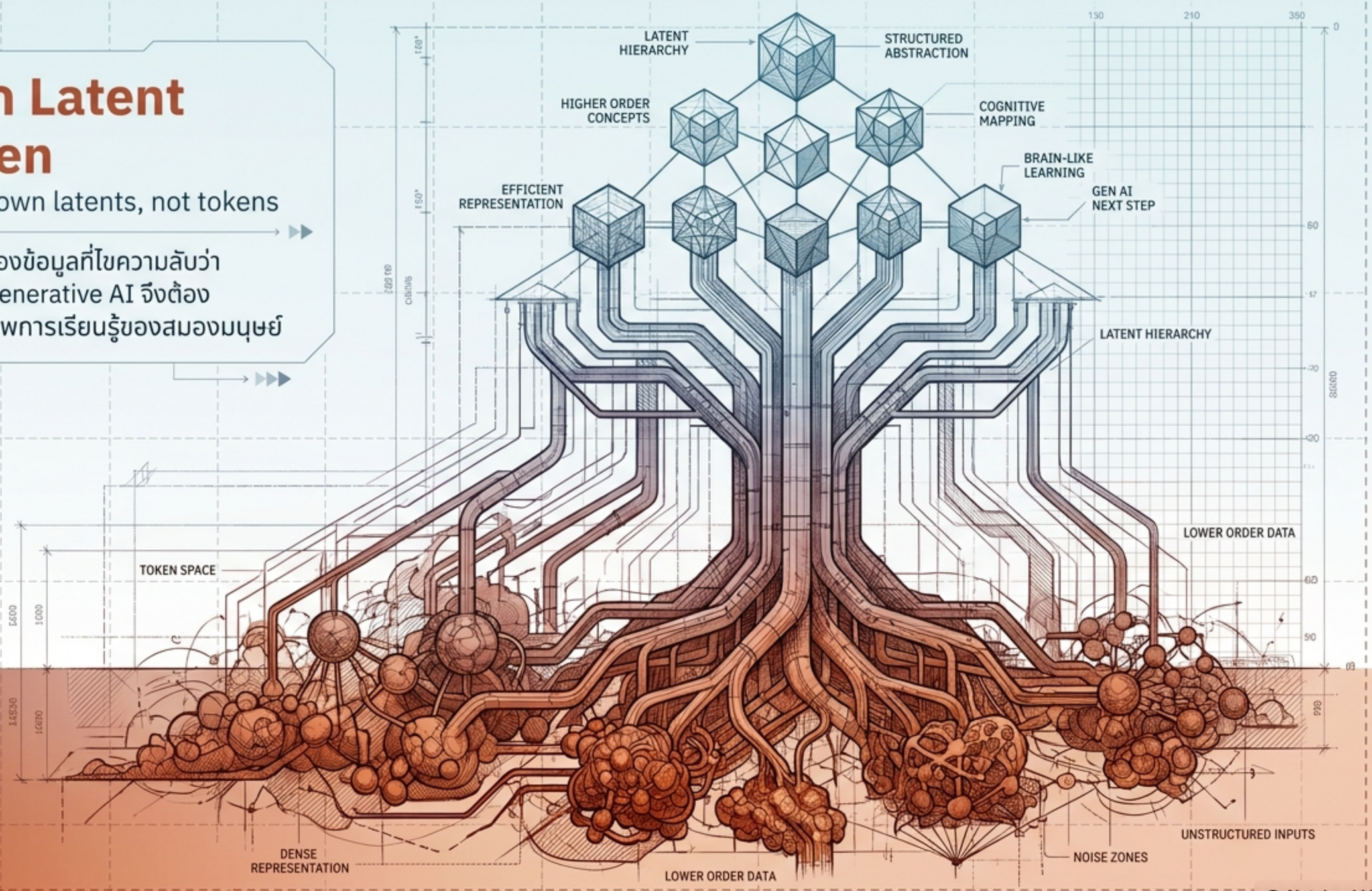
เรียนรู้จาก Latent ไม่ใช่ Token

Learn from your own latents, not tokens

ทฤษฎีความซับซ้อนของข้อมูลที่ไขความลับว่า
ทำไมก้าวต่อไปของ Generative AI จึงต้อง
เลียนแบบประสิทธิภาพการเรียนรู้ของสมองมนุษย์



NotebookLM

ช่องว่างระดับ 100,000 เท่า ระหว่าง AI และมนุษย์

AI สมัยใหม่ประสบความสำเร็จอย่างสูง แต่ต้องแลกมาด้วยความหิวโหยข้อมูลที่ผิดธรรมชาติ
ในขณะที่เด็กมนุษย์สร้างความเข้าใจโลกผ่านข้อมูลจำนวนน้อยนิด
โมเดลภาษาชั้นนำกลับต้องการข้อมูลมากกว่าถึง 5 ลำดับขั้น

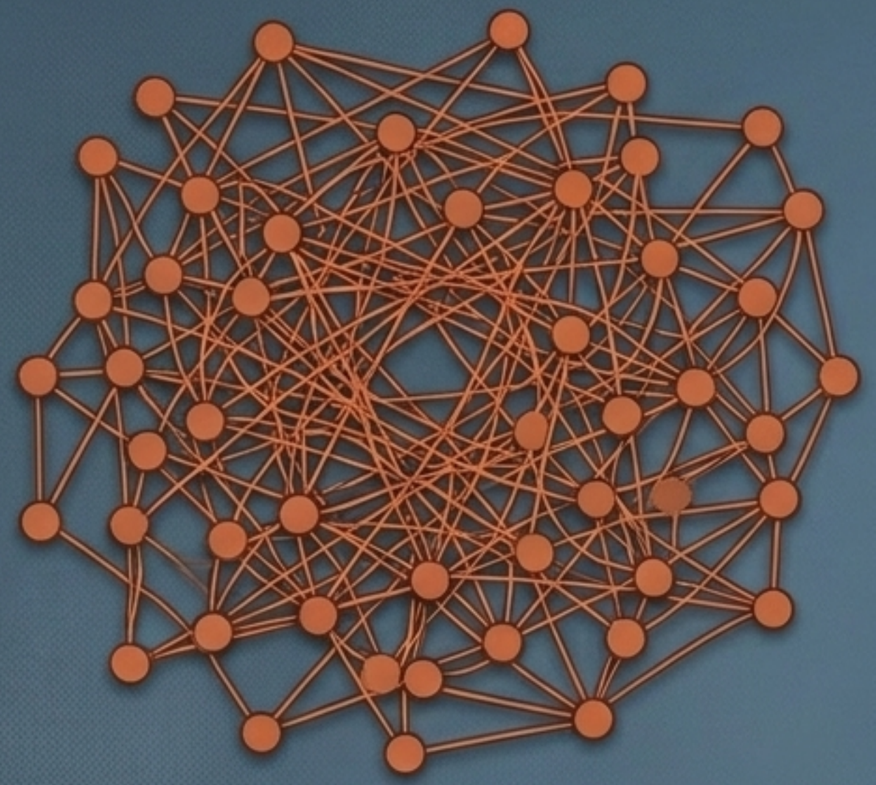
NotebookLM

การเปลี่ยนผ่านสู่ยุคแห่งการคาดเดาความหมาย (The Paradigm Shift)

แนวทางใหม่ของ AI ละทิ้งการพยายามสร้าง Input เดิมคืน แต่หันมาฝึกให้โมเดลคาดเดา Latent Representation ของตัวมันเองในระดับที่ลึกขึ้น

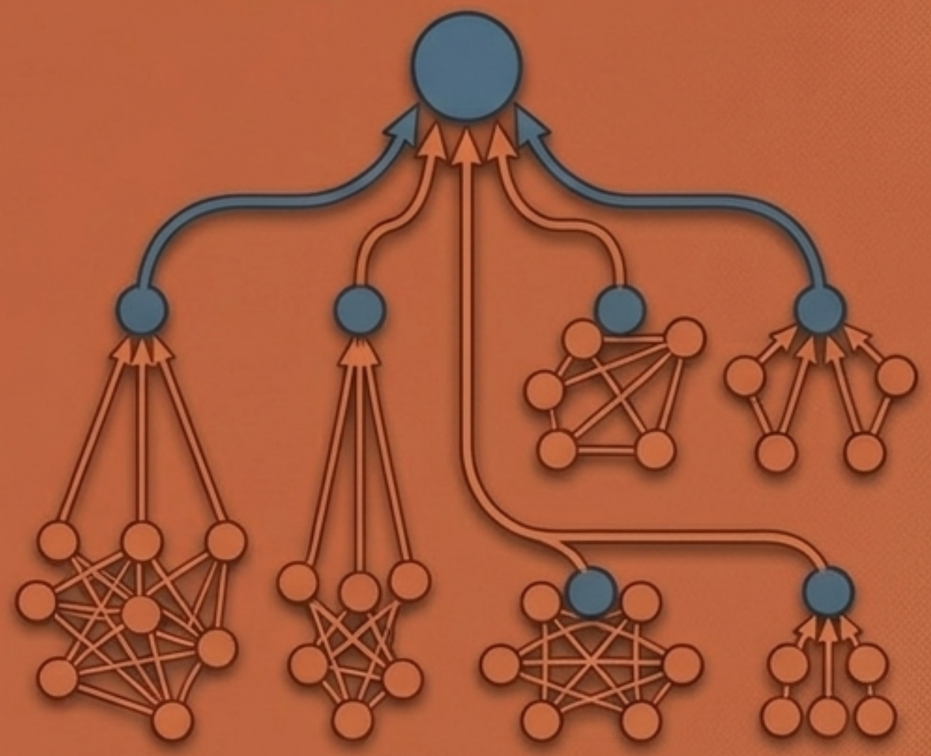
อดีต: คาดเดาคำถัดไป (Next-Token)

สัญญาณรบกวนสูง มุ่งเน้นเปลือกนอก



อนาคต: คาดเดาความหมายแฝง (Latent Prediction)

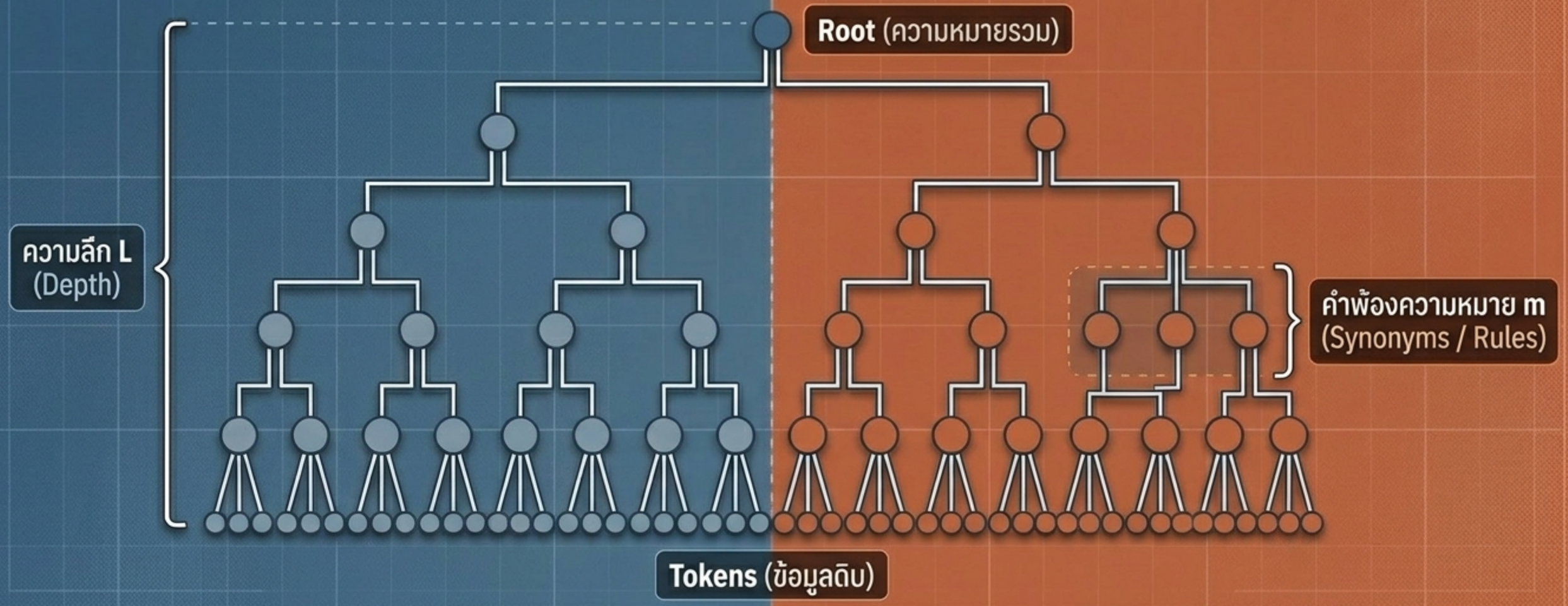
สกัดเฉพาะโครงสร้างที่เป็นระเบียบ



NotebookLM

ทำความรู้จัก Random Hierarchy Model (RHM)

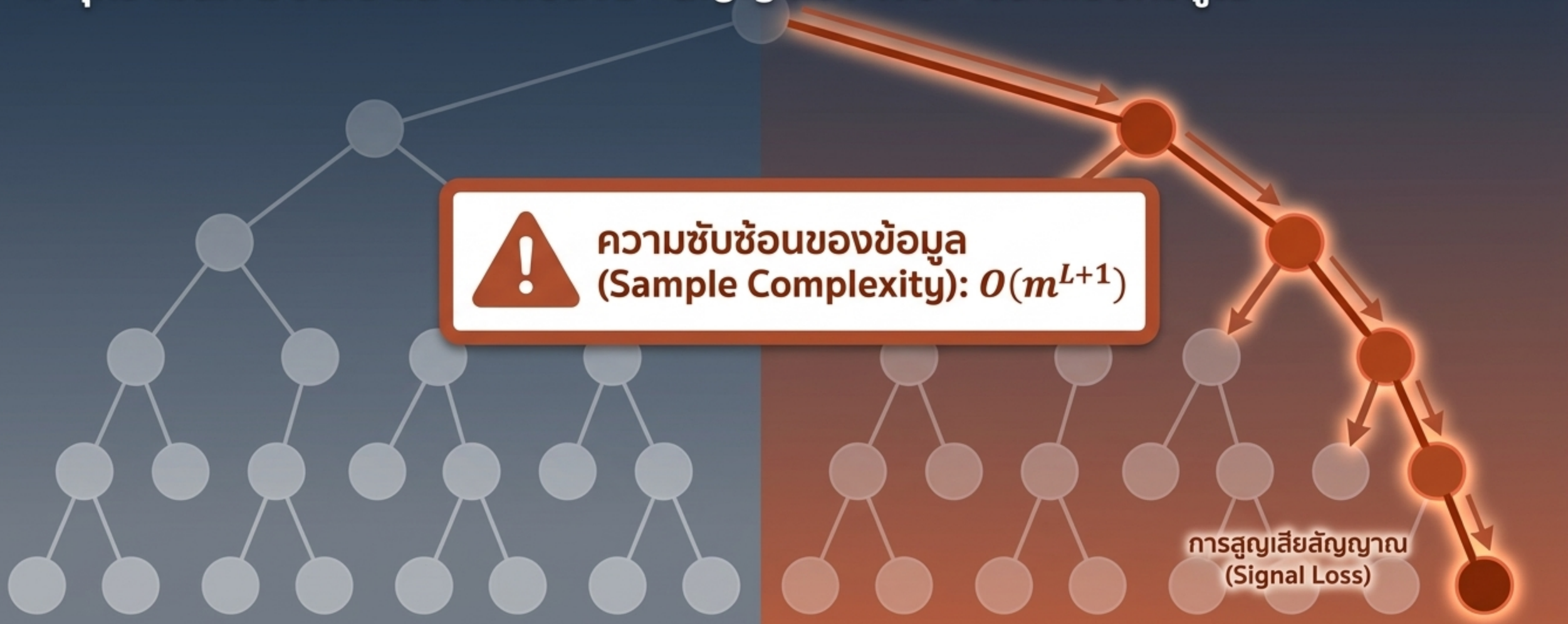
เพื่อหาข้อพิสูจน์ทางคณิตศาสตร์ เราจำลองโครงสร้างภาษาและภาพด้วย ไวยากรณ์แบบต้นไม้
ข้อมูลดิบคือใบไม้ที่ฐาน ส่วนความหมายคือกิ่งก้านที่ซ้อนทับเป็นลำดับชั้น



NotebookLM

คอขวดของการเรียนรู้ระดับ Token (The Exponential Bottleneck)

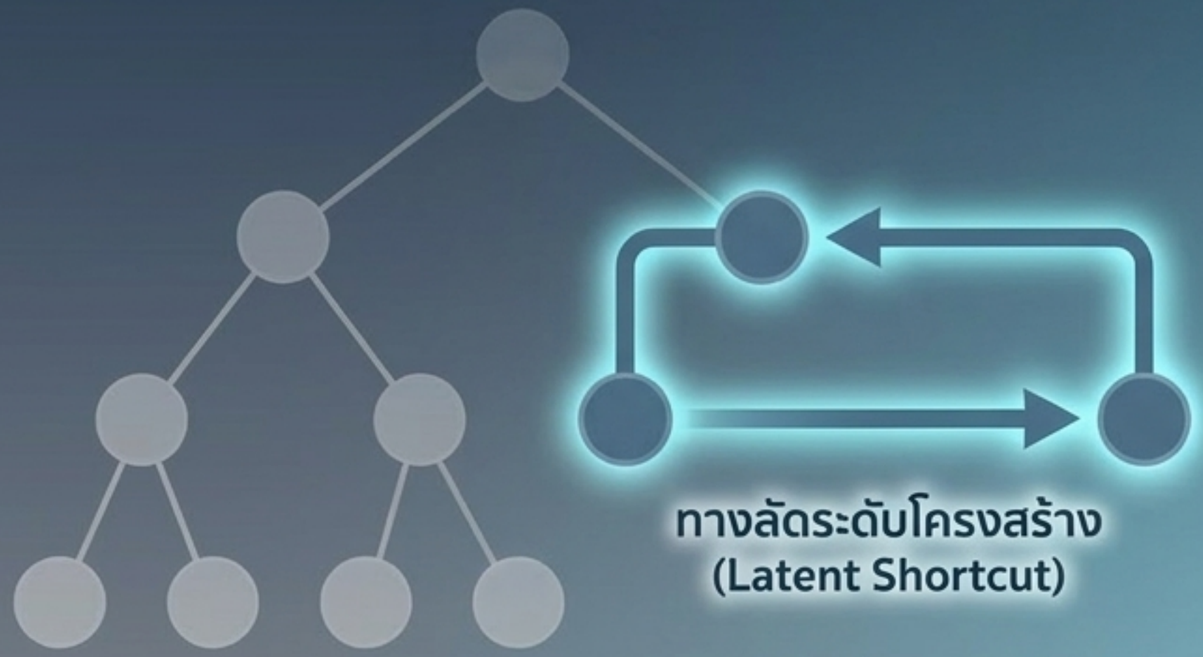
หากเป้าหมายคือการทำนาย Token ที่ถูกซ่อน โมเดลจะต้องส่งสัญญาณจากล่างสุด ทะลุความลึก L ขึ้นไป และตีกลับลงมา สัญญาณจะเจือจางลงแบบทวีคูณ



NotebookLM

ทางออกด้วยสมการคงที่: เมื่อระยะทางไม่ขึ้นกับความลึก

เมื่อโมเดลเรียนรู้โดยใช้ Latent ในระดับเดียวกันเป็นเป้าหมาย สัญญาณจะไม่ถูกเจือจางลงไปที่ฐานอีกต่อไป ความต้องการข้อมูลจึงถูกล็อคไว้ที่ระดับคงที่



ทางลัดระดับโครงสร้าง
(Latent Shortcut)

$$O(m^3)$$

คงที่ เทียบกับ $O(m^{L+1})$ ทวีคูณ

ความซับซ้อนของข้อมูล
(Sample Complexity): $O(m^3)$

NotebookLM

มิติแห่งการเรียนรู้ 3 รูปแบบ (Comparison Matrix)

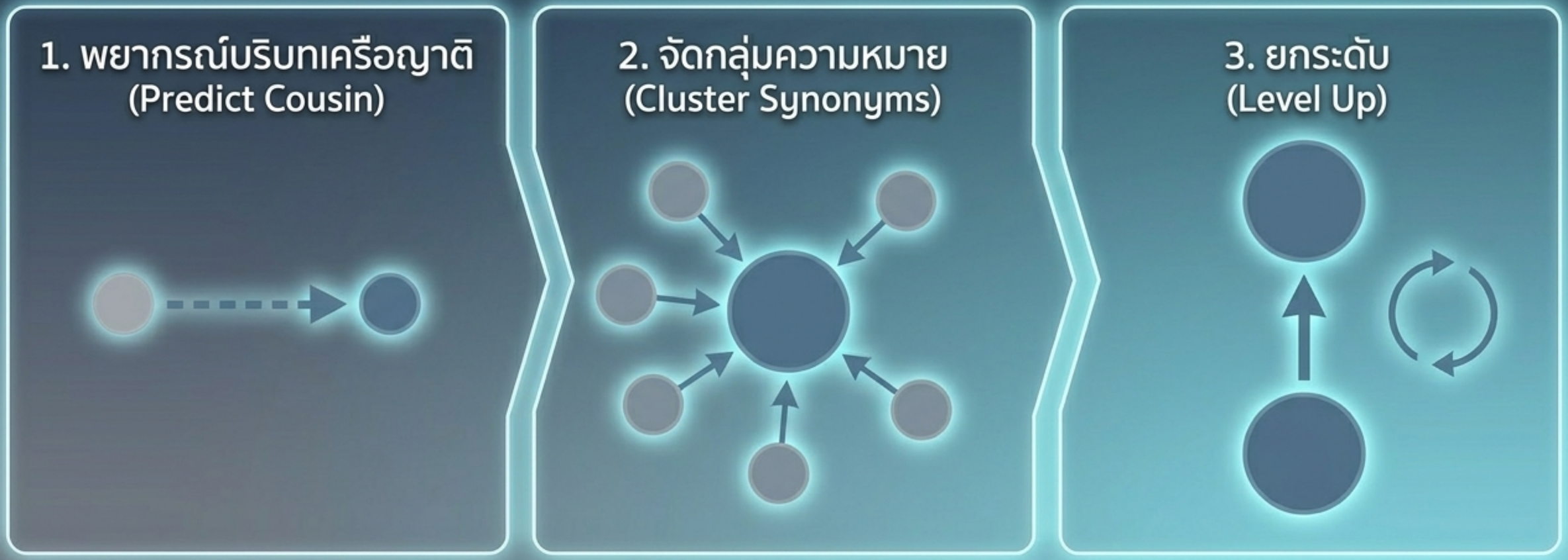
สรุปรวบยอดความแตกต่างทางคณิตศาสตร์ ซึ่งพิสูจน์ว่าจุดหมายการทำนาย (Target) คือตัวกำหนดประสิทธิภาพที่แท้จริงของโมเดล

วิธีการเรียนรู้ (Approach)	เป้าหมาย (Target)	เส้นทางสัญญาณ (Signal Path)	Sample Complexity
Supervised Classification	Root Label (บนสุด)	ล่างสุด -> บนสุด	$O(m^L)$
Token-level SSL (MLM, Diffusion)	Visible Token (ล่างสุด)	ล่างสุด -> บนสุด -> ล่างสุด	$O(m^{L+1})$
Latent-level SSL (data2vec, ILC)	**Own Latent (ระดับเดียวกัน)**	**แนวนอน (Cousin-to-Cousin)**	$\boldsymbol{O(m^3)}$

NotebookLM

บทพิสูจน์ที่ 1: คณิตศาสตร์บริสุทธิ์ (Iterative Latent Clustering)

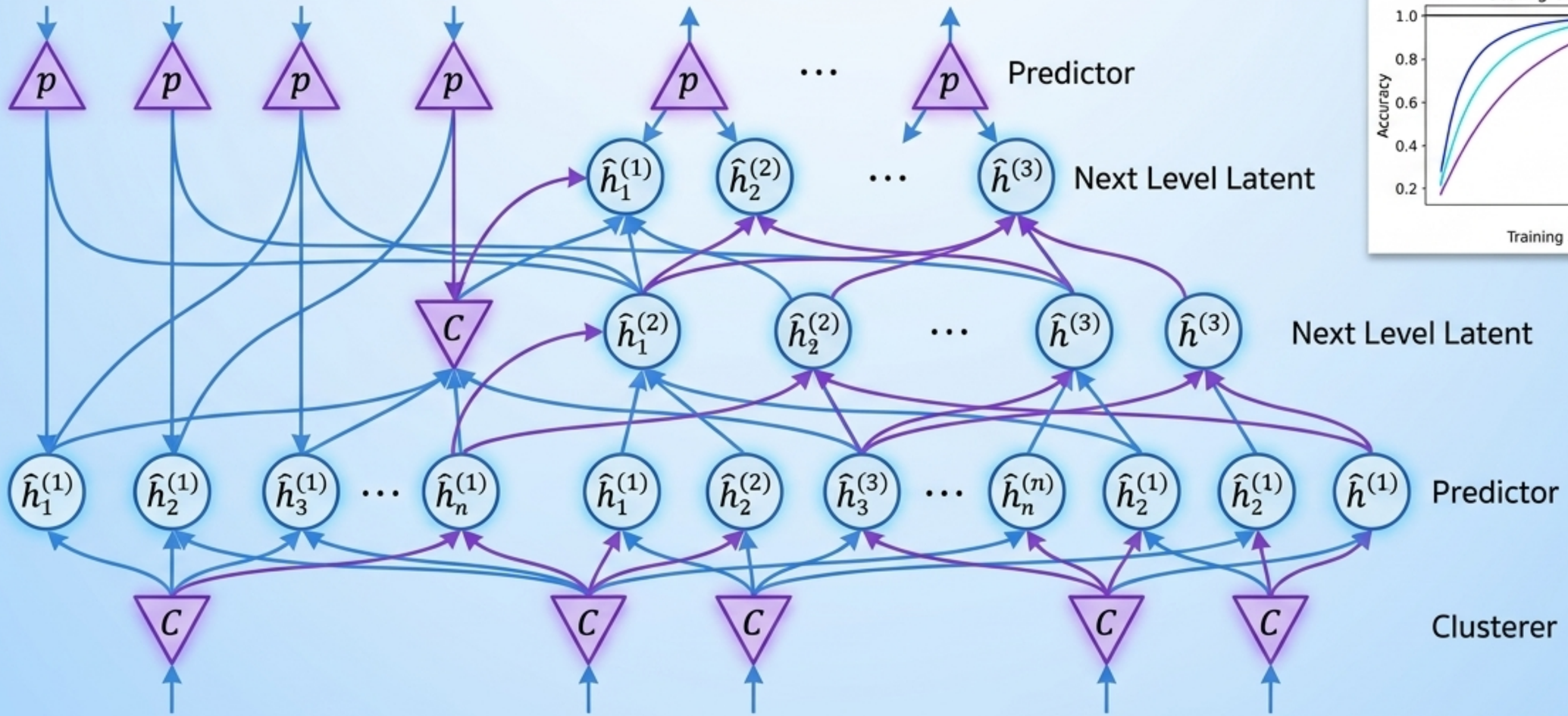
อัลกอริทึม ILC พิสูจน์ทางทฤษฎีว่า หากเราจัดกลุ่มบริบททีละชั้น โครงสร้างต้นไม้ทั้งหมดจะถูกสร้างขึ้นใหม่ ได้ด้วยปริมาณข้อมูลเพียง ~m^3 ในทุกระดับความลึก

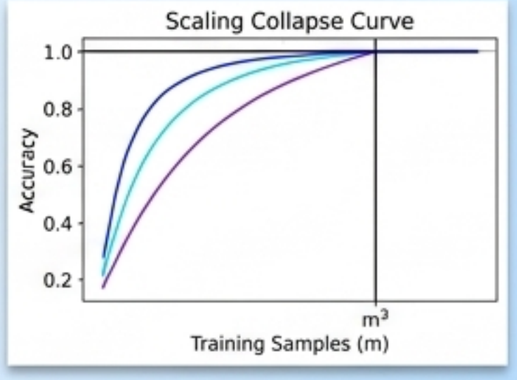


NotebookLM

บทพิสูจน์ที่ 2: การจำลองด้วย Neural Network (SLC Architecture)

สถาปัตยกรรมแบบต่อตึก (Stacked Latent-Clustering) เปลี่ยนคณิตศาสตร์ให้เป็นโครงข่ายประสาทเทียม ผลลัพธ์ยืนยันสมการ O(m^3) อย่างสมบูรณ์

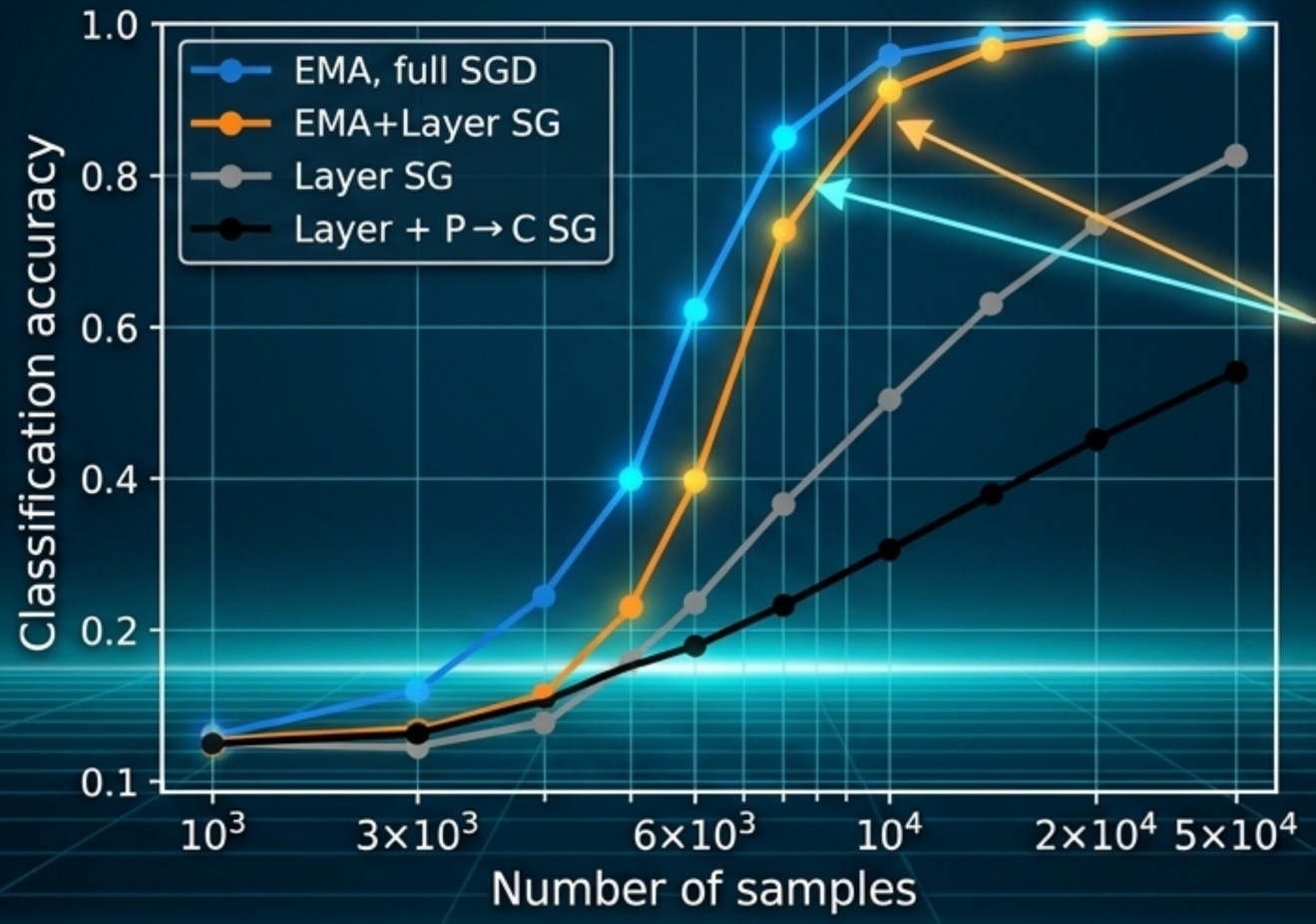




NotebookLM

บทพิสูจน์ที่ 3: กฎการเรียนรู้แบบเฉพาะที่ (Local Learning Rules)

โครงข่าย SLC สามารถเรียนรู้โครงสร้างได้แม้จะถูกตัดตอนการไหลของ Gradient (Stop-gradient) ระหว่างชั้น สอดคล้องกับกลไก Predictive Coding ของสมองมนุษย์. เรียนรู้สำเร็จโดยไม่ต้องใช้ Global Backpropagation.

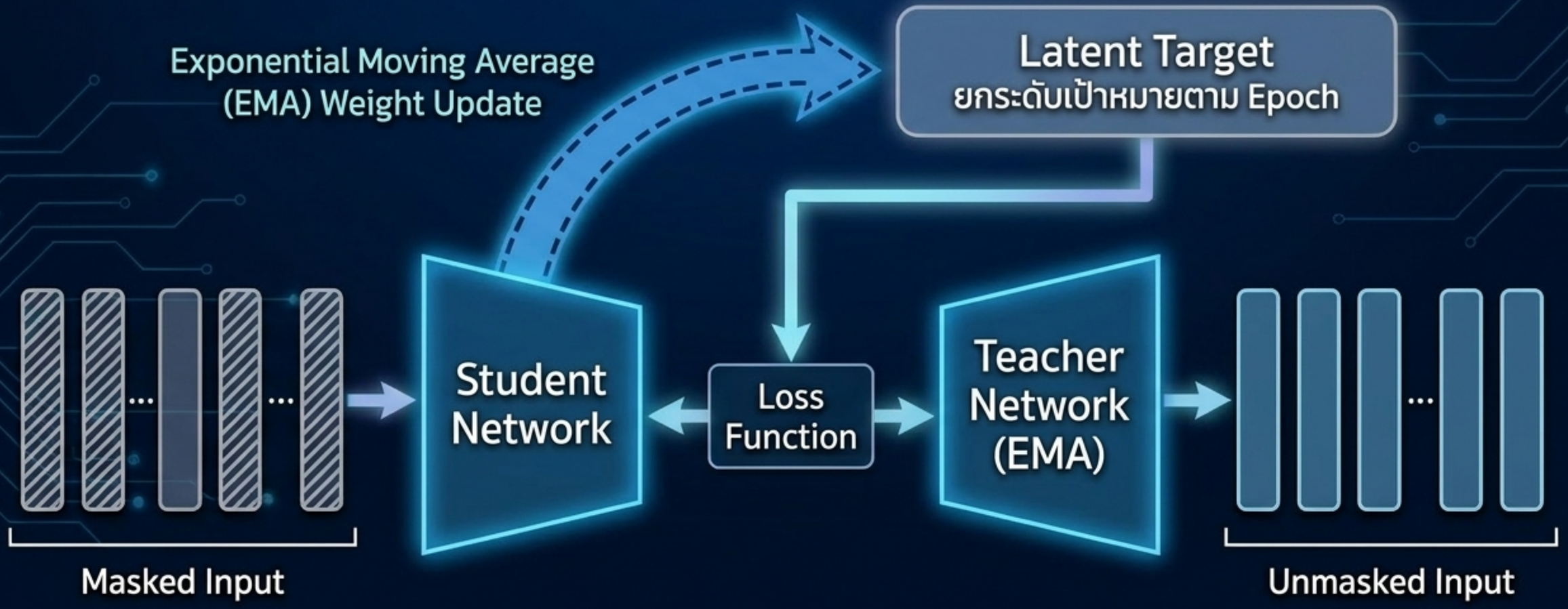


ความแม่นยำสูงใกล้เคียงกัน!
(High accuracy similarity!)

NotebookLM

บทพิสูจน์ที่ 3: ถอดรหัสกลไกของ data2vec

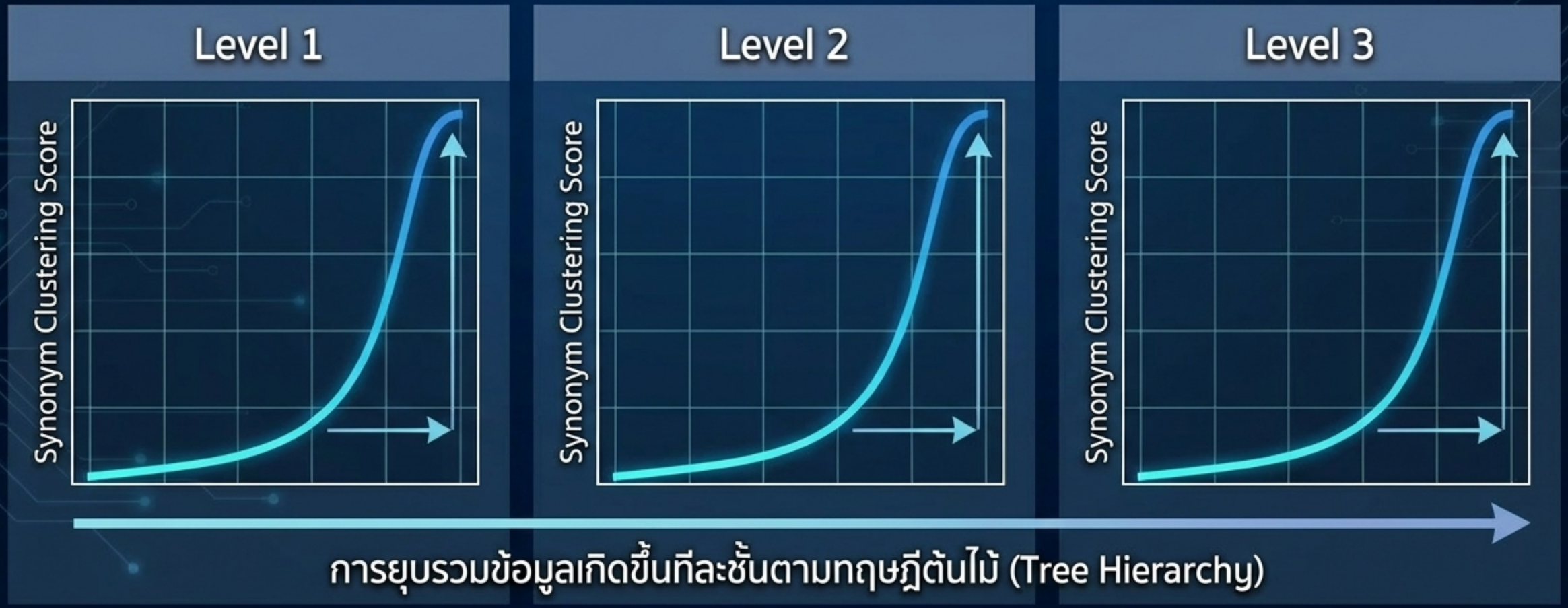
โมเดลระดับโลกอย่าง data2vec แท้จริงแล้วทำงานแบบ Hierarchical Latent Prediction อย่างไม่รู้ตัว โดยมีโครงสร้าง Teacher คอยสร้างเป้าหมายให้ Student



NotebookLM

การจัดกลุ่มคำพ้องความหมายใน data2vec

ข้อมูลจากการทดลองยืนยันว่า Encoder ของ data2vec เกิดการจัดกลุ่ม (Clustering) คำพ้องความหมายเข้าด้วยกันอย่างชัดเจนทีละชั้น



NotebookLM

จุดบรรจบของ 3 บทพิสูจน์

ไม่ว่าจะเป็นอัลกอริทึม สถาปัตยกรรมเฉพาะ หรือโมเดลระดับโลก
ทั้งหมดล้วนไขปริศนาความซับซ้อนของข้อมูลด้วยจุดตัดเดียวกัน

คณิตศาสตร์เชิงทฤษฎี
(ILC Algorithm)

การจำลองโครงข่ายเฉพาะที่
(SLC Network)

$$O(m^3)$$

การคาดเดา Latent ทำลายข้อจำกัด Exponential

โมเดล SSL ระดับโลก
(data2vec)

NotebookLM

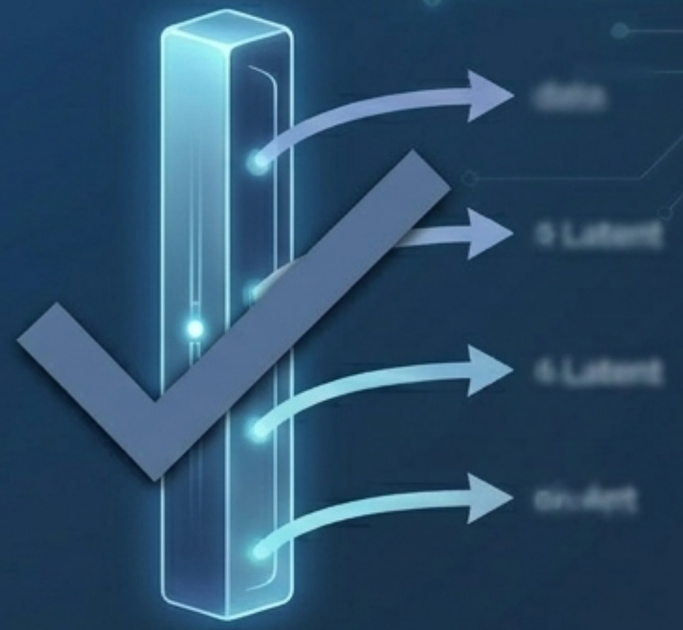
การต่อชั้นโมเดลแบบ H-JEPA ซ้ำซ้อนหรือไม่?

โมเดลเข้ารหัสเดี่ยว (Single Encoder) อย่าง data2vec สามารถทำ Multi-scale Latent Discovery ได้เอง ดังนั้นการบังคับต่อชั้นสถาปัตยกรรมซ้อนกันอาจไม่มีความจำเป็น

Explicit Stacking (H-JEPA)

ซับซ้อนและใช้ทรัพยากรสูง

Implicit Hierarchical Discovery (data2vec)



ลื่นไหลและจัดการตัวเองได้

NotebookLM

ก้าวต่อไปของ Generative AI (The Path Forward)

การพยากรณ์ Latent ของตัวเอง คือกุญแจสำคัญที่จะทลายขีดจำกัด Scaling Laws แบบเดิม สู่ยุคของ AI ที่ฉลาดขึ้นด้วยข้อมูลที่น้อยลง

1. เปลี่ยนเป้าหมาย (Shift Objectives)
จาก Token-level สู่ Own-Latent Prediction

2. ทลายคอขวด (Break the Bottleneck)
บรรลุประสิทธิภาพระดับสมการคงที่ O(m^3)

3. คิดใหม่เรื่องโครงสร้าง (Rethink Architectures)
หลีกเลี่ยงความซับซ้อนของการต่อชั้นแบบ Explicit

อนาคตของการเรียนรู้คือการสร้างความหมาย ไม่ใช่การจดจำ

NotebookLM