Kimi-VL: ประสิทธิภาพเหนือ/ระดับในโลก Multimodal AI

โมเดล **Open-Source** สถาปัตยกรรม **MoE** กับพารามิเตอร์ใช้งานจริง
เพียง **2.8 พันล้าน** (2.8B Activated Parameters)

การผสานรวมที่ลงตัวของ **‘Smart’, ‘Long’** และ **‘Clear’**

NotebookLM

ประสิทธิภาพระดับเรือธง
โดยไม่ต้องใช้ทรัพยากรระดับมหาศาล

2.8B

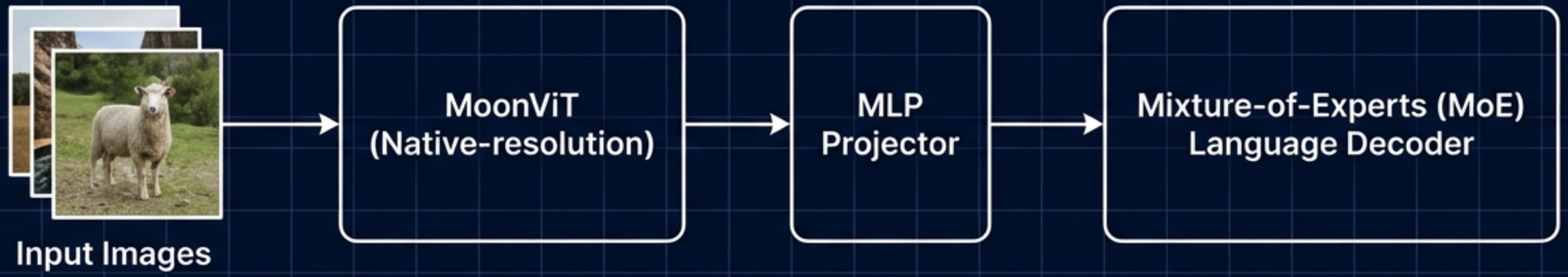
Activated Parameters (จาก 16B Total)

- **Context Window**: รองรับบริบทความยาวถึง 128k tokens
- **Visual**: รองรับ Native Resolution (MoonViT)
- **Performance**: เอาชนะ GPT-4o-mini และ Qwen2.5-VL-7B ในหลายเกณฑ์วัดผล

ประสิทธิภาพ (จากเกณฑ์วัดผล)

NotebookLM

สถาปัตยกรรม: การผสานพลังของ MoonViT และ Moonlight MoE



Vision (MoonViT):
Vision Encoder ความละเอียดดั้งเดิม
ขนาด 400M พารามิเตอร์
ไม่มีการแบ่งภาพเป็นส่วนย่อย

Bridge (MLP Projector):
ใช้ MLP 2 ชั้น เพื่อเชื่อมต่อ
Vision Features เข้ากับ LLM

Language (Moonlight MoE):
โมเดลภาษาแบบ Mixture-of-Experts
ที่มีพารามิเตอร์ใช้งานจริงเพียง 2.8B

NotebookLM

kimi-ai.

Pillar 1: ‘Clear’ – การรับรู้ภาพที่ชัดเจนด้วย Native Resolution

MoonViT ประมวลผลภาพที่ความละเอียดและอัตราส่วนภาพ (Aspect Ratio) ที่แตกต่างกันโดยธรรมชาติ โดยใช้ 2D Rotary Positional Embedding (RoPE) เพื่อระบุตำแหน่งข้อมูลภาพอย่างแม่นยำ

รองรับพิกเซลสูงสุดถึง 3.2 ล้านพิกเซล โดยไม่ต้องตัดต่อภาพ (Patching/Splitting) ทำให้เห็นรายละเอียดเล็กๆ ได้ชัดเจน

Sparkling Smiles Clinic Ratio Analysis

Ratio	Formula	2010	2020	2021	2022	Industry Average
Liguidity Ratios						
Ligudity Ratle	∨Formula = Agully Ratio	1.4%	1.8%	1.8%	1.2%	2.4%
Liguidity Ratio	∢Formula ∨ Liquidity	2.9%	2.5%	1.3%	2.6%	3.8%
Uguiblity Ratio	∢Formola ∨ Ascet Ihepumat Ratio	1.6%	1.6%	1.2%	1.6%	3.3%
Ligudity Aescmed Ratio	∨Formula ∨ Bevenned	17.6%	18.3%	18.8%	15.6%	2.7%
Liguidity Inwmeramsoot Finardal Ratio	∢Formula ∨ Reoonad Ratio	8.3%	8.6%	6.3%	6.6%	2.1%
Asset Management Ratios						
Aaset Ratio	∢Formula ∨ Liquidity	40.5%	54.4%	86.8%	22.2%	3.7%
Aaset Management Ratios	∨Aaset Baysmed Diesated	16.2%	18.9%	15.7%	23.7%	1.8%
Asset Ihawos Ratio	∨rtaset ∨ Asuot Asset	9.7%	15.2%	18.0%	18.5%	1.8%
Aaset Management Rew Ratio	∢FW ∨ Aseet	6.5%	8.6%	6.6%	6.6%	1.6%
Aaset Management Ratio	∢Formola ∨ Real Asset	3.2%	3.0%	3.7%	3.6%	0.2%
Asset Rapeiance Ratio	Formula ∨ Onifenenea IririOua.asset	36.4%	36.8%	36.8%	36.6%	2.2%
Solvency Ratios						
Soivency Ratio	∨Formola ∨ GMERNT Ehaoasid	31.7%	25.7%	64.5%	45.5%	5.5%
Solvency Ratio	∨Formula ∨ SNWOSndnnical	5.6%	5.7%	1.8%	5.2%	3.6%
Comeerser Solvency Ratio	Formula ∨ LSIMB Acset	5.6%	5.8%	5.6%	5.7%	1.9%
Solvency Ratio	Formula ∨ Underssinettivent Ratio	27.5%	30.2%	28.7%	26.6%	2.7%
Profitability Ratios						
Prelitability Ratios	B7T8 ∨ Ionoset	13.3%	10.1%	6.7%	10.6%	3.7%
Prcilitability Raties	6PP5 ∨ Inocue	11.3%	16.7%	11.3%	19.2%	4.5%
Buoness Ratio	8P81 ∨ Pkaiks	10.2%	12.5%	21.2%	20.6%	3.4%
Ressitisn Inteome Ratio	SGOM ∨ Weirbuna	8.4%	3.5%	6.2%	3.6%	0.2%
Penitene Madie Ratio	SGOM ∨ Ganns	6.3%	3.6%	6.1%	6.6%	6.7%

InfoVQA:

83.2%

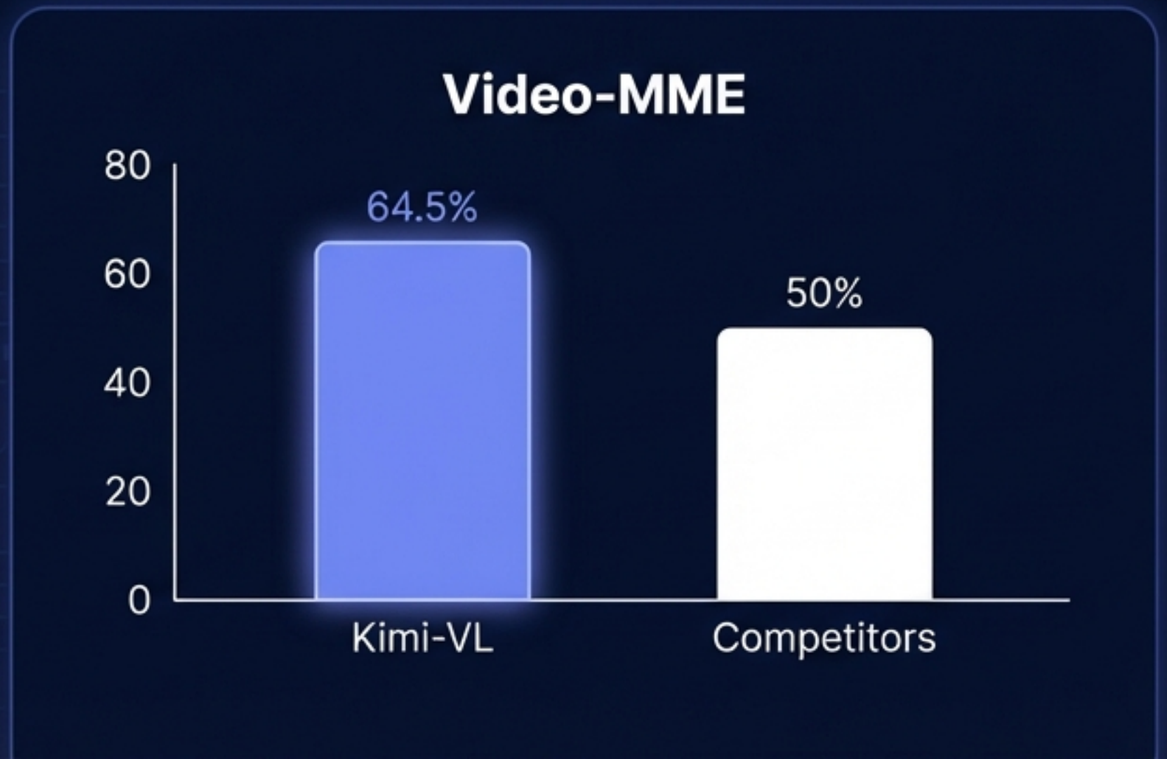
ScreenSpot-Pro:

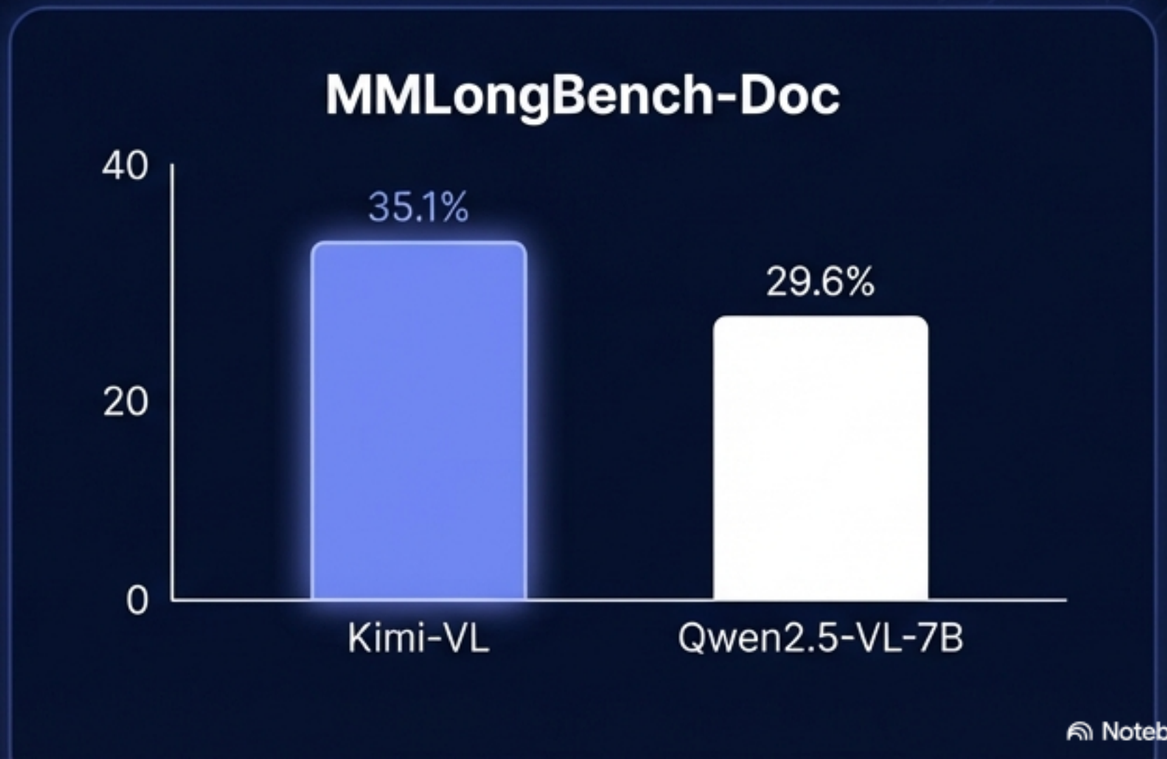
34.5%

NotebookLM

Pillar 2: ‘Long’ - รองรับบริบทข้อมูลขนาดใหญ่ถึง 128K

- Needle-in-a-Haystack: ความแม่นยำในการค้นหาข้อมูล (Recall) เกือบ 100% ในบริบท 128k ทั้งในรูปแบบข้อความและวิดีโอ
- Use Cases: วิเคราะห์วิดีโอความยาว 1 ชั่วโมง หรือเอกสารมากกว่า 100 หน้าได้ในการป้อนข้อมูลครั้งเดียว





NotebookLM

kimi-ai.

Pillar 3: ‘Smart’ - ความฉลาดในการให้เหตุผลทางคณิตศาสตร์

แม้จะมีขนาดเล็ก แต่ Kimi-VL สามารถแก้โจทย์คณิตศาสตร์และตรรกะที่ซับซ้อนได้ดีเยี่ยม แข่งขันได้กับโมเดลที่มีความหนาแน่น (Dense models) ขนาดใหญ่กว่า

Metric	Kimi-VL-A3B	GPT-4o	Qwen2.5-VL-7B
MathVista (Pass@1)	**68.7%**	63.8%	68.2%
MMMU (Val)	**57.0%**	N/A	58.6%

NotebookLM

ขั้นตอนการฝึกฝน (Training Pipeline) ที่แม่นยำ 5 ระยะ

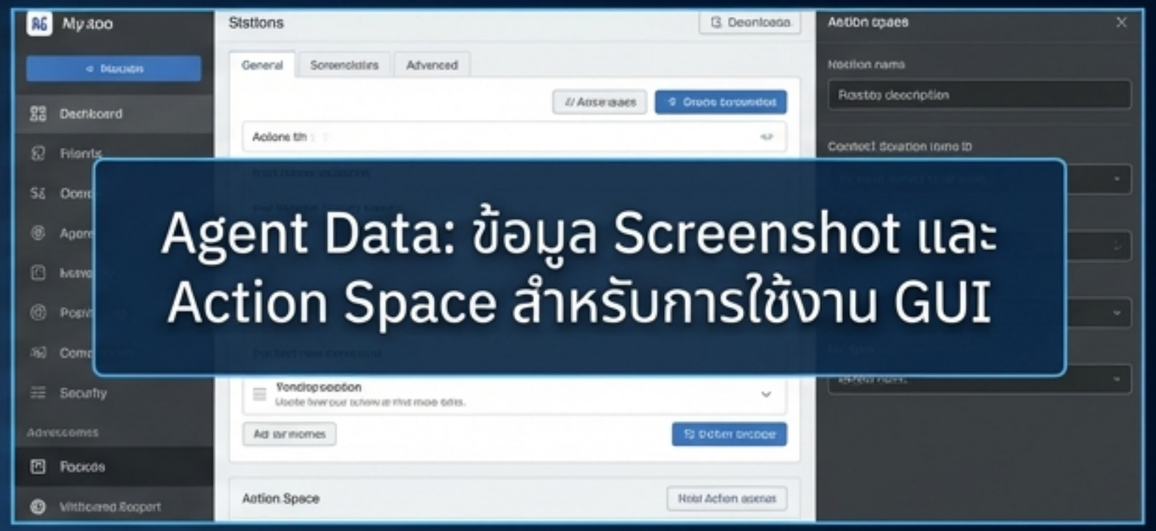
1. Text Pre-training	2. ViT Training	3. Joint Pre-training	Joint Cooldown	5. Joint Long-context
5.2T Tokens (Pure Text)	2T+ Tokens (Image-Text Pairs)	1.4T Tokens (Vision & Language)	0.6T Tokens (High Quality Data)	0.3T Tokens (128k Context)

ขั้นตอน 'Cooldown' คือกุญแจสำคัญที่ใช้ข้อมูลสังเคราะห์คุณภาพสูงเพื่อเพิ่มความแม่นยำ

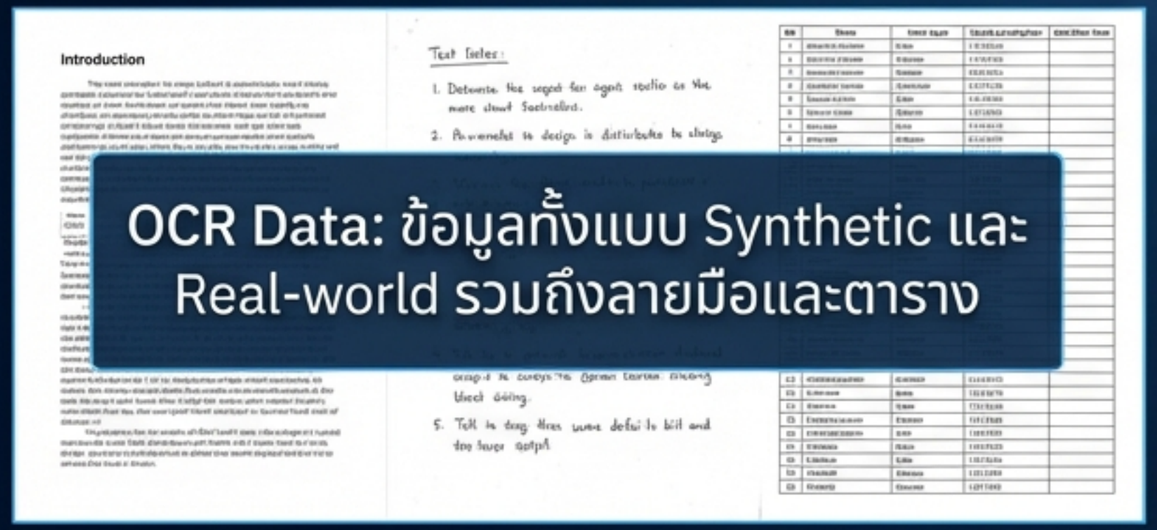
NotebookLM

กลยุทธ์การสร้างข้อมูล: หลากหลายและคุณภาพสูง

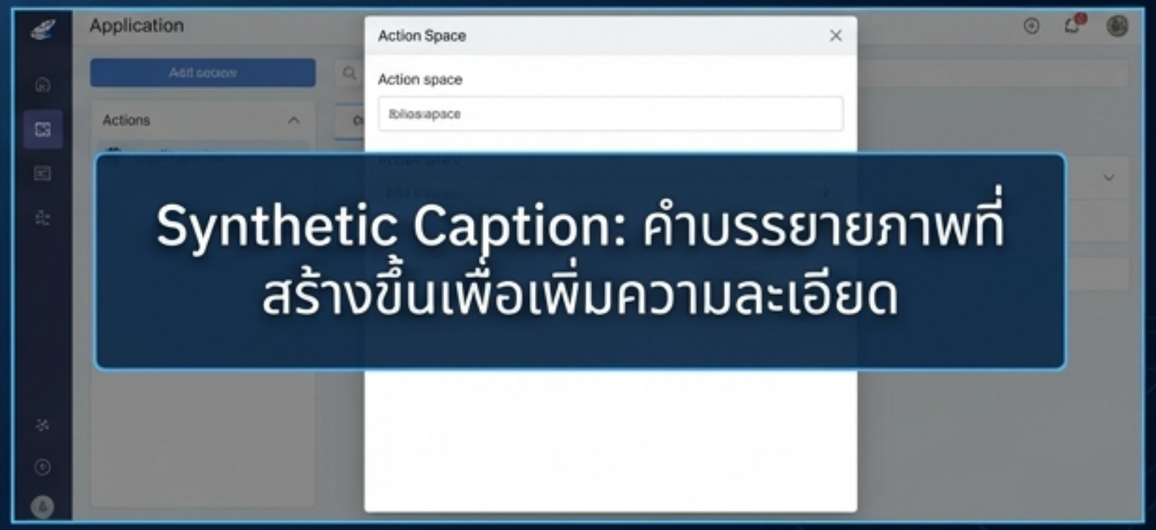
Interleaving Data: ข้อมูลภาพสลับข้อความ ช่วยรักษาความสามารถทางภาษา



Agent Data: ข้อมูล Screenshot และ Action Space สำหรับการใช้งาน GUI



OCR Data: ข้อมูลทั้งแบบ Synthetic และ Real-world รวมถึงลายมือและตาราง

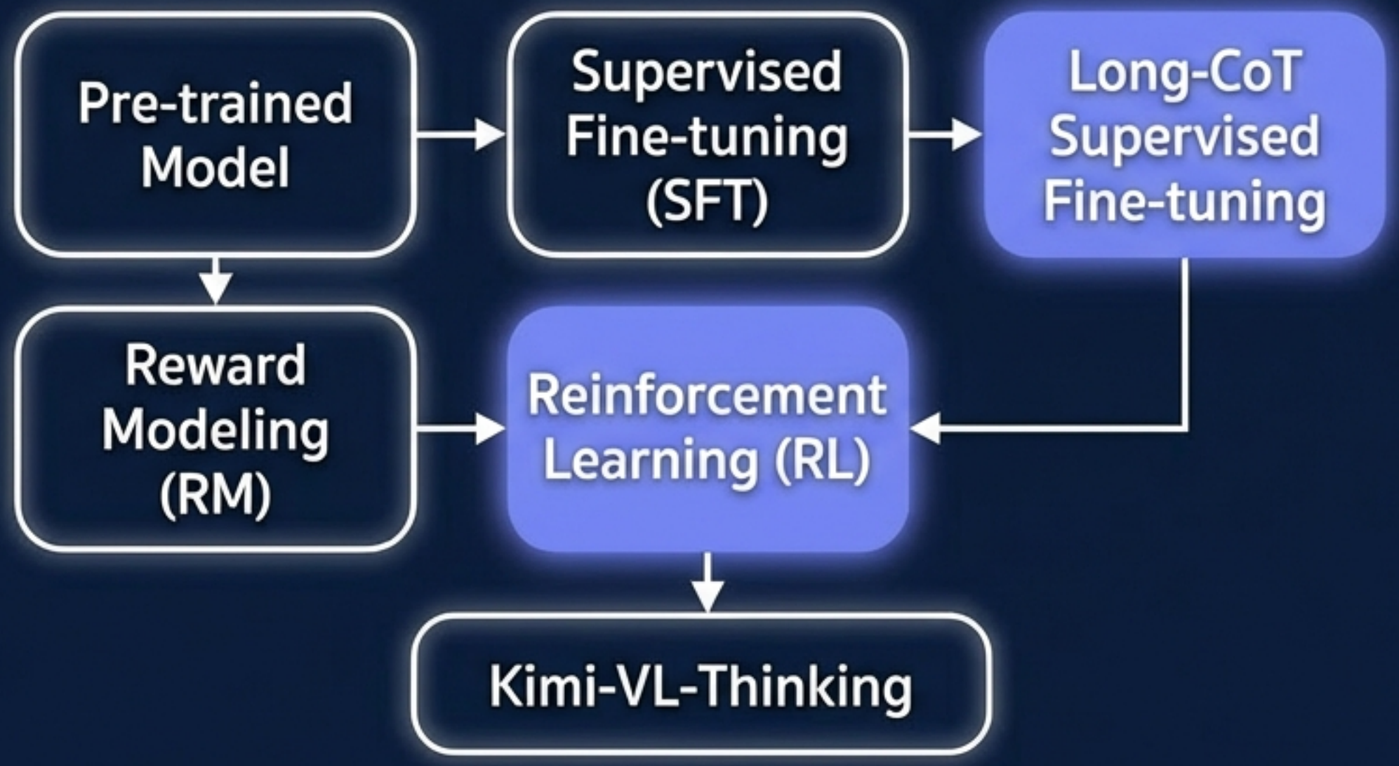


Synthetic Caption: คำบรรยายภาพที่สร้างขึ้นเพื่อเพิ่มความละเอียด

NotebookLM

ก้าวไปอีกขั้นกับ Kimi-VL-Thinking

เพิ่มระบบการคิดแบบ ‘System 2’ ให้กับ Vision-Language Model



- **Long Chain-of-Thought (CoT) SFT:** การปรับแต่งด้วยข้อมูลกระบวนการคิดที่ยาวและซับซ้อน
- **Reinforcement Learning (RL):** การเรียนรู้แบบเสริมกำลังเพื่อสอนให้โมเดล ‘วางแผน’ และ ‘ตรวจสอบตัวเอง’

NotebookLM

กระบวนการฝึกฝนระบบการคิด (RL Training)

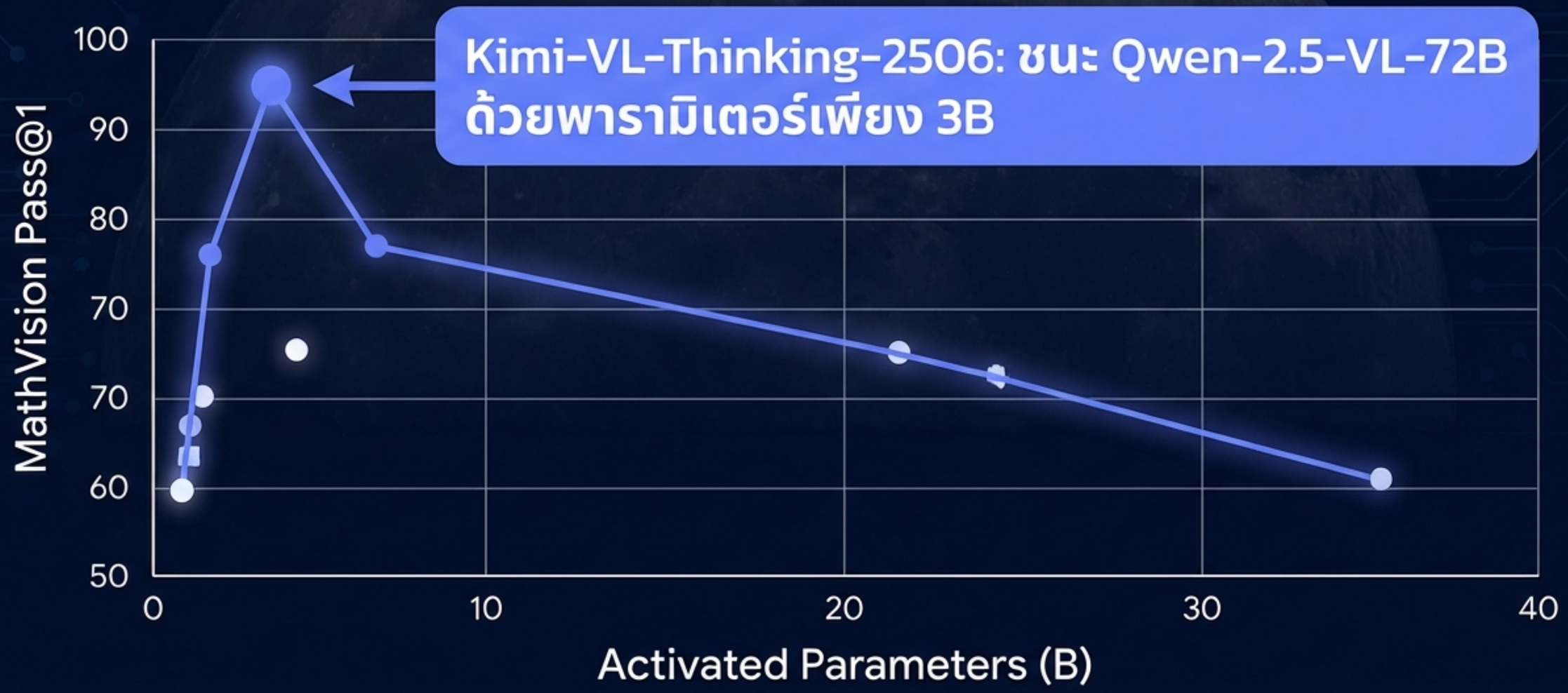
Online RL: อัปเดตโมเดลแบบวนซ้ำ (Iterative updates)

Curriculum Sampling: เลือกโจทย์ตามระดับความยากเพื่อไแ่ให้โมเดลเรียนรู้เป็นลำดับขั้น

Length Penalty: บทลงโทษเพื่อป้องกันการคิดวนซ้ำที่ยาวเกินความจำเป็น

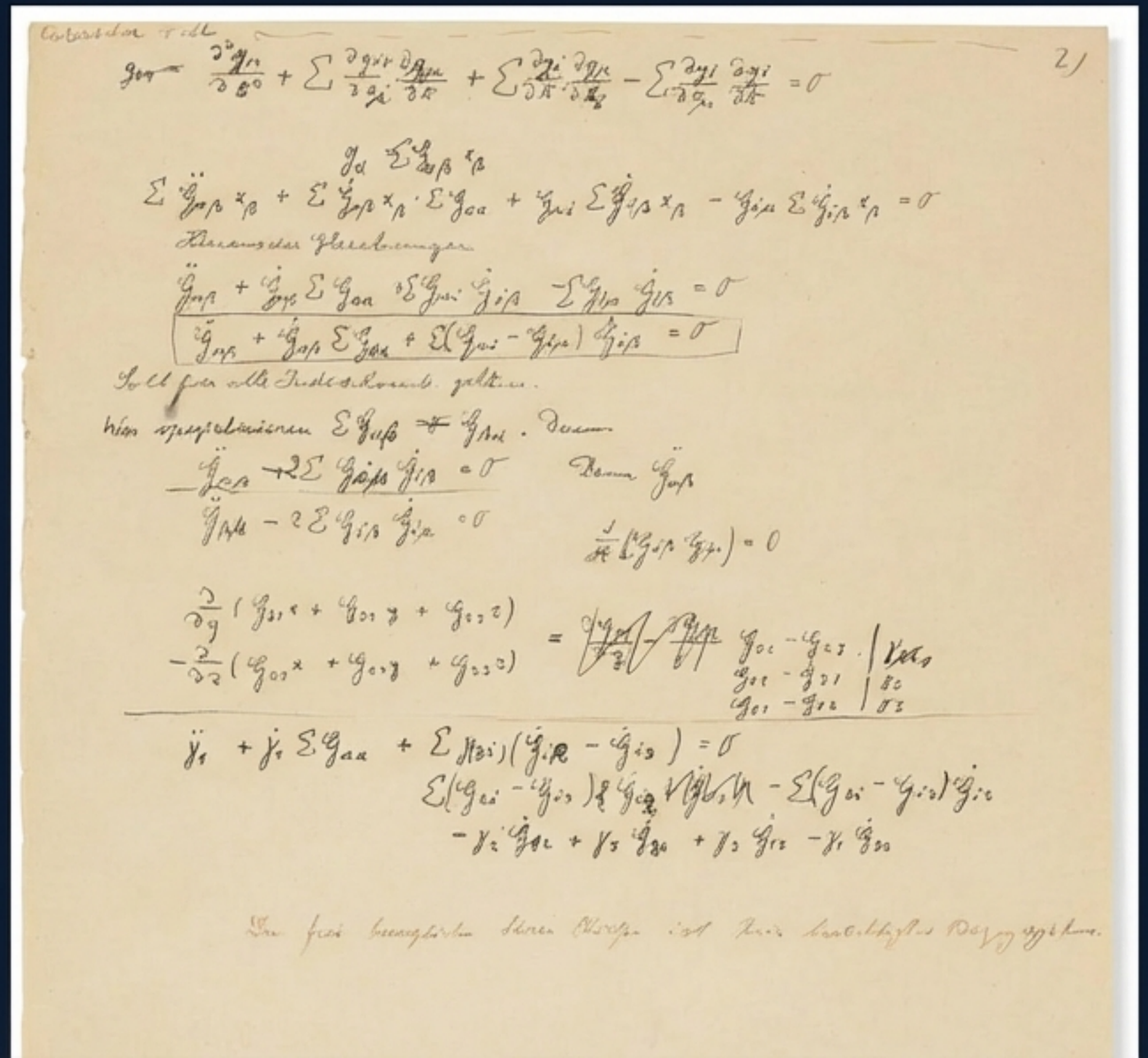
NotebookLM

ผลลัพธ์จากการ ‘คิด’ บนเกณฑ์วัดผล MathVision



NotebookLM

กรณีศึกษา 1: การอนุมานจากเอกสารลายมือเขียน (Manuscript Deciphering)

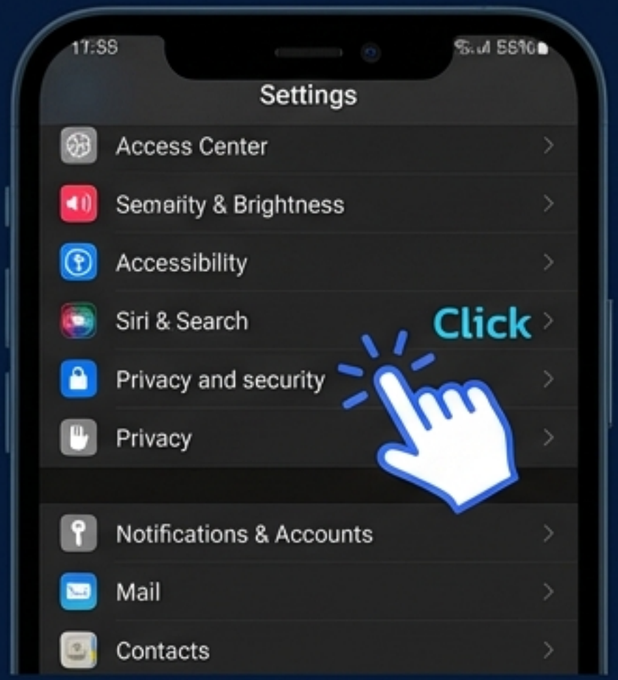


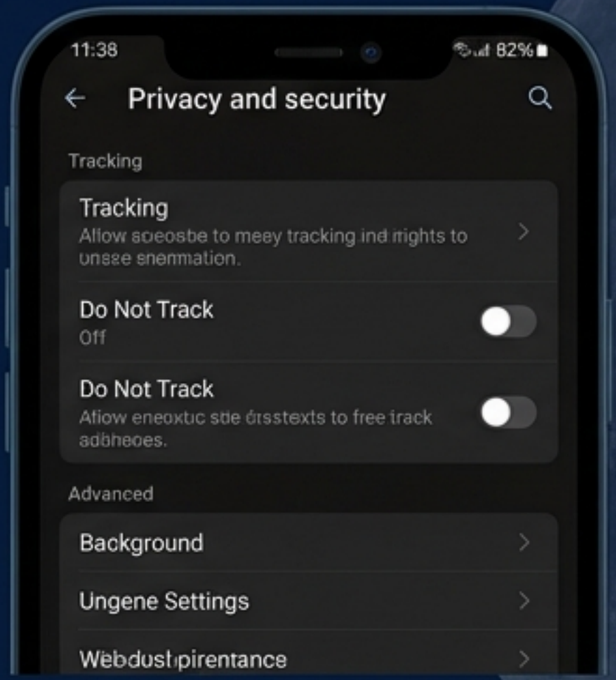
"The manuscripts belong to Albert Einstein... based on handwriting style, content analysis, and language cues."

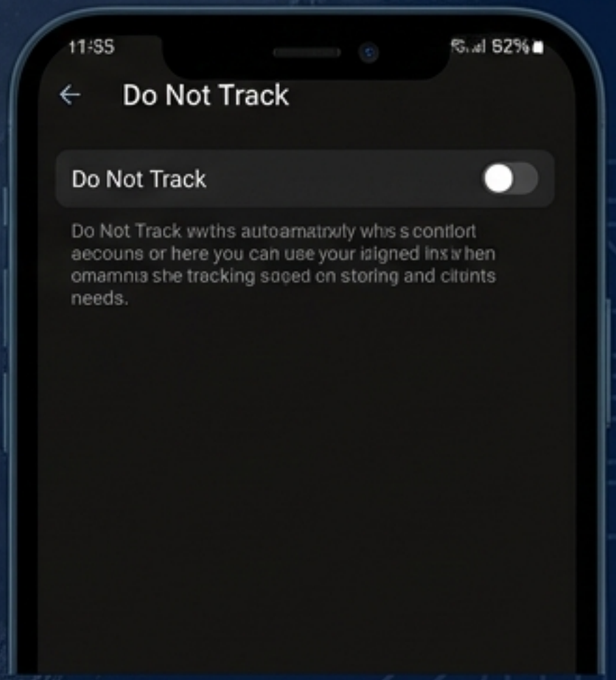
โมเดลวิเคราะห์สไตล์การเขียน, สัญลักษณ์ทางคณิตศาสตร์ (Tensor calculus), และภาษาเยอรมัน เพื่อระบุว่านี่คือต้นฉบับงานวิจัย 'General Relativity' ของ Albert Einstein

NotebookLM

กรณีศึกษา 2: AI Agent และการนำทางบนหน้าจอ (GUI Navigation)







- 1. **See:** มองเห็นองค์ประกอบบนหน้าจอ (Settings menu)
- 2. **Plan:** วางแผนลำดับขั้นตอน (คลิกเมนู -> Settings -> Privacy)
- 3. **Act:** สั่งการคลิกและเลื่อนหน้าจอ (Scroll/Click) จนสำเร็จภารกิจ

OSWorld Pass@1:
8.22% (ชนะ GPT-4o)

NotebookLM

กรณีศึกษา 3: ความเข้าใจวิดีโอขนาดยาว (Long Video Understanding)

Scene Segmentation:
แบ่งฉากวิดีโอตามเนื้อหา
ได้อย่างแม่นยำ

Timestamping:
ระบุเวลาเริ่ม-จบของ
แต่ละเหตุการณ์

Detailed Description:
บรรยายรายละเอียด
(เช่น การปีนเขา, กงล้อมเนต์)

NotebookLM

บทสรุป: มาตรฐานใหม่ของ Multimodal AI ที่ทรงพลังและประหยัด

- **Efficient:** 2.8B Activated Parameters ประหยัดทรัพยากร
- **Capable:** ครอบคลุมทั้ง Smart (Reasoning), Long (128k), และ Clear (Native Res)
- **Advanced:** ระบบ Thinking ที่ยกระดับการแก้ปัญหาซับซ้อน

Open Source: Code และ Model เผยแพร่สู่สาธารณะแล้ว
Link: https://github.com/MoonshotAI/Kimi-VL

NotebookLM