ImageNet จบแล้วหรือยัง?

Are we done with ImageNet?

การประเมินมาตรฐานทองคำใหม่
และก้าวต่อไปของ Computer Vision

สรุปงานวิจัยโดย Google Brain & DeepMind (Beyer et al., 2020)

NotebookLM

ทศวรรษแห่ง-ความก้าวหน้า (The Legacy)

2012
AlexNet
อเล็กซ์เน็ต

2014
VGG-16
วีจีจี-16

2015
ResNet-50
เรสเน็ต-50

2020
Noisy Student
นอยซี สติวเดนต์

Exhibit
มาตรฐาน
The Benchmark

ตลอด 10 ปีที่ผ่านมา ILSVRC หรือ 'ImageNet' คือสนามประลองหลักที่ผลักดันให้เกิดนวัตกรรมอย่าง AlexNet และ ResNet

Exhibit
ข้อสมมติฐาน
The Assumption

เราเชื่อมาตลอดว่า 'ความแม่นยำ' (Accuracy) ที่สูงขึ้นบน ImageNet แปลว่าโมเดลมีความฉลาดและความเข้าใจภาพทั่วไป (Generalization) มากขึ้น

Exhibit
สถานะปัจจุบัน
Status Quo

โมเดลยุคใหม่ทำคะแนนได้สูงขึ้นเรื่อยๆ จนดูเหมือนว่าเรากำลังจะแก้โจทย์นี้ได้สมบูรณ์แล้ว

NotebookLM

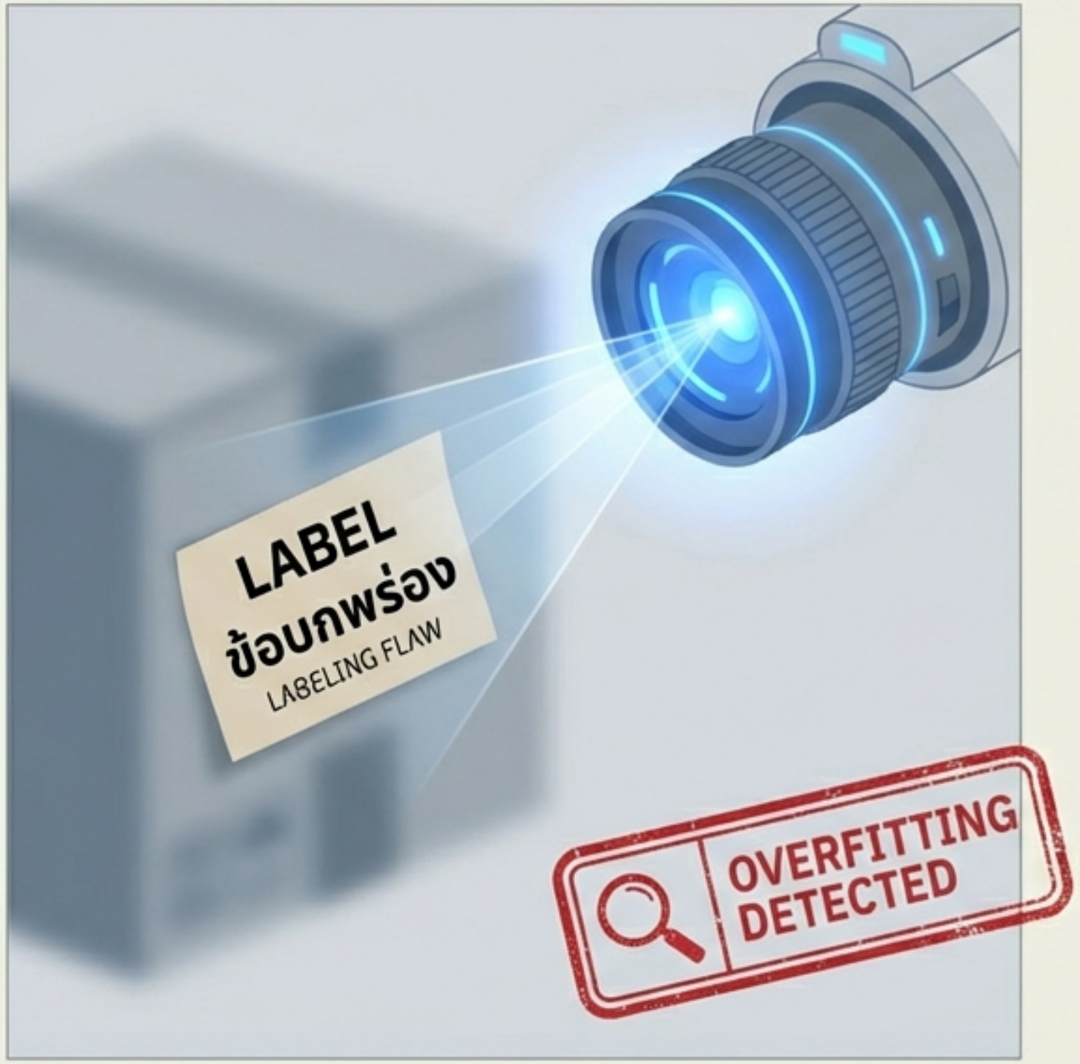


EXHIBIT B: OVERFITTING TO THE LABEL

สัญญาณเตือน: เรากำลังเดินหลงทาง? (The Suspicion)

งานวิจัยตั้งคำถามสำคัญ: ความแม่นยำที่เพิ่มขึ้นในช่วงหลัง เกิดจากการที่ AI เข้าใจภาพดีขึ้นจริง หรือแค่เรียนรู้ที่จะ ‘Overfit’ กับข้อบกพร่องของคนแปะ-ป้าย (Labeling procedure)?

The Hypothesis: เมื่อโมเดลเริ่มเก่งมากๆ คะแนนที่เพิ่มขึ้นอาจไม่ได้สะท้อนความสามารถในการมองเห็น (Perception) แต่สะท้อนความสามารถในการเดาใจคนเฉลยข้อสอบ

NotebookLM

รอยร้าวที่ 1: หนึ่งภาพไม่ได้มีแค่หนึ่งคำตอบ (Single Label Limit)

EXHIBIT A: SINGLE LABEL LIMIT

The Problem

ImageNet บังคับให้หนึ่งภาพมี ‘คำตอบที่ถูก’ เพียงข้อเดียว (Single label annotation)

The Flaw

ถ้าโมเดลตอบว่า ‘ค้อน’ แต่เฉลยในระบบคือ ‘สว่าน’ โมเดลจะถูกตัดสินว่า **ผิด ทันที** ทั้งที่มันโ่ขันมองเห็นถูกต้อง สิ่งนี้**ทำให้การวัดผลคลาดเคลื่อนจากความเป็นจริง**

NotebookLM

รอยร้าวที่ 2: ความกำกวมและกับดักภาษา (Ambiguity & Noise)

EXHIBIT A: AMBIGUITY OF OBJECTS

ImageNet Label: Quill (ปากกาขนนก) ❌

Reality: Feather boa (ผ้าพันคอขนสัตว์) ✅

EXHIBIT B: LINGUISTIC TRAPS

Ambiguity: Laptop vs. Notebook

คอมพิวเตอร์พกพาเหมือนกัน

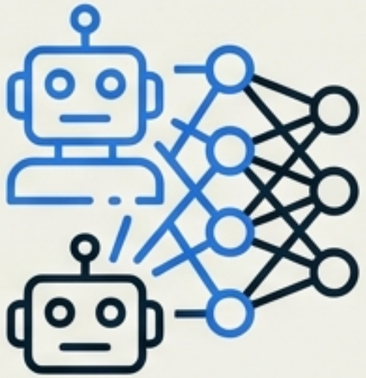
แต่มนุษย์แยกความต่างไม่ได้ชัดเจน

เมื่อข้อสอบมีความกำกวม โมเดลที่ทำคะแนนได้สูงคือโมเดลที่จำเก่ง ไม่ใช่โมเดลที่เข้าใจเก่ง

NotebookLM

สร้างไม้บรรทัดอันใหม่: ReaL Labels

ทีมวิจัยสร้างชุดเฉลยใหม่สำหรับ Validation Set ชื่อว่า ReaL (Reassessed Labels)



Step 1: Proposals
ให้โมเดล 19 ตัวช่วยกันเสนอคำตอบที่เป็นไปได้

Step 2: Human Verification
ให้มนุษย์ (Raters) ตรวจสอบว่าคำตอบไหน ‘ใช่’ บ้าง (Multi-label)

Step 3: Quality Control
ใช้สถิติกรองความผิดพลาดและใช้ Expert ในหมวดสัตว์ยากๆ

EXHIBIT C: REAL LABELS PROCESS

NotebookLM

เปรียบเทียบมาตรฐาน: ImageNet vs. ReaL

Original ImageNet Labels	ReaL (Reassessed Labels)
• มีคำตอบถูกเพียง 1 ข้อ (One-hot) • มี Noise และความกำกวมสูง • ตัดสินโมเดลว่าผิด หากตอบไม่ตรงกับเฉลยใบเดียว	• รองรับหลายคำตอบ (Multi-label) • ผ่านการกรองโดย Expert และโมเดลหลายตัว • วัดผลที่ความแม่นยำจริง: ‘ถ้าโมเดลตอบสิ่งที่อยู่ในภาพได้ ถือว่าถูก’

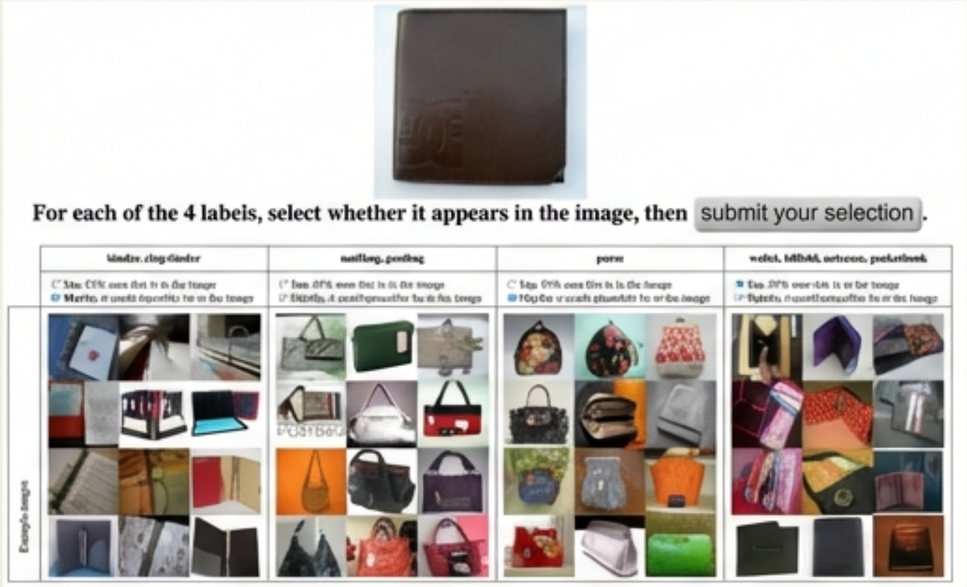
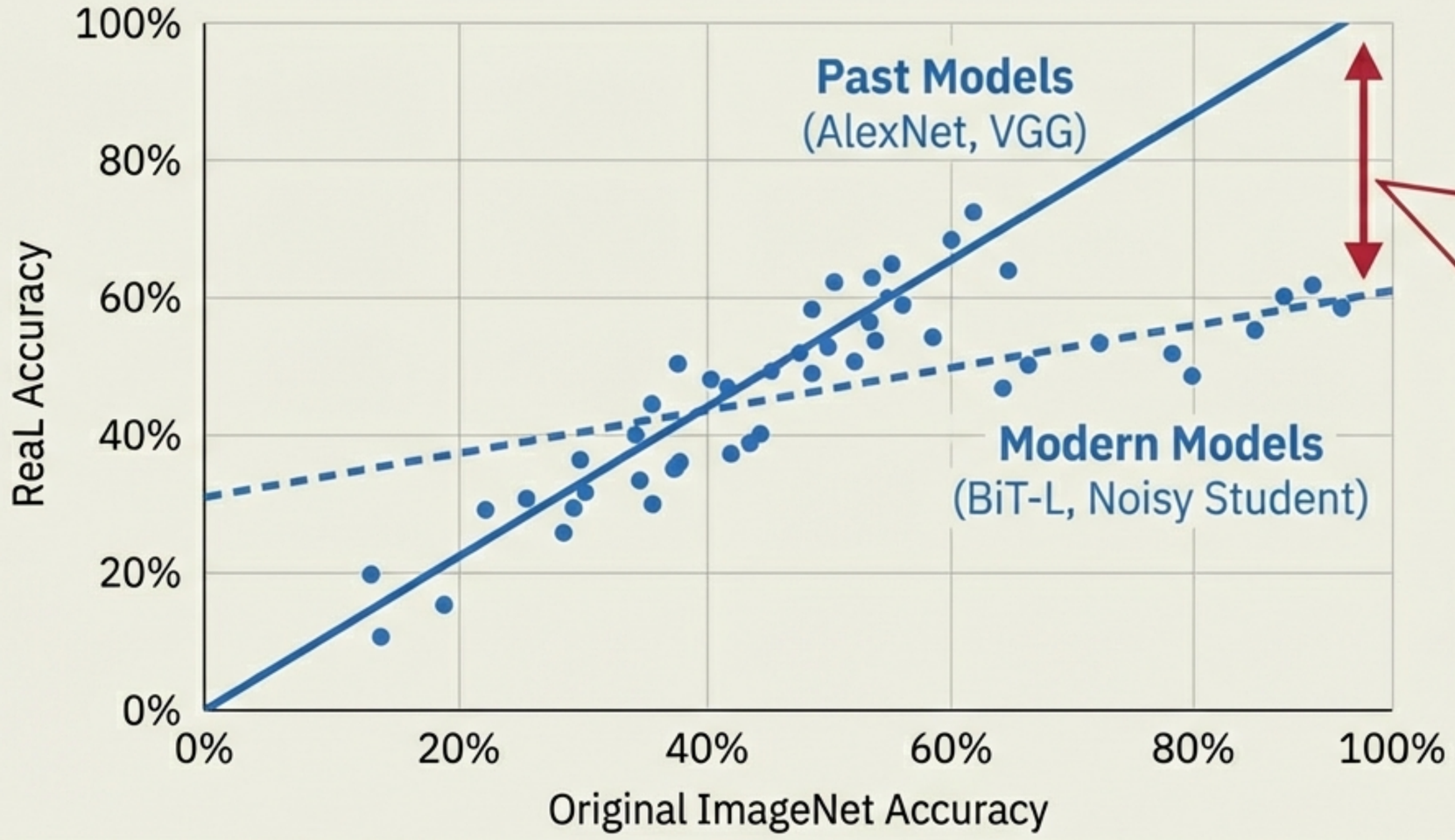


EXHIBIT D: MULTI-LABEL CONCEPT (ReaL INTERFACE)

NotebookLM

ความจริงที่ปรากฏ: กราฟที่แยกออกจากกัน (The Divergence)

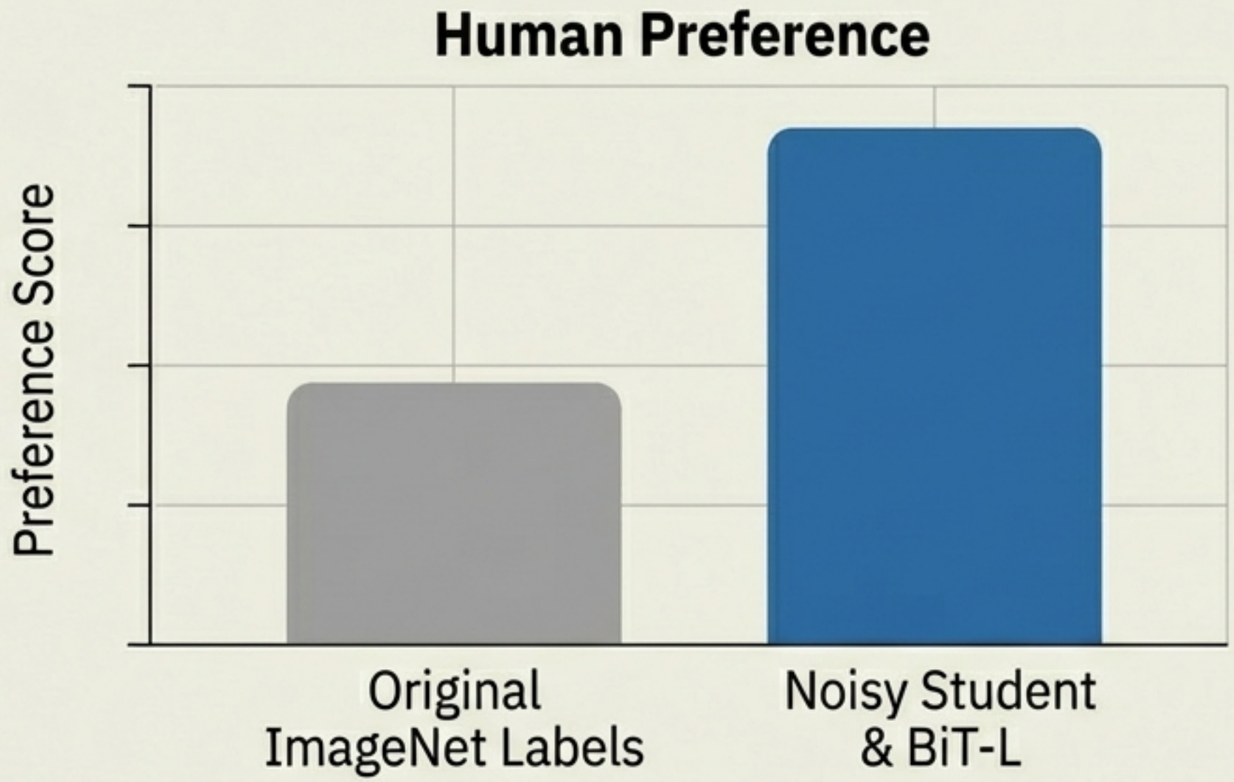


Insight: โมเดลยุคใหม่กรนเนหุรังม่ คะแนน ImageNet สูงขึ้น แต่ความถูกต้องจริง (ReaL) เริ่มไม่เพิ่มตาม
–> Overfitting

EXHIBIT E: MODEL ACCURACY DIVERGENCE (SCATTER PLOT)

NotebookLM

เมื่อ AI แม่นกว่าคนเฉลยข้อสอบ



จากการประเมินด้วยมาตรฐานใหม่ (ReaL) พบเรื่องน่าตกใจ:

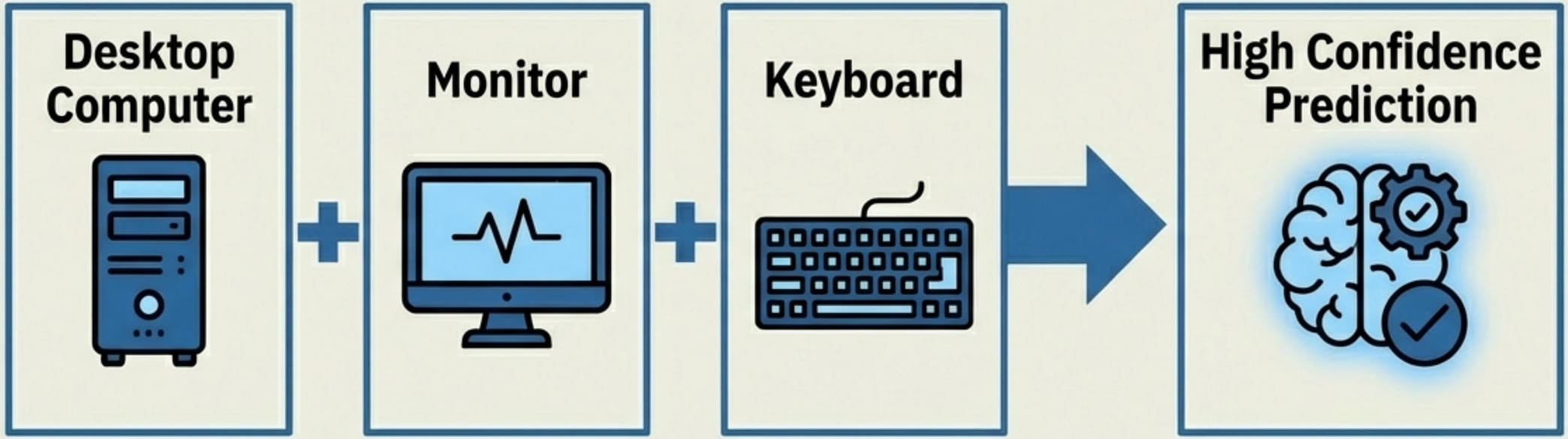
- โมเดลชั้นนำอย่าง Noisy Student และ BiT-L ทำนายสิ่งที่อยู่ในภาพได้ตรงใจมนุษย์มากกว่าป้ายกำกับเดิมเสียอีก

ฉลาก ImageNet เดิม ไม่ใช่ตัวแทนความถูกต้องอีกต่อไป

EXHIBIT F: HUMAN PREFERENCE BAR CHART

NotebookLM

กับดักความสัมพันธ์ (Co-occurrence Bias)



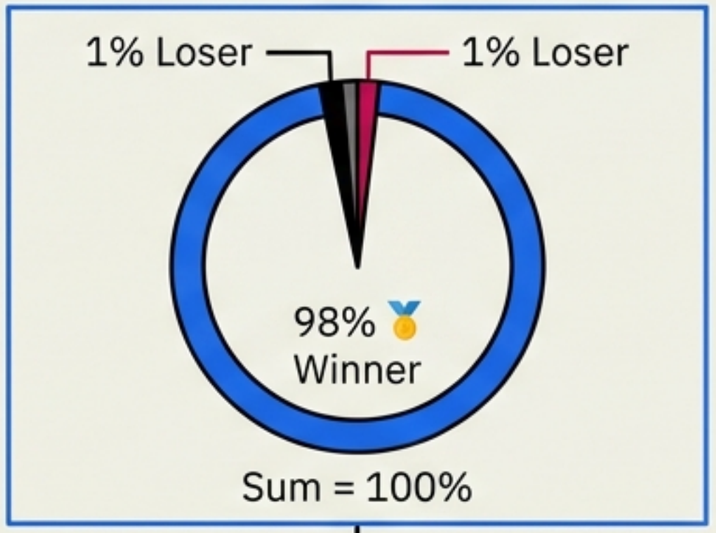
- โมเดลเรียนรู้ที่จะ ‘เดา’ จากบริบทมากกว่าการมองเห็นวัตถุนั้นจริงๆ
- ตัวอย่าง: ใน ImageNet รูป ‘Desktop Computer’ มักมาคู่กับ ‘Monitor’ (87%)
- Result: โมเดลทายถูกถึง 71.1% เพราะมันจำ Pattern ว่าถ้าเห็นจอคอมฯ ให้ทายว่า Desktop ไว้ก่อน แม้ในภาพนั้น Desktop อาจจะถูกบังอยู่ก็ตาม

EXHIBIT G: CO-OCCURRENCE BIAS SCHEMATIC

NotebookLM

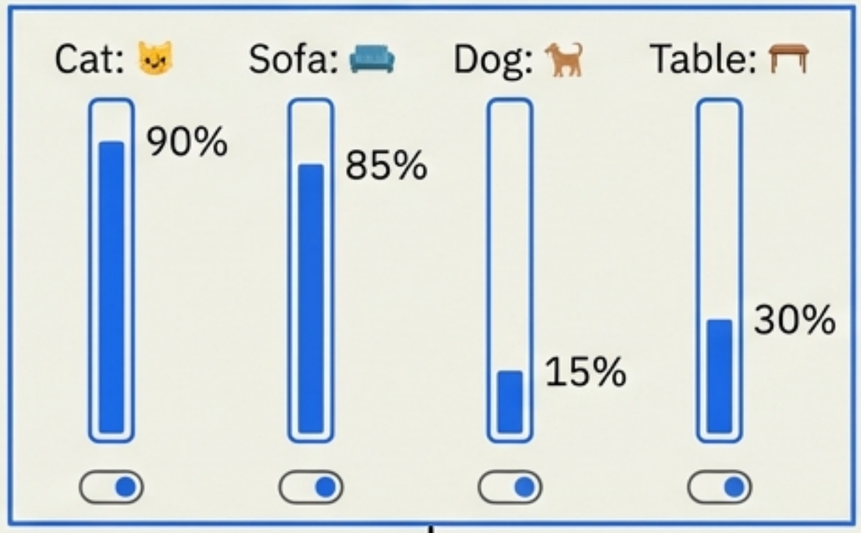
ทางออกที่ 1: เลิกบังคับเลือกด้วย Sigmoid Loss

Old Way (Softmax)



บังคับเลือก 1 คำตอบ
(Winner takes all)

New Way (Sigmoid Loss)



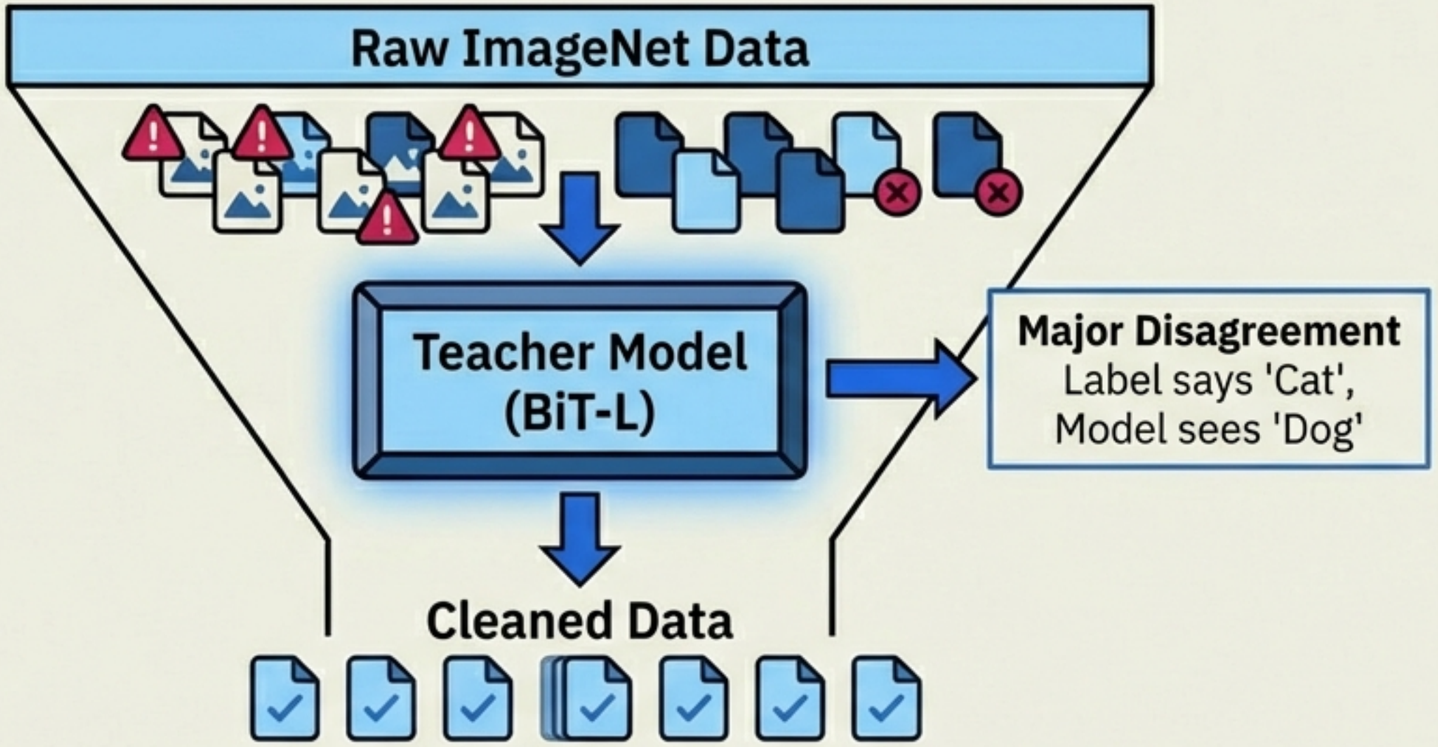
อิสระต่อกัน
(Independent Binary Classification)

อนุญาตให้โมเดลทำนายหลายวัตถุพร้อมกันได้
ลดการลงโทษเมื่อตอบถูกแต่ไม่ตรงเฉลยใบเดียว

EXHIBIT H: SOFTMAX VS. SIGMOID LOSS COMPARISON

NotebookLM

ทางออกที่ 2: ล้างข้อมูลให้สะอาด (Clean Training Data)



Remove ~10% of noisy data

- Garbage In, Garbage Out: การลบข้อมูลที่ผิดพลาดช่วยให้โมเดลเรียนรู้ได้ดีขึ้น โดยเฉพาะเมื่อเทรนนานๆ

EXHIBIT I: CLEAN TRAINING DATA FUNNEL

NotebookLM

ผลลัพธ์: เก่งขึ้นจริง ไม่ใช่แค่คะแนนเฟ้อ

EXHIBIT J: RESNET-50 (900 EPOCHS) RESULTS

78.5% Accuracy

with Sigmoid + Clean Data

vs 76.0% Baseline

- No Overfitting: ยิ่งเทรนนาน ยิ่งเก่งขึ้น แม้จะผ่านไป 900 epochs
- ReaL Score Improvement: ความถูกต้องบนมาตรฐาวใหม่เพิ่มขึ้น +1.6%

EXHIBIT J: RESNET-50 (900 EPOCHS) RESULTS

NotebookLM

บทสรุป: มาตรฐานใหม่แห่งยุคถัดไป

EXHIBIT 1:

The Limit

โมเดล AI ฉลาดเกินกว่าที่ ImageNet ฉบับดั้งเดิมจะวัดผลได้เที่ยงตรงแล้ว

EXHIBIT 2:

The Reality

ReaL Labels ให้ภาพความเป็นจริงได้ดีกว่า Accuracy แบบเดิม

EXHIBIT 3:

The Solution

ใช้ Sigmoid Loss และ Clean Data เพื่อปลดล็อกศักยภาพที่แท้จริง

NotebookLM

ก้าวต่อไปของ Computer Vision (Future Outlook)

ImageNet ยังไม่ตาย แต่เราต้องใช้มันอย่างระมัดระวัง

ถึงเวลาที่นักพัฒนาจะเลิกไล่ล่าตัวเลขบนกระดาษ และหันมาสร้างโมเดลที่เข้าใจความซับซ้อนของโลกความจริง (ReaL World) อย่างแท้จริง

NotebookLM