

ความลับที่ซ่อนอยู่ในสถิติ:
อักษรวอยนิชคือภาษาลับ
หรือเป็นเพียงเกมลูกเต๋าลอง?
จากสมุดโบราณอายุ 600 ปี สู่คณิตศาสตร์เบื้องหลัง AI

[ความน่าจะเป็น ม.5]

[ตัวแปรสุ่ม ม.6]

[ทฤษฎีข้อมูล]



Cosham
Manuscript of Yoon



อายุ (Age): ค.ศ. 1404-1438
(ตรวจด้วยคาร์บอน 14)
ยุคเรเนซองส์ตอนต้น



ภาพประกอบ (Visuals):
พืชพันธุ์ที่ไม่มีอยู่จริงบนโลก,
แผนภูมิจักรวาลประเภท



หมึก (Ink): น้ำตึ๊ดไอ้
เขียนเส้นไหล ไม่มีรอยลบหรือชะงัก
(ผิดวิสัยการจารึกที่สลับที่
ที่ต้องหยุดคิดเทียบตาราง)



สถานะ (Status):
นักถอดรหัสระดับโลกยอมแพ้!
ยังไม่มีใครแปลได้แม้แต่ประโยคเดียว

ความขัดแย้งทางภาษาศาสตร์ (The Linguistic Paradox)

✓ เหมือนภาษาจริง

- กฎของซิปฟ์ (Zipf's Law):
ความถี่ของคำแปรผกผันกับ
อันดับความถี่
- ความยาวคำ (Word Length):
แจกแจงแบบโค้งปกติ
เฉลี่ย 5-6 ตัวอักษร

ขัดแย้งกับภาษาจริง ✖

- การซ้ำคำผิดธรรมชาติ (Reduplication): พบคำซ้ำติดกัน
บ่อยมาก เช่น qokedy qokedy dal
(แทบไม่พบในภาษายุโรป)
- กฎตำแหน่งที่แข็งทื่อ (Rigid Morphology): อักษรบางตัวถูก
บังคับให้อยู่แค่หัวคำ หรือท้ายคำ
เท่านั้น ขาดเครื่องหมายวรรคตอน

Apophenia & Pareidolia:
สมองมนุษย์วิวัฒนาการมาเพื่อ
'มองหาความหมาย' จาก 'ข้อมูลสุ่ม'
(Pattern-Seeking)
เพื่อเอาชีวิตรอด

ปหททกธส
๒๐๙๖๓๗
๕๓๘๓๓
๙๙๙๕๕๔

กลุ่มพรานนักวิจัย:
นักวิจัยกว่า 50 คนอ้างว่าแปลออก
(ฮีบรู, ละตินยุคกลาง, นาอิวเติล)
แต่ไม่มีใครแปลตรงกันเลย
เพราะสมองหลอกให้
เราเติมความหมายไปเอง!

เหตุการณ์อิสระ (Independent – ลูกเต๋า)

การทอยแต่ละครั้งไม่สนอดีต
ทอยได้หน้าเดิม โอกาสยังเท่าเดิมเสมอ

$$P(A \cap B) = P(A) \cdot P(B)$$

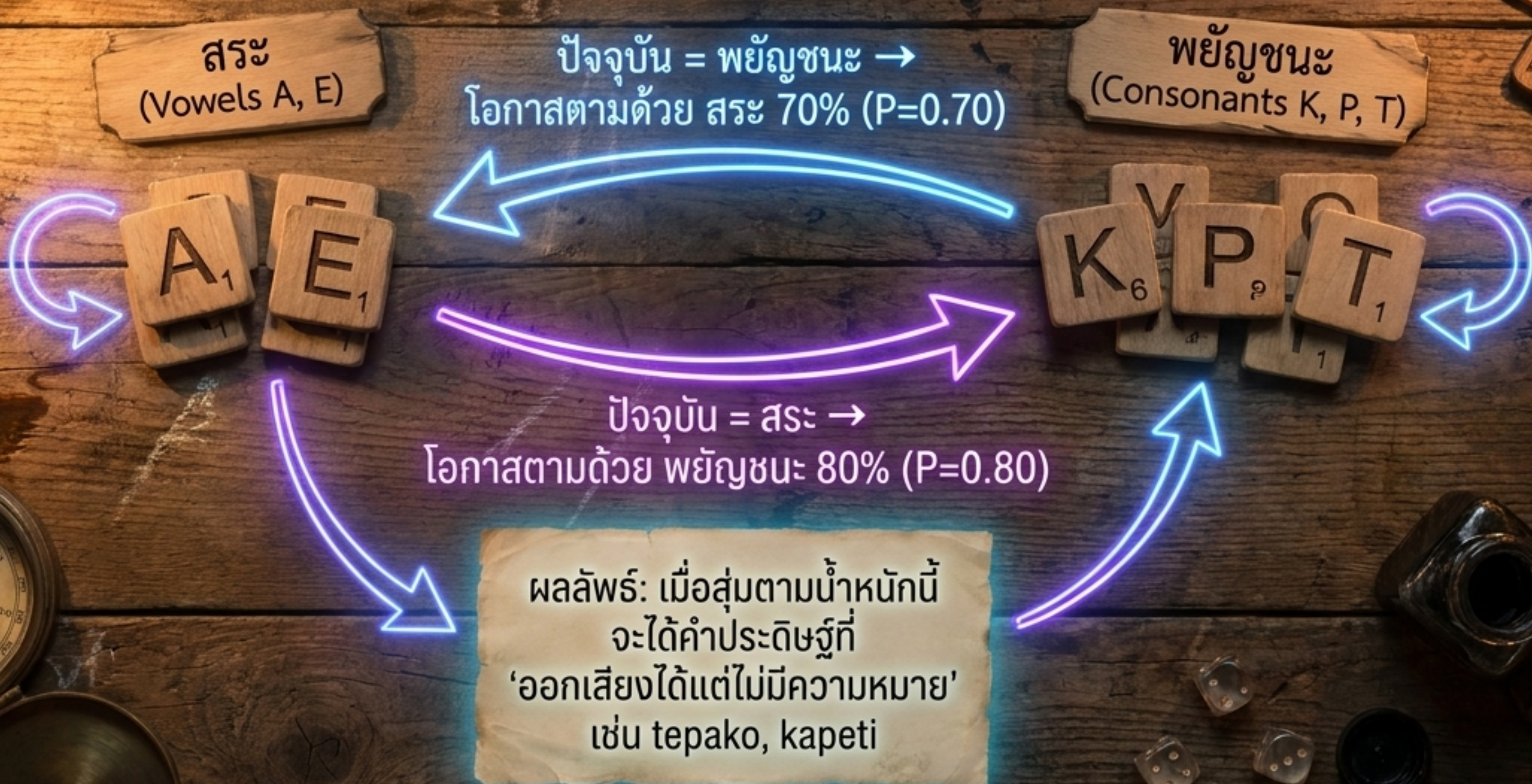
เหตุการณ์ที่มีเงื่อนไข (Dependent – จั่วไพ่)

การจั่วไพ่ (ไม่ใส่คืน)
ทำให้โอกาสเกิดสิ่งถัดไป
"ขึ้นอยู่กับ" อดีต

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

กระบวนการมาร์คอฟ (Markov Chain)

กฎเหล็ก: สัญลักษณ์ถัดไปถูกกำหนดโดยปัจจุบันเท่านั้น



สมมติฐานตะแกรงการ์ดั้น

(The Cardan Grille Hypothesis)

กลไก:

เลื่อนแผ่นเจาะรูสัมพันธ์
ทางไปบนตารางพยางค์

ความเร็ว:

สร้างข้อความคล้าย
วอยนิซได้เร็วถึง
1-2 หน้า/ชั่วโมง

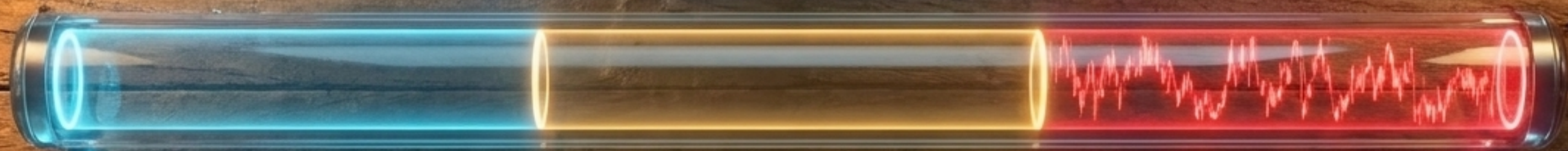
Prefix		Root		Suffix	
au	aú	aut	hi	bu	asa
bu	reil	hut	me	jał	buł
lio	hit	dac	mo	iut	liq
bod	ca	it	m	d	htg
eit	hi	coo	m	a	ni
ghm	ca	it	m	a	adom
aist	n	it	en	n	mau

ผลลัพธ์:

- อธิบาย 'กฎของชิปป์' ได้สมบูรณ์
- อธิบาย 'การซ้ำคำ' ได้ (เกิดเมื่อเลื่อนแผ่นไปตกช่องเดิมซ้ำสองครั้งพอดี!)

ทฤษฎีสารสนเทศ: มาตรวัดความไม่แน่นอน (Shannon Entropy: H)

$$H(X) = -\sum P(x_i) \log_2 P(x_i)$$

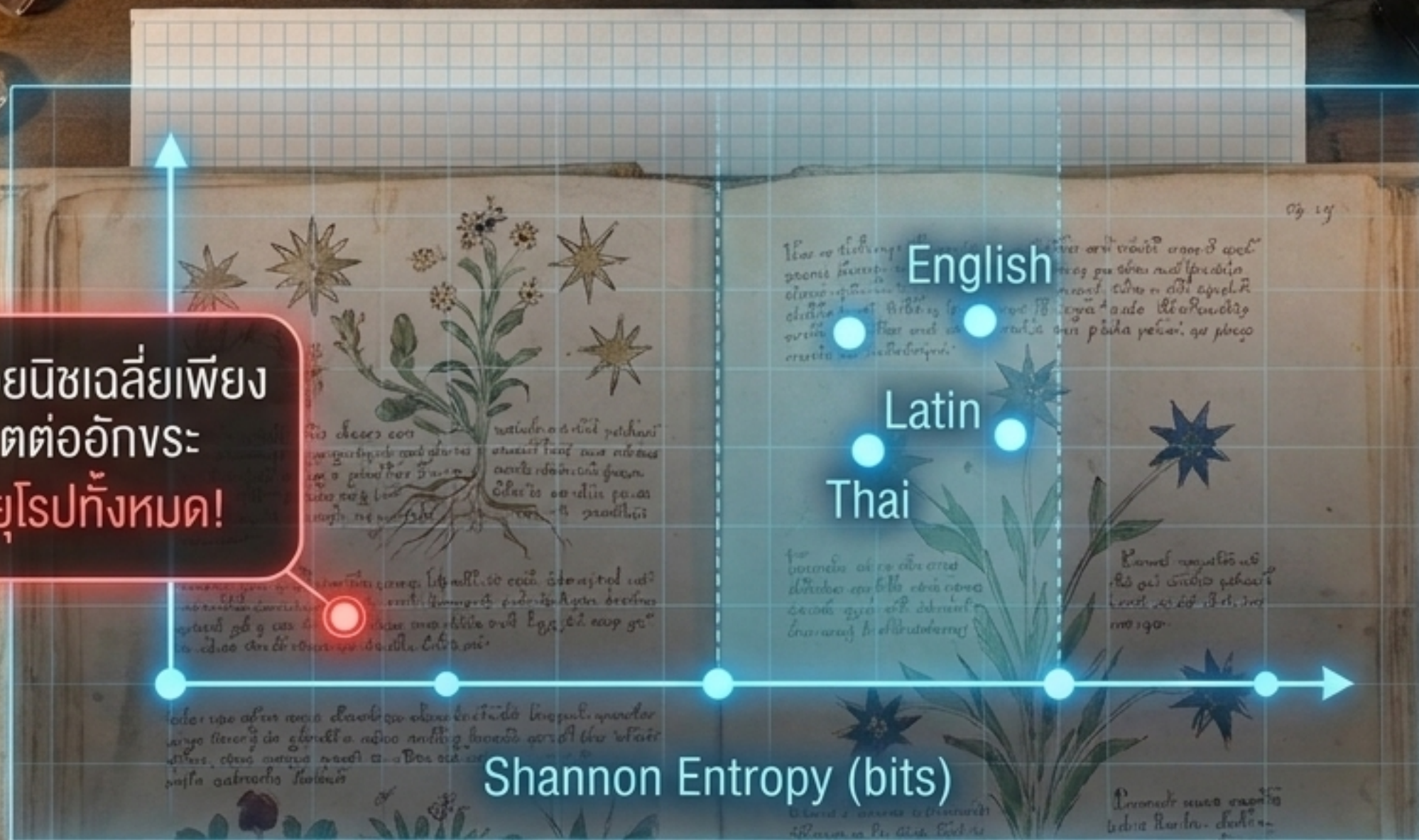


Low Entropy
โครงสร้างตายตัว
เช่น ababab (เดาง่ายสุด)

Moderate Entropy
ภาษารandomธรรมดา
(มีไวยากรณ์ผสมความหลากหลาย)

High Entropy
ขี้ส่ม White Noise
(เดาไม่ได้เลย ทุกตัวอักษร $P=1/n$)

เอนโทรปีของวอยนิชเฉลี่ยเพียง
3.5 - 3.8 บิตต่ออักขระ
ต่ำกว่าภาษายุโรปทั้งหมด!



ความหมาย: เมื่อรู้ตัวอักษรปัจจุบัน
ตัวถัดไปถูกบังคับเกือบตายตัว
= หลักฐานเชิงประจักษ์ว่ามันคือ
'การสุมผ่านเครื่องมือเชิงกล'
ไม่ใช่ภาษาที่มนุษย์ใช้สื่อสารจริง

จากวอยนิชสู่ AI ยุคใหม่ (Next-Token Prediction)

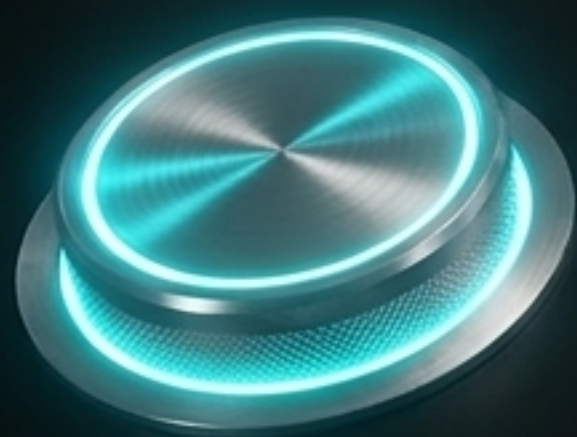


หลักการมาร์คอฟในศตวรรษที่ 15
คือรากฐานโดยตรงของโมเดลภาษาขนาดใหญ่
(LLMs) อย่าง ChatGPT ในปัจจุบัน!

โมเดลไม่ได้ 'เข้าใจ' ความหมาย แต่คำนวณ
'การแจกแจงความน่าจะเป็นของคำถัดไป'
จากบริบททั้งหมดในอดีต

สมการ Softmax:
$$P(w) = \text{Softmax}(z_w/T)$$

ควบคุมลูกเต๋าของ AI (ตัวแปร μ, σ)



อุณหภูมิ
(Temperature - T)

$T \rightarrow 0$: แม่นยำ
ตายตัว ไม่สร้างสรรค์
 $T > 1.2$: สุ่มกระจาย
เดายาก อาจหลุดบริบท
(High Entropy)

ตัดหางแถวทิ้ง
เลือกแค่คำที่มี
ความน่าจะเป็น
สูงสุด K อันดับแรก

Top-K

เลือกเฉพาะกลุ่มคำที่พรวน
ความน่าจะเป็นสะสมถึง
เกณฑ์ (เช่น 90%) ปิดกั้น
AI พ้นคำไร้สาระ

Top-P





THE HOAX

จุดร่วม: ทั้งสองสิ่งสร้างขึ้นจาก
“สถิติความน่าจะเป็น” ล้วนๆ ไม่ใช่
“ความเข้าใจในความหมายที่แท้จริง”



THE HALLUCINATION

เมื่อ AI ขาดข้อมูลจริง มันจะประกอบ
คำที่ไวยากรณ์สละสลวยแต่มีข้อออกมา
เหมือนที่อักษรรอยนิชหลอนนักวิชาการ
ระดับโลกมาได้ 600 ปี!

AI Hallucination = The Modern Voynich

บทสรุปทางวิชาการ: ความลับที่ถูกเปิดเผย

1. รหัสลับ (Encrypted)

❌ โต๊ะแย้งด้วยสถิติ:
รหัสลับต้องมีเอนโทรปีสูง
แต่อย่างน้อยก็ต่ำมาก

2. ภาษาที่สาบสูญ (Lost Language)

❌ โต๊ะแย้งด้วยสถิติ:
มีการซ้ำคำผิดธรรมชาติ
พยางค์แข็งที่ต่อเนื่องไป

3. กลลวง (Hoax / Procedural)

✅ สนับสนุนด้วยสถิติ:
ตรงกับหลักการ Markov
และ Cardan Grille
ทุกประการ (เอนโทรปีต่ำ
+ กฎของชิปป์ + คำซ้ำ)

คณิตศาสตร์คือแว่นขยายส่องหาความจริง

Łożyż mylihae o wynich

```
probability = model.predict_proba(X_test)[:; 1]
```

$P(b)$

$P(ib)$

ไม่ว่าจะเป็นกลองยูกกลาง หรือข้อมูลเทียบจาก AI...
ความเข้าใจเรื่อง 'ความน่าจะเป็นและสถิติ (Probability & Statistics)'
คืออาวุธสำคัญที่สุด ที่จะช่วยให้เราแยกแยะ 'ความจริง'
ออกจาก 'ภาพลวงตาทางสถิติ' ได้เสมอ

ทดลองและเรียนรู้เพิ่มเติมกับ
Panya AI Tutor

Math & AI Cheatsheet (สรุปสาระสำคัญสำหรับทบทวน)

Q1: ความน่าจะเป็น ม.5

- Independent vs Dependent Events
- Markov Chain (สถานะถัดไปขึ้นกับปัจจุบันเท่านั้น: $\sum P_{ij} = 1$)

Q2: ตัวแปรสุ่ม ม.6

- Next-Token Prediction & Softmax Equation
- Temperature / Top-K / Top-P (ควบคุมความน่าจะเป็นสะสม)

Q3: ทฤษฎีข้อมูล (Information Theory)

- Shannon Entropy (H) วัดความไม่แน่นอน (วอยนิช = 3.5 - 3.8 บิต)
- Zipf's Law ($f(r) \propto 1/r$)

Q4: จิตวิทยาการรู้คิด (Cognitive Science)

- Apophenia (สมองหาความหมายในความสุ่ม)
- AI Hallucination (ภาพหลอนจากการอิงสถิติโดยไร้ความจริง)