



# จริยธรรมในการใช้ AI เพื่อการวิจัย

๑๐ กันยายน ๒๕๖๘

มหาวิทยาลัยเทคโนโลยีสุรนารี

วิโรจน์ อรุณมานะกุล

สถาบันภาษาไทยสิรินธร จุฬาลงกรณ์มหาวิทยาลัย

## ตารางแสดงข้อแตกต่างระหว่างจริยธรรมและกฎหมาย<sup>11</sup>

จริยธรรม	กฎหมาย
1. เป็นเครื่องมือควบคุมพฤติกรรมระดับสูงของมนุษย์	1. เป็นเครื่องมือควบคุมพฤติกรรมระดับต่ำของมนุษย์
2. ไม่มีการลงโทษตามกฎหมายแต่ลงโทษโดยสังคม	2. มีบทลงโทษชัดเจน
3. เป็นการควบคุมพฤติกรรมจากภายใน	3. เป็นการควบคุมพฤติกรรมภายนอก
4. เป็นข้อบังคับของสังคมที่ไม่เป็นลายลักษณ์อักษร	4. เป็นข้อบังคับของรัฐที่เป็นลายลักษณ์อักษร
5. เป็นเรื่องของจิตสำนึกที่ทำเพราะเห็นว่าถูกต้องและ ภูมิใจที่ได้ทำ	5. เป็นบทบัญญัติว่าด้วยการที่ต้องทำหรือต้องละเว้น ไม่สามารถหลีกเลี่ยงได้



RESEARCH  
MISCONDUCT

### จริยธรรม (Ethics)

### จรรยาบรรณ (Code of Conduct)

หลักการและแนวทางในการประพฤติปฏิบัติที่ดีงามตามหลักศีลธรรมที่สังคมยอมรับโดยทั่วไป เป็นเรื่องของความดี-ความชั่ว ถูก-ผิด

ประมวลกฎเกณฑ์ความประพฤติหรือมาตรฐานทางจริยธรรมเฉพาะที่กำหนดขึ้นสำหรับผู้ประกอบการวิชาชีพนั้นๆ

เกิดจากค่านิยม วัฒนธรรม ศาสนา และปรัชญาของสังคม มีลักษณะกว้างและเป็นนามธรรม

เกิดจากการกำหนดเป็นลายลักษณ์อักษรโดยองค์กรวิชาชีพ มีลักษณะเฉพาะเจาะจงและเป็นรูปธรรมมากกว่า

ไม่มีบทลงโทษที่เป็นทางการ แต่อาจมีการลงโทษทางสังคม เช่น การไม่ยอมรับ การประณาม

มีบทลงโทษที่ชัดเจนสำหรับผู้ละเมิด เช่น การตัดเตือน การพักใบอนุญาต หรือการเพิกถอนใบอนุญาตผู้ประกอบการวิชาชีพ

# จรรยาบรรณนักวิจัย สภาวิจัยแห่งชาติ ๒๕๔๑

**นักวิจัย** หมายถึง ผู้ที่ดำเนินการค้นคว้าหาความรู้ อย่างเป็นระบบ เพื่อตอบประเด็นที่สงสัย โดยมีระเบียบวิธีอันเป็นที่ยอมรับ ในแต่ละศาสตร์ที่เกี่ยวข้อง ซึ่งครอบคลุมทั้งแนวคิด มโนทัศน์ และวิธีการที่ใช้ในการรวบรวม และวิเคราะห์ข้อมูล

**จรรยาบรรณ** หมายถึง หลักความประพฤติอันเหมาะสม แสดงถึงคุณธรรมและจริยธรรม ในการประกอบอาชีพ ที่กลุ่มบุคคลแต่ละสาขาวิชาชีพ ประมวลขึ้นไว้เป็นหลัก เพื่อให้สมาชิกในสาขาวิชานั้น ๆ ยึดถือปฏิบัติ เพื่อรักษาชื่อเสียง และส่งเสริมเกียรติคุณ ของสาขาวิชาชีพของตน

**จรรยาบรรณนักวิจัย** หมายถึง หลักเกณฑ์ควรประพฤติปฏิบัติของนักวิจัยทั่วไป เพื่อให้การดำเนินงานวิจัย ตั้งอยู่บนพื้นฐานของจริยธรรม และหลักวิชาการที่เหมาะสม ตลอดจนประกันมาตรฐาน ของการศึกษาค้นคว้า ให้เป็นไป อย่างสมศักดิ์ศรี และเกียรติภูมิของนักวิจัย

## ข้อ 1. นักวิจัยต้องซื่อสัตย์และมีคุณธรรมในทางวิชาการและการจัดการ

นักวิจัยต้องมีความซื่อสัตย์ต่อตนเอง ไม่นำผลงานของผู้อื่นมาเป็นของตน ไม่ลอกเลียนงานของผู้อื่น (Plagiarism) ต้องให้เกียรติ และอ้างถึงบุคคล หรือแหล่งที่มาของข้อมูล ที่นำมาใช้ในงานวิจัย (Citation) ต้องชื่อตรงต่อการแสวงหาทุนวิจัย และมีความเป็นธรรม เกี่ยวกับผลประโยชน์ ที่ได้จากการวิจัย

## ข้อ 2. นักวิจัยต้องตระหนักถึงพันธกรณี ในการทำวิจัย ตามข้อตกลงที่ทำไว้ กับหน่วยงานที่สนับสนุนการวิจัย และต่อหน่วยงาน ที่ตนสังกัด

นักวิจัยต้องปฏิบัติตามพันธกรณี และข้อตกลงการวิจัย ที่ผู้เกี่ยวข้องทุกฝ่ายยอมรับร่วมกัน อุทิศเวลาทำงานวิจัย ให้ได้ผลดีที่สุด และเป็นไปตามกำหนดเวลา มีความรับผิดชอบ ไม่ละทิ้งงาน ระหว่างดำเนินการ

# จรรยาบรรณนักวิจัย สภาวิจัยแห่งชาติ ๒๕๔๑

## ข้อ 3. นักวิจัยต้องมีพื้นฐานความรู้ในสาขาวิชาการที่ทำวิจัย

นักวิจัยต้องมีพื้นฐานความรู้ ในสาขาวิชาการที่ทำวิจัยอย่างเพียงพอ และมีความรู้ความชำนาญ หรือมีประสบการณ์ เกี่ยวเนื่องกับเรื่องที่ทำวิจัย เพื่อนำไปสู่งานวิจัยที่มีคุณภาพ และเพื่อป้องกัน ปัญหาการวิเคราะห์ การตีความ หรือ การสรุป ที่ผิดพลาด อันอาจก่อให้เกิดความเสียหาย ต่องานวิจัย

## ข้อ 4. นักวิจัยต้องมีความรับผิดชอบต่อสิ่งที่ศึกษาวิจัย ไม่ว่าจะจะเป็นสิ่งที่มีชีวิต หรือไม่มีชีวิต

นักวิจัยต้องดำเนินการด้วยความรอบคอบระมัดระวัง และเที่ยงตรงในการทำวิจัย ที่เกี่ยวข้องกับคน สัตว์ พืช ศิลป วัฒนธรรม ทรัพยากร และสิ่งแวดล้อม มีจิตสำนึก และมีปณิธานที่จะอนุรักษ์ ศิลปวัฒนธรรม ทรัพยากร และสิ่งแวดล้อม

## ข้อ 5. นักวิจัยต้องเคารพศักดิ์ศรี และสิทธิของมนุษย์ที่ใช้เป็นตัวอย่างในการวิจัย (จริยธรรมการวิจัยในคน)

นักวิจัยต้องไม่คำนึงถึง ผลประโยชน์ทางวิชาการจนละเลย และขาดความเคารพในศักดิ์ศรี ของเพื่อนมนุษย์ ต้องถือ เป็นภาระหน้าที่ ที่จะอธิบายจุดมุ่งหมายของการวิจัย แก่บุคคลที่เป็นกลุ่มตัวอย่าง โดยไม่ หลอกลวง หรือบีบบังคับ และไม่ละเมิดสิทธิส่วนบุคคล

# จรรยาบรรณนักวิจัย สภาวิจัยแห่งชาติ ๒๕๔๑

ข้อ 6. นักวิจัยต้องมีอิสระทางความคิด โดยปราศจากอคติในทุกขั้นตอนของการทำวิจัย

นักวิจัยต้องมีอิสระทางความคิด ต้องตระหนักว่า อคติส่วนตนหรือความลำเอียง ทางวิชาการ อาจส่งผลให้มีการบิดเบือนข้อมูล (Falsification) และข้อค้นพบทางวิชาการ อันเป็นเหตุให้เกิดผลเสียหายต่องานวิจัย

ข้อ 7. นักวิจัยพึงนำผลงานวิจัยไปใช้ประโยชน์ในทางที่ชอบ

นักวิจัยพึงเผยแพร่ผลงานวิจัย เพื่อประโยชน์ทางวิชาการ และสังคมไม่ขยายผลข้อค้นพบ จนเกินความเป็นจริง (Fabrication) และไม่ใช่ผลงานวิจัย ไปในทางมิชอบ

ข้อ 8. นักวิจัยพึงเคารพความคิดเห็นทางวิชาการของผู้อื่น

นักวิจัยพึงมีใจกว้าง พร้อมที่จะเปิดเผยข้อมูล และขั้นตอนการวิจัย ยอมรับฟังความคิดเห็น และเหตุผลทางวิชาการของผู้อื่น และพร้อมที่จะปรับปรุงแก้ไข งานวิจัยของตนให้ถูกต้อง

ข้อ 9. นักวิจัยพึงมีความรับผิดชอบต่อสังคมทุกระดับ

นักวิจัยพึงมีจิตสำนึก ที่จะอุทิศกำลังสติปัญญาในการทำวิจัย เพื่อความก้าวหน้าทางวิชาการ เพื่อความเจริญ และประโยชน์สุข ของสังคมและมวลมนุษยชาติ

# Research misconduct



Plagiarism. ลักลอกงานวิชาการ ไม่อ้างอิงที่มางาน

Falsification. ปลอมแปลงบิดเบือนข้อมูล

Fabrication สร้างข้อมูลเท็จในงานวิจัย

Authorship ไม่ควรมีชื่อกลับมี ควรมีกลับไม่มี

Duplicate publication ส่งงานเดียวตีพิมพ์หลายที่

Salami publication งานวิจัยเดียวแยกบทความโดยไม่จำเป็น

# การขอตำแหน่งวิชาการ

ประกาศ ก.พ.อ. เรื่อง หลักเกณฑ์และวิธีการพิจารณาแต่งตั้งบุคคล

ให้ดำรงตำแหน่งผู้ช่วยศาสตราจารย์ รองศาสตราจารย์ และศาสตราจารย์ พ.ศ. ๒๕๖๓

หน้า ๒๓ ๕.๑.๔ จริยธรรมและจรรยาบรรณทางวิชาการในการพิจารณากำหนดตำแหน่ง....ต้องคำนึงถึงจริยธรรมและจรรยาบรรณทางวิชาการ ดังนี้

(๑) ต้องมีความซื่อสัตย์ทางวิชาการไม่นำผลงานของผู้อื่นมาเป็นผลงานของตนเองและไม่ลอกเลียนผลงานของผู้อื่น (Plagiarism) ไม่นำผลงานของตนเองในเรื่องเดียวกันไปเผยแพร่ ในวารสารวิชาการมากกว่าหนึ่งฉบับ (Duplicate publication, Salami publication) ไม่คัดลอกข้อความใด ๆ จากผลงานเดิมของตนโดยไม่อ้างอิงผลงานเดิมตามหลักวิชาการ (Self-plagiarism) ทั้งนี้ในลักษณะที่จะทำให้เข้าใจผิดว่าเป็นผลงานใหม่

(๒) ต้องอ้างอิงถึงบุคคลหรือแหล่งที่มาของข้อมูลที่นำมาใช้ในผลงาน (Citation) ทางวิชาการของตนเองเพื่อแสดงหลักฐานของการค้นคว้า

# การขอตำแหน่งวิชาการ

ประกาศ ก.พ.อ. เรื่อง หลักเกณฑ์และวิธีการพิจารณาแต่งตั้งบุคคล

ให้ดำรงตำแหน่งผู้ช่วยศาสตราจารย์ รองศาสตราจารย์ และศาสตราจารย์ พ.ศ. ๒๕๖๓

(๓) ต้องไม่คำนึงถึงผลประโยชน์ทางวิชาการจนละเลยหรือละเมิดสิทธิ ส่วนบุคคลของผู้อื่นหรือสิทธิมนุษยชน (จริยธรรมการวิจัยในคน)

(๔) ผลงานทางวิชาการต้องได้มาจากการศึกษาโดยใช้หลักวิชาการเป็นเกณฑ์ ปราศจากอคติ และเสนอผลงานตามความเป็นจริง ไม่จงใจเบี่ยงเบนผลการศึกษาหรือวิจัย (Falsification) โดยหวังผลประโยชน์ส่วนตัว หรือเพื่อก่อให้เกิดความเสียหายแก่ผู้อื่น และเสนอผลงานตามความเป็นจริง ไม่ขยายข้อค้นพบโดยปราศจากการตรวจสอบยืนยันในทางวิชาการ (Fabrication)

# กรณีตัวอย่าง

**Plagiarism** : 2011 Germany's Defense Minister ลักลอกงานในวิทยานิพนธ์ป.เอกตัวเอง ผลคือถูกถอดถอนปริญญญา ลาออกจากตำแหน่ง

**Self-plagiarism** : 2005 A professor of mechanical and materials engineering at Queen's University ถูกพบว่าลักลอกงานตัวเอง เอาเนื้อหาบทความเดิมใช้ซ้ำในบทความใหม่โดยไม่ได้อ้างอิงให้ถูกต้อง. ถูกตำหนิเป็นทางการจากมหาวิทยาลัย ต่อมา 4 บทความถูกถอดถอนหลังจากการตรวจสอบ

**Duplicate publication, Salami publication** : 2011 Economics professor ที่ Switzerland ตีพิมพ์ 4 บทความ nearly identical เกี่ยวกับกรณีเรือไททานิก, Journal of Economic Perspectives issued a formal censure, วารสาร JEBO เอาชื่อขึ้นบัญชีดำ, สร้างความเสื่อมเสีย

**Falsification** : 1998 a British gastroenterologist (แพทย์โรคทางเดินอาหาร), falsified data เลือกใช้เลือกตัดบางข้อมูล เพื่อ claim ว่ามีความสัมพันธ์ระหว่าง MMR vaccine กับ autism, บทความถูกถอด, ถูกลอน medical license

**Fabrication** : 2004-2005 stem-cell scientist เกาหลีใต้, ถูกตรวจสอบพบว่า the core results were wholly made-up, บทความถูกถอด, ลาออกจากงาน, ถูกดำเนินคดี

# Harvard professor fired following claims she falsified ethics research data

Business school professor's dismissal is first time Harvard has removed a tenured instructor in about 80 years



Ramon Antonio Vargas

Tue 27 May 2025 13:52 BST

A [Harvard](#) professor known for researching honesty before being accused of extensive data fraud has been fired, the first time the Ivy League institution has dismissed a tenured instructor in about 80 years.

[Francesca Gino](#) was initially put on administrative leave by the Harvard Business School (HBS) in 2023 after multiple allegations of falsifying data related to her research, which focused on ethical behavior. On Tuesday, a university spokesperson confirmed that Gino's tenure had been revoked, terminating her employment.



## Harvard fires professor for fabricating research on dishonesty

By [KEITH GRIFFITH](#) and [EMMA RICHTER FOR DAILYMAIL.COM](#)

**PUBLISHED:** 22:38 BST, 26 May 2025 | **UPDATED:** 19:02 BST, 29 May 2025

Harvard University fired a professor after **finding that she fabricated research** used in studies on dishonesty.

Francesca Gino, a star business professor at the Ivy League, was stripped of her title this past week after Harvard administrators informed business faculty of their decision, **GHB reported**.

An investigation into her work **was launched in 2023** after a trio of data bloggers - Uri Simonsohn, Leif Nelson and Joe Simmons - presented what they said was evidence of academic fraud in four studies co-authored by Gino, noting that they also 'believe that many more Gino-authored papers contain fake data.'

# Detection

**Plagiarism ; Self-plagiarism** : reviewer อ่านแล้วเห็นภาษาที่สะดุดไม่ต่อเนื่อง, โปรแกรมอย่าง Turnitin ตรวจสอบข้อความที่เขียนเหมือนกัน (exact copy) หรือถอดจากความเดียวกัน (paraphrase) & ไม่มีการอ้างอิงเหมาะสม

**Duplicate publication, Salami publication** : โปรแกรมอย่าง Turnitin ช่วยหาได้ หรือ reviewer คำนวณว่าเคยอ่านจากไหนมาก่อน

**Falsification ; Fabrication** : ไม่มีงานเก่าที่คล้าย เป็นความไม่ถูกต้องของข้อมูล จะรู้ได้อย่างไร? กรณี falsification ในอังกฤษ มีนักข่าวสงสัย ตรวจสอบหลักฐานการรักษาผู้ป่วยพบว่าไม่ตรงกับที่รายงานในบทความ นักวิจัยไม่ได้เผยแพร่รับทุนจากทนายที่กำลังฟ้องร้องบริษัทยาอยู่

กรณี fabrication ในเกาหลีใต้ ผลที่รายงานสวยเกินไป ไม่มีคนอื่นทำซ้ำได้ นักข่าวเริ่มสงสัย ตามสัมภาษณ์ผู้เกี่ยวข้อง เมื่อเป็นข่าว มหาวิทยาลัยตั้งกรรมการสอบงานวิจัยปกติสามารถ falsified คือสามารถตรวจสอบทำซ้ำได้

# The age of AI

\*\* นอกจาก misconduct ที่กล่าวมา มีอะไรเกิดขึ้นได้จากการนำ AI มาใช้

นักวิจัยนำ AI มาใช้ในงาน

AI เขียนงานแทน, AI ช่วยแก้ภาษา, AI อ่านและเขียนบททวนวรรณกรรม, AI คิดหัวข้อวิจัย, AI วางแผนการวิจัย, AI ทำการทดลอง, AI วิเคราะห์ตีความผล, AI สรุปและอภิปรายผล, AI เขียนรายการอ้างอิง, AI คำนวณสถิติ, AI สร้างกราฟ ตาราง, ....

\*\* เป็น misconduct หรือไม่ อะไรที่ทำได้ อะไรที่ไม่ควรทำ

\*\* เทียบเคียงได้กับ ghost writer และ paper mill. ถือได้ว่าไม่ถูกจริยธรรม

ghost writer : จ้างวาน/ใช้ผู้อื่นเขียนบทความให้ ผู้เขียนไม่มีชื่อเป็น author นั้น

paper mill : จ่ายเงินเพื่อใส่ชื่อบทความที่ผลิตมาเพื่อตีพิมพ์ ปี 2023 มีรายงานพบนักวิชาการไทยหลายมหาวิทยาลัยมีส่วนเกี่ยวข้องกับ paper mill ดูจากผลงานมากผิดปกติ co-author กับต่างชาติคนละสาขา

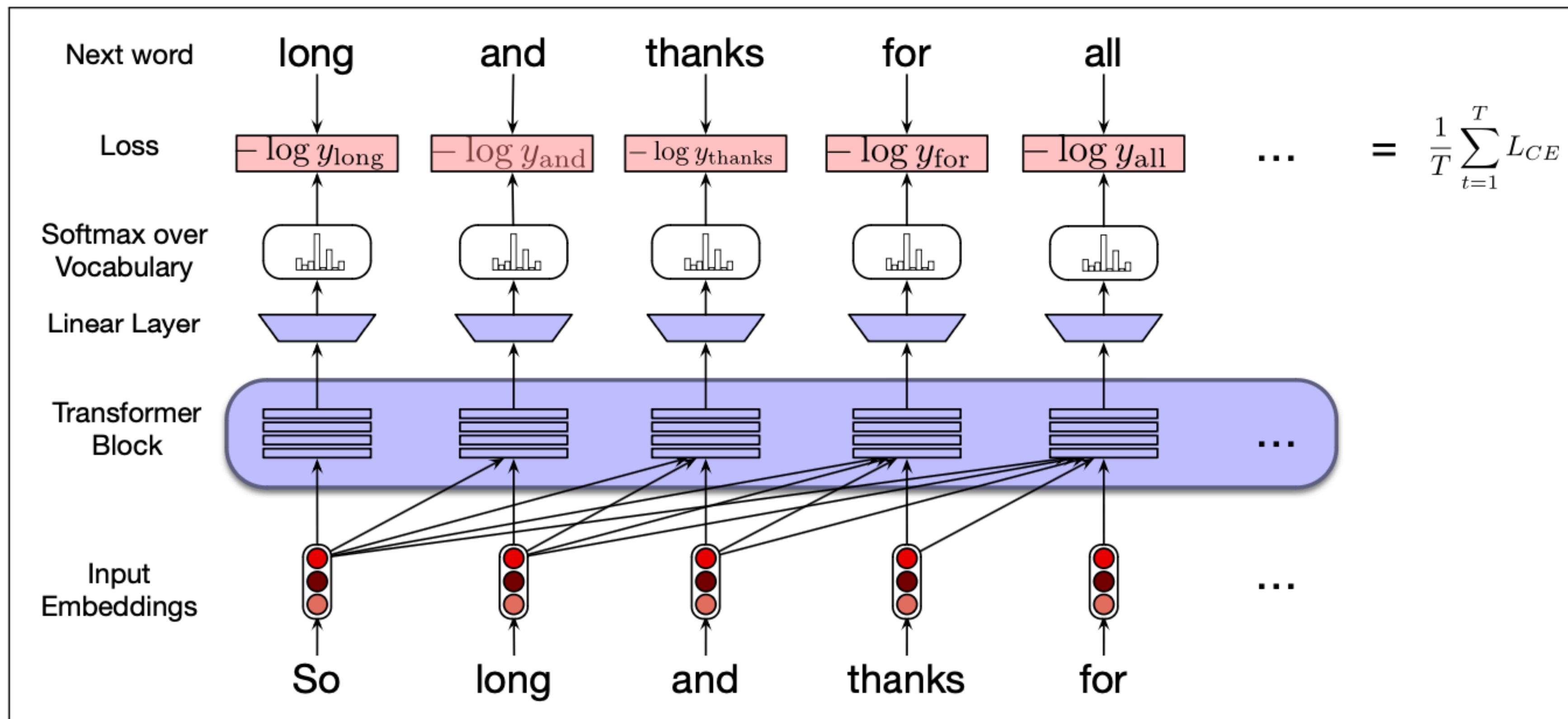
ใช้โปรแกรมอย่าง Turnitin ตรวจได้ใหม่ทำให้ AI เขียนงานให้



ทำไม AI จึงฉลาด

AI ไม่ได้จำเพียงสถิติแต่สร้างความเข้าใจภาษา

Predict next word from previous words



**Figure 10.7** Training a transformer as a language model.

Try not to let your dog ride with his head out of the window.



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย

พยายาม ไม่ ให้ สุนัข ของ คุณ นั่ง หัว ของ เขา ออก จาก  
หน้าต่าง.

Google Translate 2009

พยายามอย่าปล่อยให้สุนัขของคุณนั่งกับหัวของเขาออกไปนอกหน้าต่าง

พยายามอย่าให้สุนัขของคุณขี่ โดยเอาหัวของเขาออกไป  
นอกหน้าต่าง

พยายามอย่าให้สุนัขของคุณขี่ โดยเอาหัวออกไปนอก  
หน้าต่าง

พยายามอย่าให้สุนัขของคุณขี่หัวออกไปนอกหน้าต่าง

พยายามอย่าปล่อยให้สุนัขขี่ โดยเอาหัวออกไปนอก  
หน้าต่าง

2021

ความนี้เป็นภาษาไทย

Try not to let your dog ride with his head out of the window.

ChatGPT 2025

Google Translate 2025

พยายามอย่าให้สุนัขของคุณขึ้นหัวออกนอกหน้าต่างรถ

พยายามอย่าให้สุนัขของคุณนั่ง โดยเอาหัวออกไปนอกหน้าต่าง



# ChatGPT 2025

"Take the pet for a few short rides before your trip so it gets accustomed," says Greg Hammer, a veterinarian and the president of the American Veterinary Medical Association. "You can't take your pet on a four-hour trip if he's never been in the car before."

"พา สัตว์ เลี้ยง สำหรับ ขี่ สั้น น้อย ก่อน การ เดินทาง ของ คุณ จึง ได้ รับ ชื่น" Greg ค้อน, สัตวแพทย์ และ ประธาน ของ สมาคม แพทย์ อเมริกัน Veterinary กล่าว. "คุณ ไม่ สามารถ นำ สัตว์ เลี้ยง ของ คุณ ใน การ เดินทาง สี่ ชั่วโมง ถ้า เขา ไม่ เคย ใน รถ ก่อน.

2009

"ควรพาสัตว์เลี้ยงของคุณไปทดลองนั่งรถระยะสั้น ๆ สักสองสามครั้งก่อนการเดินทาง เพื่อให้มันคุ้นเคย" เกร็ก แฮมเมอร์ สัตวแพทย์และประธานสมาคมสัตวแพทย์อเมริกันกล่าว "คุณไม่สามารถพาสัตว์เลี้ยงของคุณเดินทางสี่ชั่วโมงได้หากมันไม่เคยขึ้นรถมาก่อน"

2021

"พาสัตว์เลี้ยงไปขี่ระยะสั้น ๆ ก่อนออกเดินทางเพื่อให้มันคุ้นเคย" Greg Hammer สัตวแพทย์และประธานสมาคมการแพทย์สัตวแพทย์แห่งสหรัฐอเมริกากล่าว "คุณไม่สามารถพาสัตว์เลี้ยงไปเที่ยวสี่ชั่วโมงได้ถ้าเขาไม่เคยอยู่ในรถมาก่อน"

"สัตว์เลี้ยงสำหรับขี่สั้น ๆ ไม่กี่ครั้งก่อนการเดินทางของคุณดังนั้นจึงได้รับคุ้นเคย" เกร็ก แฮมเมอร์สัตวแพทย์และประธานของสมาคมแพทย์อเมริกันสัตวแพทย์กล่าวว่า "คุณไม่สามารถใช้สัตว์เลี้ยงของคุณในการเดินทางสี่ชั่วโมงถ้าเขาไม่เคยอยู่ในรถมาก่อน"

เกร็กแฮมเมอร์สัตวแพทย์และประธานสมาคมการแพทย์สัตวแพทย์อเมริกันกล่าวว่า" พาสัตว์เลี้ยงไปสักครู่ก่อนเดินทาง "คุณไม่สามารถพาสัตว์เลี้ยงของคุณเดินทางสี่ชั่วโมงถ้าเขาไม่เคยอยู่ในรถมาก่อน"

พาสัตว์เลี้ยงไปสักครู่ก่อนการเดินทางของคุณเพื่อให้มันคุ้นเคย "Greg Hammer สัตวแพทย์และประธานสมาคมการแพทย์สัตวแพทย์อเมริกันกล่าว" คุณไม่สามารถพาสัตว์เลี้ยงของคุณเดินทางสี่ชั่วโมงได้หากเขาไม่เคยเคยอยู่ในรถมาก่อน



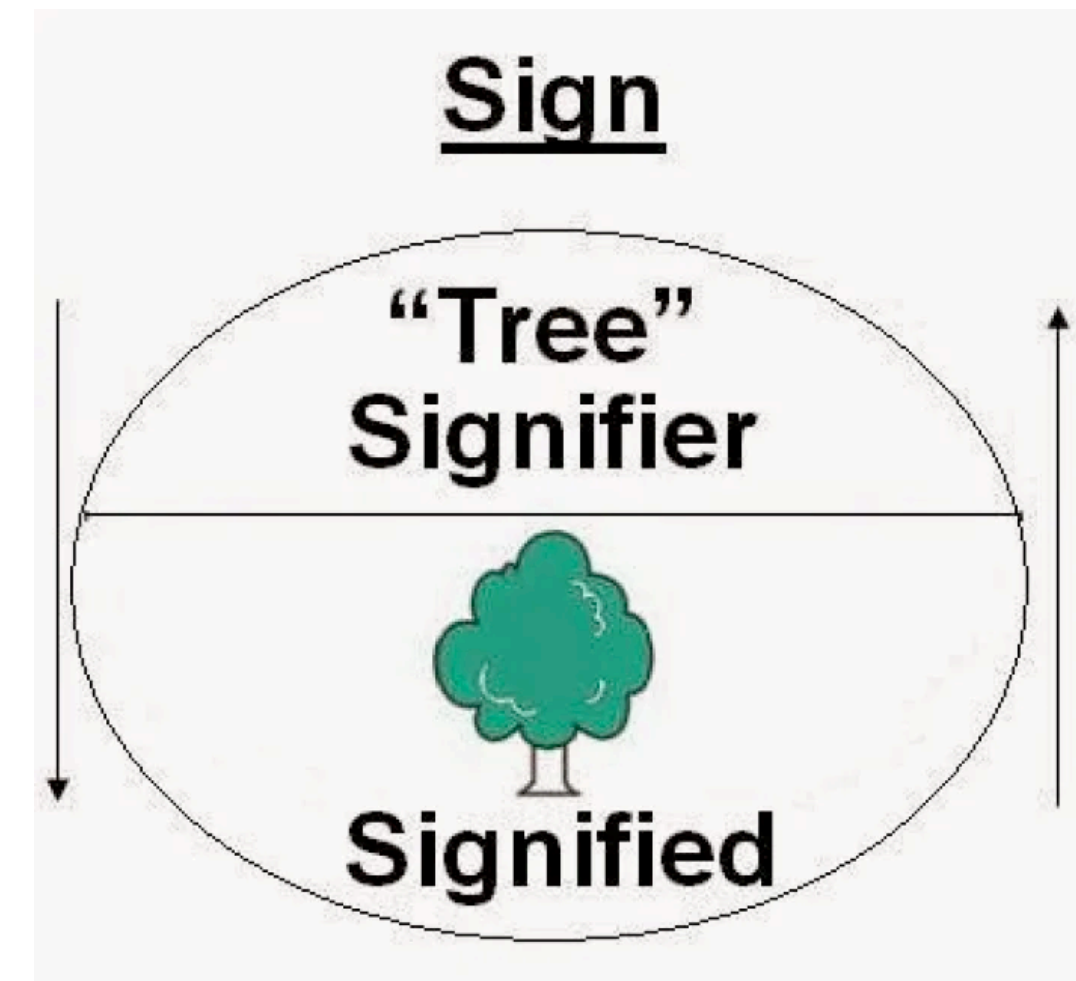
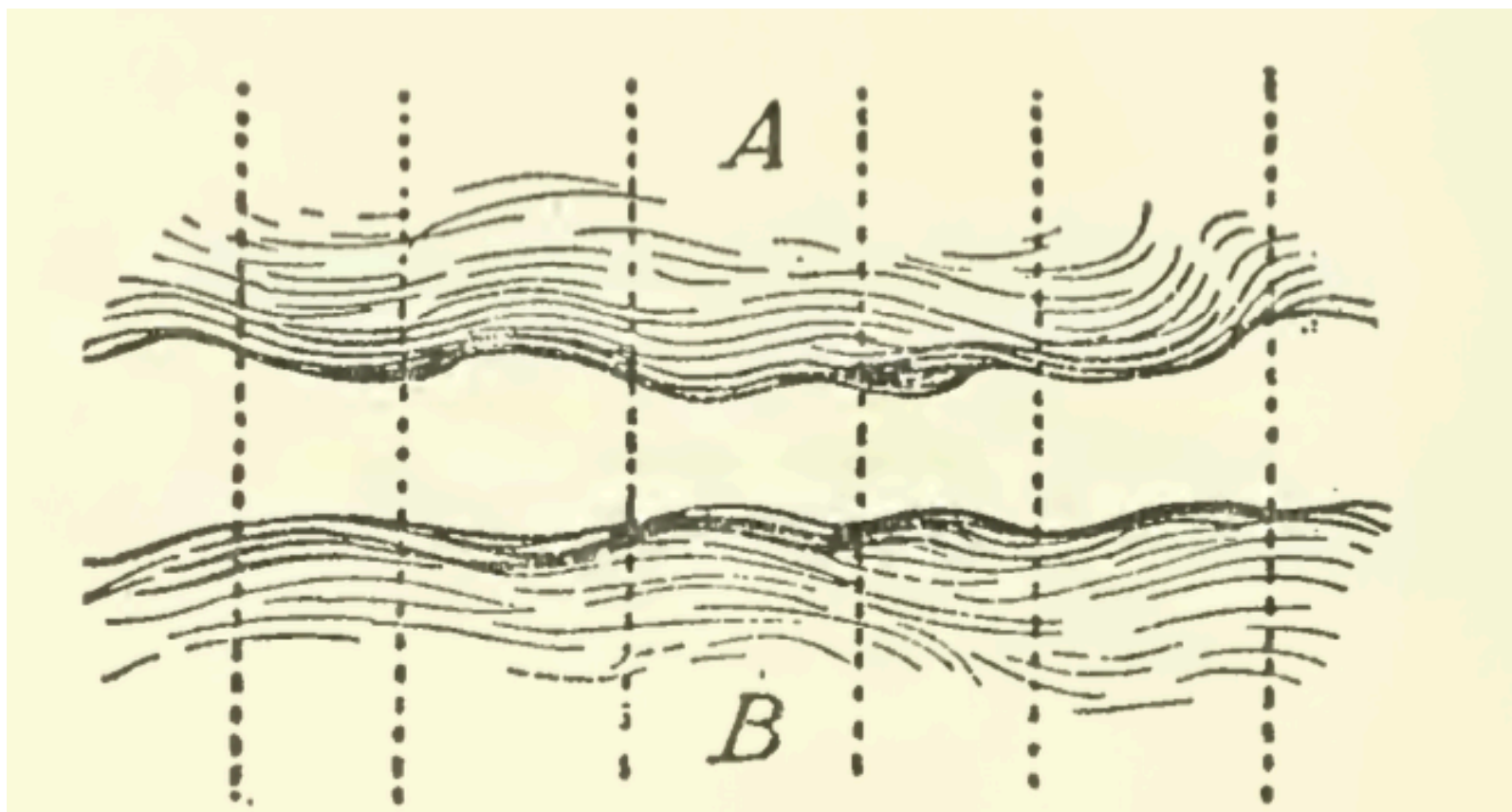
Statistical Machine Translation	Neural Machine Translation
<p>word = “หมา”, “แมว”, ... คำแต่ละคำถูกมองแยกขาดกัน</p>	<p>หมา = (-0.11,0.32,-0.09,0.17,-0.24,0.28,0.41,...,0.21) แมว = (-0.08,0.29,-0.15,0.21,-0.19,0.24,0.37,...,0.18) คำแต่ละคำถูกมองผ่านความสัมพันธ์กับคำอื่น ๆ ทั้งหมด</p>
<p>พื้นฐานมาจากค่าสถิติ สถิติของหน่วยภาษาในระดับต่าง ๆ</p>	<p>พื้นฐานมาจาก embedding embedding = sign ในระดับต่าง ๆ</p>
<p>ข้อมูลภาษา =&gt; corpus สำหรับคำนวณสถิติ</p>	<p>ข้อมูลภาษา =&gt; corpus สำหรับสร้าง system of sign</p>



**“Without language, though is a vague, uncharted nebula. There are no pre-existing ideas, and nothing is distinct before the appearance of language. (Saussure, 1959, p. 112)**

**“.. in language, one can neither divide sound from thought nor thought from sound” (Saussure, 1959, p. 113)**

**"Language is a system of interdependent terms in which the value of each term results solely from the simultaneous presence of the others" (Saussure, 1959, p. 114).**

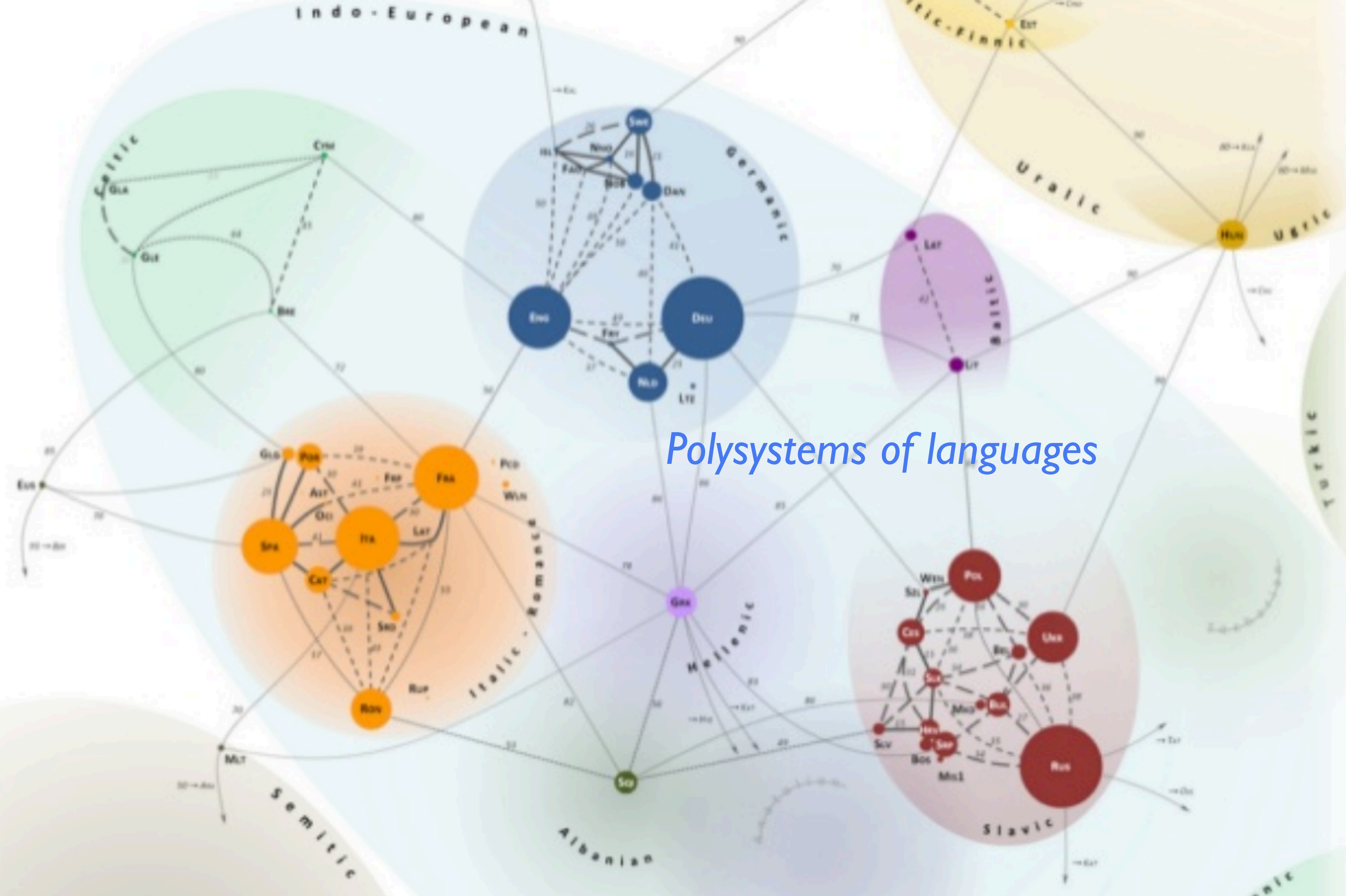




# The Distributional Hypothesis is that **words that occur in the same contexts tend to have similar meanings** (Harris, 1954).

minced garlic 8 slices (about in. thick ) <b>crusty Italian</b>	bread	<b>3 ripe medium</b> tomatoes , sliced 1 lb. fresh mozzarella cheese ,
other half for whole wheat flour . This <b>will give your</b>	bread	<b>a good nutritional</b> boost without sacrificing lightness of
we decreased the diversity of grain varieties ? # <b>is sourdough</b>	bread	<b>a pro-biotic food</b> ? Could some of the allergy problems
home , as if we were Hansel and <b>Gretel scattering stale</b>	bread	<b>across the country</b> . As if by the time I return home
A baker in Woodstock , New York , <b>and author of</b>	Bread	<b>Alone</b> : <b>Bold</b> Fresh Loaves From Your Own Hands (William Morrow
this year , baristas handed out samples of <b>the new banana</b>	bread	<b>along with index</b> cards listing the ingredients . " It 's so
concluded we exchange the kiss of peace . <b>Then someone brings</b>	bread	<b>and a cup</b> of water and wine mixed together to him who
said , we will give you food , <b>two loaves of</b>	bread	<b>and a kilo</b> of margarine , and you will settle in a
to open a jar of sandwich spread . <b>A loaf of</b>	bread	<b>and a pack</b> of bologna lay on the table next to a
# By the time we dressed and broke <b>our fast with</b>	bread	<b>and ale</b> , a great to-do had arisen in the innyard .
Salad , along with specialty brunch <b>cocktails like Banana</b>	Bread	<b>and Apple Pie</b> Mimosa . Ledger is located at 125 Washington St
editor of Stalin 's Letters to Molotov , <b>the author of</b>	Bread	<b>and Authority in</b> Russia 1914-1921 . He has taught political
a ham and cheese sandwich made with <b>crustless , buttered</b>	bread	<b>and broiled or</b> pan-fried in butter . This spring version makes a
instead of iceberg lettuce ; whole grain <b>instead of white</b>	bread	<b>and brown instead</b> of white rice . At this age , the
not like I 'm Robert De Niro . <b>Music is the</b>	bread	<b>and butter for</b> me , for us . And even if one
during wedding season and the holidays , <b>however their</b>	bread	<b>and butter is</b> oftentimes the corporate world . Corporations of
the Hollywood tentpole flicks while the smaller <b>fare were the</b>	bread	<b>and butter of</b> a movie studio . # I could easily see
many business travelers or vacationing <b>families -- the</b>	bread	<b>and butter of</b> the accomodations business -- sneaking around

*The distribution of an element will be understood as the **sum of all its environments**. An environment of an element A is an **existing array of its co-occurents**, i.e. the other elements, each in a particular position, with which A occurs to yield an utterance. (Harris, 1954)*



*Polysystems of languages*

I was angry with my friend:  
I told my wrath, my wrath did end.  
I was angry with my foe:  
I told it not, my wrath did grow.  
  
And I watered it in fears,  
Night & morning with my tears:  
And I sunned it with smiles,  
And with soft deceitful wiles.  
  
And it grew both day and night,  
Till it bore an apple bright,  
And my foe beheld it shine,  
And he knew that it was mine.  
  
And into my garden stole,  
When the night had veild the pole;  
In the morning glad I see:  
My foe outstretched beneath the tree.

จำโกรธเมยคำไว้	จิตสลับระงับพต้น
ศัตรูโกรธผูกพัน	จิตมิเผยแพร่บายเคย
เพาะโกรธด้วยเกรงกลัว	คละระวีมิราเคย
เจ้าค้ำน้ำตาเปรย	หลอกทะเต้าด้วยเทศนัย
โกรธนั้นเติบโตใหญ่หิ่ง	รสีรุ่งมิเหือดไป
ออกผลคนสงขำไพ	ผลิประกาศประจักษ์ตา
ศัตรูเห็นพร่าวพราย	ประจจุจิตมโนมา
รู้ผลเป็นของเจ้า	จิตมิแกคังมิเคคือนคลาย
ย่องเจ้าสวนราตรี	ทูลทีกระทำหมาย
ยามเจ้าเห็นร่างกาย	มฤตินอน ณ ได้พฤกษ์
ยิ้มสุขสบายใจ	ดูจะได้สมจิตนึก
ศัตรูล้มได้พฤกษ์	มฤตินั้นจำปริดี



AW

**You**

สมชายเอาขนมผิงไปซ่อนในไมโครเวฟ สมศรีจะใช้ไมโครเวฟอุ่นอาหารกิน เลยนำขนมผิงออกมาใส่ใน  
กระเป๋าสะพาย สมศรีกินอาหารเวิร์จก็เดินออกจากห้องพักไปที่โรงเรียน ตอนนี้ขนมผิงอยู่ที่ไหน ถ้าสมชาย  
กลับมาที่ห้องเขาจะไปหาขนมผิงที่ไหน

## สร้างความเข้าใจเกี่ยวกับโลก



**ChatGPT**

จากคำบอกเล่า ขนมผิงถูกย้ายจากไมโครเวฟไปใส่ในกระเป๋าสะพายของสมศรี จากนั้นสมศรีก็เดินออกจาก  
ห้องพักไปที่โรงเรียน ดังนั้น ขนมผิงอยู่ที่กระเป๋าสะพายของสมศรี ซึ่งขณะนี้อยู่กับสมศรีที่โรงเรียน

ถ้าสมชายกลับมาที่ห้องและต้องการหาขนมผิง เขาอาจจะหาในไมโครเวฟก่อน แต่ไม่พบขนมผิงที่นั่น เพราะมัน  
ถูกย้ายไปที่กระเป๋าสะพายของสมศรี

# พบการใช้ AI ในงานวิจัยต่าง ๆ

- Analyzing data and images [34, 43, 65, 88, 106, 115, 122, 124, 149, 161].
- Interpreting data and images [13, 14, 21, 41].
- Generating hypotheses [32, 37, 41, 107, 149].
- Modelling complex phenomena [32, 41, 43, 122, 129].
- Designing molecules and materials [15, 37, 43, 205].
- Generating data for use in validation of hypotheses and models [50, 200].
- Searching and reviewing the scientific literature [30, 72].
- Writing and editing scientific papers, grant proposals, consent forms, and institutional review board applications [3, 53, 54, 82, 163].
- Reviewing scientific papers and other research outputs [53, 54, 98, 178, 212].

<sup>66</sup>The ethics of using artificial intelligence in scientific research<sup>99</sup> 2025

# AI Policy

Elsevier

Where authors use generative AI and AI-assisted technologies in the writing process, these technologies should only be used to improve readability and language of the work. Applying the technology should be done with human oversight and control and authors should carefully review and edit the result, because AI can generate authoritative-sounding output that can be incorrect, incomplete or biased. The authors are ultimately responsible and accountable for the contents of the work.

Association of Computational Linguistics

- **Literature search.** Generative text models may be used as search assistants, e.g. to identify relevant literature. However, we expect the authors to read and discuss such references, just like the references identified by a regular search engine or a semantic literature recommendation tool. The usual requirements for citation accuracy and thoroughness of literature reviews apply; beware of the possible biases in suggested citations.

# AI Policy

Taylor & Francis

**Generative AI tools must not be listed as an author**, because such tools are unable to assume responsibility for the submitted content or manage copyright and licensing agreements. Authorship requires taking accountability for content, consenting to publication via a publishing agreement, and giving contractual assurances about the integrity of the work, among other principles. These are uniquely human responsibilities that cannot be undertaken by Generative AI tools.

Taylor & Francis strives for the highest standards of editorial integrity and transparency. Editors' and peer reviewers' use of manuscripts in Generative AI systems may pose a risk to confidentiality, proprietary rights and data, including personally identifiable information. Therefore, **editors and peer reviewers must not upload files, images or information from unpublished manuscripts into Generative AI tools**. Failure to comply with this policy may infringe upon the rightsholder's intellectual property.

**Taylor & Francis currently does not permit the use of Generative AI in the creation and manipulation of images and figures, or original research data for use in our publications**. The term "images and figures" includes pictures, charts, data tables, medical imagery, snippets of images, computer code, and formulas. The term "manipulation" includes augmenting, concealing, moving, removing, or introducing a specific feature within an image or figure. For additional information on Taylor & Francis' image policy for journals, please see Images and figures.

# AI Policy

Taylor & Francis supports the **responsible use** of Generative AI tools that **respect high standards of data security, confidentiality, and copyright protection** in cases such as:

- Idea generation and idea exploration
- Language improvement
- Interactive online search with LLM-enhanced search engines
- Literature classification
- Coding assistance

## Association of Computing Machinery

Generative AI tools and technologies, such as ChatGPT, may not be listed as authors of an ACM published Work. The use of generative AI tools and technologies to create content is permitted but must be fully disclosed in the Work. For example, the authors could include the following statement in the Acknowledgements section of the Work: ChatGPT was utilized to generate sections of this Work, including text, tables, graphs, code, data, citations, etc. If you are uncertain about the need to disclose the use of a particular tool, err on the side of caution, and include a disclosure in the acknowledgements section of the Work.

# ส่วนใหญ่ยอมให้ใช้หรือห้าม AI ทำอะไร

- ปรับปรุงแก้ภาษาให้ถูกต้อง ถูกไวยากรณ์ "improve readability and language of the work"
- ช่วยเริ่มต้นสำรวจหาความคิดที่น่าสนใจ Idea generation/exploration
- สำรวจ รวบรวมและจัดประเภทงานวิจัย แต่ต้องอ่านงานต่าง ๆ เองด้วย
- ส่วนใหญ่มอง AI เป็นเครื่องมือช่วยในการเขียนและปรับปรุงภาษา ไม่ใช่มาแทนที่บทบาทและความรับผิดชอบของผู้เขียน ผู้เขียนยังคงต้องรับผิดชอบต่อความถูกต้อง ความสมบูรณ์ และความเป็นต้นฉบับของผลงานทั้งหมด
- ห้ามให้ AI เขียน review เนื้อหาให้
- ผู้ทบทวนและบรรณาธิการไม่ควรใช้ AI ในการสร้างรายงานการทบทวนหรือจดหมายตัดสินใจ หรือ upload เนื้อหาที่เป็นความลับของต้นฉบับเข้าไปในเครื่องมือ AI

# ส่วนใหญ่ยอมให้ใช้หรือห้าม AI ทำอะไร

- ห้ามใส่ AI เป็น author, co-author  
เนื้อหาและความถูกต้องเป็นความรับผิดชอบของ author
- การใช้ AI ในการสร้างหรือจัดการข้อมูลและรูปภาพมักถูกห้าม  
เว้นแต่เป็นส่วนหนึ่งของวิธีการวิจัยและมีการเปิดเผยอย่างละเอียด
- ห้ามใช้ AI สร้าง original research data
- นักวิจัยต้องมีความรู้ในเรื่องที่ทำเป็นอย่างดี

# บางที่ให้ใช้ AI ได้ บางที่ไม่ให้ใช้

**Association of Computing Machinery Work.** “The use of generative AI tools and technologies to create content is permitted but must be fully disclosed in the Work.” เช่น “ChatGPT was utilized to generate sections of this Work, including text, tables, graphs, code, data, citations, etc.”

**SAGE** : “generating ideas, accelerating research discovery, synthesising, or analysing findings, polishing language, or structuring a submission”.

“Clearly indicate the use of language models in the manuscript, including which model was used and for what purpose.”

---

**Taylor & Francis** "does not permit the use of Generative AI in the creation and manipulation of images and figures, or original research data for use in our publications. The term “images and figures” includes pictures, charts, data tables, medical imagery, snippets of images, computer code, and formulas.

# เหตุผลเบื้องหลัง AI policy

- AI เป็น author ไม่ได้ เพราะ AI ไม่สามารถรับผิดชอบงานได้
- ปัญหา Hallucination, Inconsistency, Recall. ทำให้คำตอบอาจผิด ไม่คงที่ ไม่ครบถ้วน ทำให้ต้องตรวจสอบสิ่งที่ AI สร้าง/เขียนให้
- จะตรวจสอบความถูกต้องได้ นักวิจัยต้องมีความรู้ในเรื่องนั้นเป็นอย่างดี
- เพื่อความโปร่งใส ตรวจสอบได้ จึงต้องแจ้งว่าใช้ AI อย่างไร หรือแบบ prompt ที่ใช้และผลที่ได้
- เพื่อป้องกันความลับงานวิจัย จึงไม่สามารถ upload บทความที่ยังไม่เผยแพร่ให้ AI อ่าน
- คำตอบของ AI มักเขียนแบบกว้าง ๆ ไม่ได้ผิด แต่ไม่เจาะลึกเท่าประสบการณ์นักวิจัยในสาขา จึงควร review ด้วยความรู้ของ reviewer เอง



# ปัญหาของ AI

Hallucination : ให้ข้อมูลไม่จริง

Inconsistency : ตอบคำถามไม่เหมือนเดิม

Recall : หาคำตอบไม่ครบถ้วน

AI ถูก train มาให้ตอบได้สอดคล้องกับคำถามและดูน่าเชื่อถือ

\*Bias : อคติ/สรุปความเข้าใจจากข้อมูลที่ train มา

students by the AI systems. Additionally, it is important to have a clear and transparent governance framework in place, to ensure that the development, deployment and use of AI in education is guided by ethical considerations.

While AI has the potential to revolutionize the way we teach and learn, it is important to consider the ethical implications of this integration, and to take steps to ensure that AI is used in an inclusive and fair manner. This includes ensuring that AI systems are developed and implemented in a transparent and accountable manner. and that they are guided by the principles

Attali, Y., & **Burstein, J. (2006)**. Automated Essay Scoring With e-rater® V.2. *The Journal of Technology, Learning and Assessment*, **4**(3).

Attali, Y., & Cohen, E. (2016). "Automated essay scoring with e-rater v.2." *Journal of Technology, Learning, and Assessment*, 14(3), 1-22.

Attali, Y., & Cohen, G. (2016). "Automated essay scoring with e-rater v.2: Development, scoring, and validation." *Journal of Technology, Learning, and Assessment*, 14(1), 1-35.

Baker, R. S. J. d., Corbett, A. T., & Koedinger, K. R. (2004). "Cognitive Tutors: Technology bringing learning science to the classroom."

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2018). Cognitive Tutor Authoring Tools: A user-friendly authoring environment for creating intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 28(1), 3-17.

Brantley, J. (2017). Designing and leading professional development for AI in education. *Journal of Technology and Teacher Education*, 25(2), 191-208.

Chen, W., & Chen, Y. (2018). Automated essay scoring with e-rater v.2. *Journal of Technology and Teacher Education*, 26(1), 87-110.

students by the AI systems. Additionally, it is important to have a clear and transparent governance framework in place, to ensure that the development, deployment and use of AI in education is guided by ethical considerations.

While AI has the potential to revolutionize the way we teach and learn, it is important to consider the ethical implications of this integration, and to take steps to ensure that AI is used in an inclusive and fair manner. This includes ensuring that AI systems are developed and implemented in a transparent and accountable manner, and that they are guided by the principles of inclusiveness, fairness, and respect for human autonomy.

#### REFERENCES<sup>1</sup>

- Attali, Y., & Cohen, E. (2016). "Automated essay scoring with e-rater v.2." *Journal of Technology, Learning, and Assessment*, 14(3), 1-22.
- Attali, Y., & Cohen, G. (2016). "Automated essay scoring with e-rater v.2: Development, scoring, and validation." *Journal of Technology, Learning, and Assessment*, 14(1), 1-35.
- Baker, R. S. J. d., Corbett, A. T., & Koedinger, K. R. (2004). "Cognitive Tutors: Technology bringing learning science to the classroom."
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2018). *Cognitive Tutor Authoring Tools: A user-friendly authoring environment for creating intelligent tutoring systems*. *International Journal of Artificial Intelligence in Education*, 28(1), 3-17.
- Brantley, J. (2017). Designing and leading professional development for AI in education. *Journal of Technology and Teacher Education*, 25(2), 191-208.
- Chen, W., & Chen, Y. (2018). Automated essay scoring with e-rater v.2. *Journal of Technology and Teacher Education*, 26(1), 87-110.

paper does not exist



**You**

Can you explain J R Firth idea of context of situation?

**information mostly is correct**



**ChatGPT**

J.R. Firth, a prominent British linguist, introduced the concept of "context of situation" as a fundamental aspect of his linguistic theory, which emphasizes the importance of context in understanding language use. Firth's approach to linguistics is part of a broader movement known as British Contextualism, which sought to understand language in relation to its use in specific social and situational contexts. His ideas were instrumental in the development of what later became known as Systemic Functional Linguistics (SFL), primarily through the work of his student M.A.K. Halliday.

The "context of situation" refers to the idea that the meaning of any piece of language—be it a word, phrase, or sentence—cannot be fully understood without considering the environment in which it is used. This environment includes not just the physical setting, but also the social relationships between speakers, the purpose of the interaction, and the cultural norms that influence communication. Firth proposed that the context of situation is composed of three main elements:



1. **Field of Discourse:** This aspect concerns what is happening, the nature of the social action that is taking place. It involves the subject matter or the topic being discussed. Understanding the field of discourse helps to clarify why certain words and structures are chosen in communication.
2. **Tenor of Discourse:** This refers to the participants in the communication process, their roles, and relationships. It includes considerations of social status, power dynamics, degree of formality, and other relational factors between the speakers and listeners. The tenor influences the tone and style of the language used. **but these three ideas are not from Firth**
3. **Mode of Discourse:** This element focuses on the channel of communication (spoken, written, or some other form) and the manner in which the language is organized. It involves how the message is conveyed and the linguistic features that are typical for different modes of discourse, such as the use of certain grammatical structures in spoken vs. written language.

Firth's idea of the context of situation is foundational for understanding that language cannot be seen as a set of isolated units or structures but must be analyzed in terms of its function and use in specific situations. This perspective encourages a holistic view of language, recognizing that words gain their meaning not merely from their intrinsic semantic properties but from their deployment within specific social and situational contexts.



**You**

Are you sure that Firth talked about field, tenor, mode?

## Revision after clarification that the previous answer was incorrect



**ChatGPT**

Upon reflection, I must clarify that the specific categorization of context into field, tenor, and mode is more directly attributed to M.A.K. Halliday, a student of J.R. Firth, and not to Firth himself. This tripartite framework is a key component of Halliday's Systemic Functional Linguistics (SFL), which built upon and extended Firth's initial concepts.

J.R. Firth introduced the concept of context of situation as a crucial element for understanding language in use, emphasizing that the meaning and function of language can only be fully understood within the specific social and situational contexts in which it occurs. He posited that language is inherently tied to the situations in which it is used, highlighting the importance of context in linguistic analysis.

Building on Firth's foundational ideas, M.A.K. Halliday developed the SFL theory, which includes a detailed analysis of how language functions within social contexts. It was Halliday who explicitly articulated the concepts of field (what is happening, the subject matter of the discourse), tenor



## Senior lawyer apologises after filing AI-generated submissions in Victorian murder case

Rishi Nathwani KC has apologised for filing AI-generated information. (ABC News: Kristian Silva)

Aug 15, 2025

### In short:

A senior lawyer has apologised to a Victorian judge for filing submissions in a murder case that included AI-generated errors.

The fake submissions included fabricated quotes from a speech to the state legislature and non-existent case citations purportedly from the Supreme Court.

The AI-generated errors caused a 24-hour delay in resolving a case.

[abc.net.au/news/victoria-lawyer-apologises-aft...](https://www.abc.net.au/news/victoria-lawyer-apologises-aft...)



Share article 

<https://www.abc.net.au/news/2025-08-15/victoria-lawyer-apologises-after-ai-generated-submissions/105661208>

# AI Hallucination Cases

This database tracks legal *decisions*<sup>1</sup> in cases where generative AI produced hallucinated content – typically fake citations, but also other types of arguments. It does not track the (necessarily wider) universe of all fake citations or use of AI in court filings.

While seeking to be exhaustive (327 cases identified so far), it is a work in progress and will expand as new examples emerge. This database has been featured in news media and online posts.<sup>2</sup> (Readers may also be interested in [this project](#) regarding AI use in academic papers.)

If you know of a case that should be included, feel free to [contact me](#).

Click to Download CSV

<https://www.damiencharlotin.com/hallucinations/>

# Feedback and scoring are reliable and consistent?

01

Using the same essay with ChatGPT Plus in 4 chats

คะแนนตรวจงานเดียวกัน แต่ละครึ่งได้ผลต่างกัน

Score : 48, 67, 46, and 55

Deduct scores from errors. 31, 8, 18, and 15

02

03

Evaluate 4 different essays in one chat

คะแนนตรวจงานหลายชิ้นในคราวเดียว ผลสะท้อนความต่างได้

Essay 1 : (55/100 errors -15) Essay 2 :(56/100 errors -13)

Essay 3 :(36/100 errors -23) Essay 4 : (69/100 no error)

E1 ,E2 advanced learner. E3 intermediate learner. E4 professional user

04

## Aug 2025: 2 conversations with the same task

01

First conversation using GPT-5

ตรวจงาน 3 ชั้น rubric เดียวกัน

Score : 68, 56, 77

Deduct scores from errors. 18, 23, 6

02

03

Second conversation using GPT-5.

ตรวจงาน 3 ชั้นเต็ม rubric เต็ม เริ่ม conversation ต่อจากอันก่อน

Score : 50, 27, 65

Deduct scores from errors. 20, 29, 12

04

## Senses of *keep* (1st time)

1. To retain possession of: To have or hold onto something.
2. To continue to have: To not lose or give away.
3. To maintain in a specified condition: To preserve or protect.
4. To store: To place or hold in a location for future use.
5. To stay in a specified state or position: To remain or stay.
6. To continue doing something: To persist or carry on.
7. To fulfill or observe (a promise, commitment, etc.): To adhere to.
8. To prevent from leaving or escaping: To restrain or detain.
9. To manage or conduct (a business, household, etc.): To oversee or maintain.
10. To provide food and lodging for: To support or care for.
11. To continue having a specified temperature: To maintain a certain condition.
12. To record or maintain a record of: To document.
13. To safeguard or protect: To defend or take care of.

## Senses of *keep* (2nd time)

1. To retain possession of: To hold or continue to have something.
2. To maintain in a specified state or condition: To preserve or maintain something in its current state.
6. To save for future use: To store or set aside
8. To remain in a certain condition: To stay in a particular state or place.
3. To continue doing something: To persist in an activity or state.
5. To fulfill or adhere to: To stick to a promise, agreement, or duty.
12. To delay or prevent: To cause someone to stay in a place or situation.
7. To manage or conduct: To carry on or manage a business or activity.
9. To provide food and lodging for: To support or maintain someone or something.
11. To record: To write down or register.
4. To guard or protect: To take care of or safeguard.
10. To stay in one's memory: To remember.
13. To raise or care for animals: To breed or look after animals.

# LLM : เด็กที่ถูกขังในห้องสมุดอ่านแต่หนังสือ



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย





แก้ปัญหาอย่างไร

How to solve the problems



- ไม่ใช่เพียงข้อมูลจากตัวโมเดล LLM
- สามารถค้น internet ให้ได้ข้อมูลปัจจุบัน อ้างอิงได้
- ค้นจากข้อมูลเฉพาะที่เตรียมให้ หรือฐานข้อมูลบทความวิจัย
- ให้ผู้ใช้ upload ข้อมูลเพิ่มเติมได้และใช้ข้อมูลนั้นตอบคำถาม
- คิดมากขึ้น ไม่รีบตอบ ใช้ chain of thoughts , reasoning model
- เลือกโมเดลที่เหมาะสมกับงาน เพราะ LLM แต่ละตัวมีความต่างกัน



ใช้ AI ช่วยทำอะไรได้บ้าง

What can AI help with?



# ใช้ตรวจแก้ภาษา

Proof reading and editing language

make a correction on this text:

Given that AI can generate texts very well like a native speaker. Many articles published on the web nowadays are written by AI. They are not easily distinguished from texts written by human. This might pose a problem for creating a corpus because texts included in a corpus must be authentic and should be produced by a native speaker. A corpus that is compiled from texts produced by a non-native is called a learner corpus. This type of corpus is useful for studying languages produced by learners. Errors and misuses by learners can be identified from the learner corpus. In addition, translation texts even it is produced by a native speaker are considered not naturally occur. They are collected as a part of parallel corpus, in which nature and problems of translation can be studied. Therefore, whether texts generated by AI should be included in a corpus is still an open question. In this paper, we investigate this problem by comparing two comparable corpora. One is a collection of abstracts in an academic journal. Another is a collection of abstracts generated by ChatGPT using the same titles. Whether there are differences in languages of both corpora will be revealed.

## ช่วยตรวจแก้ภาษา



Here is the corrected version of your text:



Given that AI can generate texts very well, similar to a native speaker, many articles published on the web nowadays are written by AI. These are not easily distinguishable from texts written by humans. This might pose a problem for creating a corpus because texts included in a corpus must be authentic and should be produced by a native speaker. A corpus that is compiled from texts produced by a non-native speaker is called a learner corpus. This type of corpus is useful for studying language produced by learners. Errors and

ได้งานภาษาตัวเอง

ได้เนื้อหาตามความคิดเรา  
แต่ภาษาเป็นของ AI

ยอมรับได้?

AW

write an introduction of a paper. texts generated by AI can be found easily on the web. Many corpora collect data from the web. Data can be collected from AI generated texts. If a corpus must be authentic texts written by a native speaker. This might be a problem. need to investigate whether language generated by AI is similar or different from those written by human. This is the aim of this paper. Cite research papers that can be found on the internet using APA style. Write the introduction using style of non-native speakers like Thai.

list all the points, ask ChatGPT to write



### Introduction



In the digital age, artificial intelligence (AI) has become a significant part of life, including the generation of text. Texts generated by AI can be found easily on the web, and many corpora collect data from these sources (Smith, 2019). However, if a corpus must be authentic texts written by a native speaker, this might be a problem.

The question of whether language generated by AI is similar or different from those written by human is an important issue. This issue has significant implications for the construction and use of corpora, the validity of linguistic analyses based on these corpora, and the broader understanding of language and communication in the digital age (Smith & Johnson, 2020).

The aim of this paper is to investigate this critical issue. It seeks to compare AI-generated texts with those written by human authors, focusing on their similarities and differences. Through this investigation, the paper aims to contribute to the ongoing discourse on the role

ได้เนื้อหาตามความคิดเรา  
แต่ภาษาเป็นของ AI

ยอมรับได้?

AW

rewrite this in English like it is written by a non-native  
ปัจจุบัน AI สามารถเขียนบทความต่าง ๆ ได้ดีไม่แพ้มนุษย์ บทความต่าง ๆ จำนวนมากที่เผยแพร่ใน  
อินเทอร์เน็ตเป็นบทความที่เขียนด้วย AI. และคนอ่านจำนวนมากก็อาจจะไม่รู้ ไม่สามารถแยกออกได้ว่า  
เขียนด้วยมนุษย์หรือ AI. เรื่องนี้เป็นปัญหาสำหรับคนที่ต้องการสร้าง corpus ด้วยการรวบรวมข้อมูลตัว  
บทจากแหล่งต่าง ๆ ในอินเทอร์เน็ต เพราะหลักในการสร้าง corpus ดีคือการรวบรวมตัวบทที่เขียนด้วย  
คนและควรเป็นผู้พูดภาษานั้น ข้อมูลที่เขียนโดยคนที่ไม่ใช่เจ้าของภาษาหรือเป็นผู้เรียนภาษานั้นจะเรียก  
ว่า learner corpus. หรือแม้แต่ข้อมูลที่ได้มาจากการแปลก็มองว่าเป็นภาษาที่ไม่เป็นธรรมชาติและอาจ  
ได้รับอิทธิพลจากภาษาต้นฉบับ เพราะหลักพื้นฐานของการรวบรวมข้อมูลงานเขียนจากผู้พูดภาษานั้นก็  
เพื่อให้ได้ภาษาที่สะท้อนลักษณะและการใช้ที่แท้จริงของภาษานั้นเอง  
บทความวิจัยนี้สนใจหาคำตอบว่า ข้อมูลภาษาที่รวบรวมจากตัวบทที่สร้างด้วย AI กับข้อมูลภาษาที่  
เขียนด้วยเจ้าของภาษานั้นมีความเหมือนหรือความต่างกันหรือไม่ โดยจะทดลองเปรียบเทียบ corpus  
ของบทความย่อจากบทความในวารสารวิชาการ เทียบกับ corpus ของบทความย่อที่สร้างด้วย ChatGPT  
จากชื่อบทความเดียวกัน

wrote in Thai, then asked ChatGPT to rewrite in English



Nowadays, AI can write various articles as well as humans. Many articles published on the  
internet are written by AI, and many readers may not know or distinguish whether they are  
written by humans or AI. This is a problem for those who want to create a corpus by  
collecting text data from various sources on the internet. Because the principle of creating a  
good corpus is to collect texts written by people and should be native speakers of that  
language. Data written by non-native speakers or language learners is called a learner  
corpus. Even data obtained from translation is considered unnatural language and may be

# ใช้ AI อ่านและสรุปบทความได้



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย

## Use AI to read and summarize articles



# ให้สรุปรายงานวิจัยและวิจารณ์งาน



ให้สรุปสาระสำคัญของรายงานวิจัยนี้ และเขียนวิจารณ์พร้อมทั้ง  
แนะแนวทางการทำวิจัยต่อไป

Reasoned for 21 seconds >

รายงานวิจัย “เรียนจำต้นแบบเพื่อนำบัณฑิตแก้ไขผู้กระทำผิดทางเพศ ปีที่ 2” มีสาระสำคัญอยู่ที่การพัฒนาและทดสอบ  
โปรแกรมบำบัดแก้ไขสำหรับผู้กระทำผิดเกี่ยวกับเพศในบริบทของประเทศไทย โดยมีเป้าหมายหลักเพื่อสร้าง  
มาตรฐานและแนวทางการบำบัดแก้ไขที่เหมาะสมกับลักษณะเฉพาะของผู้กระทำผิดกลุ่มนี้ เพื่อลดอัตราการกระทำผิด  
ซ้ำในอนาคต จากการดำเนินงานของโครงการ พบว่า

ในช่วงแรกของการวิจัย ผู้วิจัยได้ดำเนินการเลือกพื้นที่และเรียนจำที่มีศักยภาพ (เช่น เรียนจำกลางบางขวางและเรียน  
จำกลางขอนแก่น) รวมทั้งกำหนดกลุ่มเป้าหมายเป็นผู้ต้องขังชายที่กระทำผิดเกี่ยวกับเพศในช่วง 30 เดือนก่อนปล่อย  
ตัว โดยได้มีการพัฒนาเครื่องมือประเมินโปรแกรมบำบัดแก้ไขจากงานวิจัยในปีที่ 1 เพื่อตรวจสอบความน่าเชื่อถือ  
ของเครื่องมือดังกล่าว

ในขั้นตอนต่อมา ทีมวิจัยได้พัฒนาหลักสูตรและจัดการฝึกอบรมผู้ดำเนินการบำบัดแก้ไข โดยใช้หลักสูตรเฉพาะทางที่  
ออกแบบให้สอดคล้องกับมาตรฐานสากล มีการฝึกอบรมทั้งในด้านการประเมินความเสี่ยงและการจัดการโปรแกรม  
บำบัดแก้ไข โดยผู้เข้ารับการอบรมมีคุณสมบัติด้านจิตวิทยาและสุขภาพจิตที่มีประสบการณ์อย่างน้อย 1 ปี

ส่วนการทดสอบโปรแกรมบำบัดแก้ไขนั้นแบ่งเป็น 3 ส่วนหลักคือ แผนงานเตรียมความพร้อม (โดยเน้นการประเมิน  
ความเสี่ยงและความต้องการของผู้ต้องขังก่อนเข้าร่วม โปรแกรม) แผนงานหลัก (ซึ่งแบ่งเป็น โปรแกรมแบบพื้นฐาน  
และแบบเข้มข้นที่เหมาะสมกับระดับความเสี่ยงของผู้ต้องขัง) และแผนงานสรุปปิด โปรแกรม (เพื่อประเมินผลการ  
เปลี่ยนแปลงด้านความคิดและบุคลิกภาพของผู้ต้องขังหลังการบำบัด) ผลการทดสอบพบว่า ผู้ต้องขังที่ได้รับการ  
บำบัดแก้ไขในรูปแบบทั้งสองมีการลดลงของคะแนนที่วัดความคิดสนับสนุนการกระทำผิดและลักษณะบุคลิกภาพ  
ต่อต้านสังคมอย่างมีนัยสำคัญเมื่อเทียบกับกลุ่มควบคุมที่ไม่ได้รับการบำบัด

อย่างไรก็ตาม จากการวิจัยครั้งนี้ยังมีข้อจำกัดที่ควรได้รับการพิจารณาในเชิงวิจารณ์ เช่น ระยะเวลาการบำบัดแก้ไขที่  
ค่อนข้างสั้น (เพียง 18–54 ชั่วโมง) เมื่อเทียบกับ โปรแกรมบำบัด ในมาตรฐานสากลที่มัก ใช้เวลานานกว่า (ประมาณ  
200–300 ชั่วโมง) นอกจากนี้ กลุ่มตัวอย่างที่ใช้ในการทดลองจำกัดอยู่ที่เรียนจำสองแห่ง ซึ่งอาจไม่ครอบคลุมความ  
หลากหลายของผู้กระทำผิดทางเพศในระดับประเทศอย่างเต็มที่ รวมถึงการติดตามผลในระยะยาวยังมีความจำกัดใน



## ถามรายละเอียดในรายงานได้

Reasoned for 33 seconds >

กลุ่มตัวอย่างในงานวิจัยนี้ประกอบด้วยผู้ต้องขังชายที่กระทำผิดเกี่ยวกับเพศ โดยมีคุณลักษณะที่กำหนดไว้จากการหารือร่วมกับเจ้าหน้าที่ของเรือนจำและหน่วยงานที่เกี่ยวข้อง เป้าหมายหลักคือผู้ต้องขังชายที่ยังอยู่ในระยะเวลาก่อนปล่อยตัวประมาณ 30 เดือน ซึ่งถูกพิจารณาว่าอยู่ในช่วงที่มีโอกาสเปลี่ยนแปลงพฤติกรรมได้ โดยในขั้นตอนแรกจากการสำรวจและหารือ พบว่ามีผู้ต้องขังคดีเพศที่มีคุณลักษณะตามเกณฑ์ที่กำหนดรวมทั้งสิ้น 71 ราย ซึ่งครอบคลุมกลุ่มผู้กระทำผิดที่มีลักษณะเฉพาะแตกต่างกันตามประเภทของเหยื่อ ได้แก่ ผู้กระทำผิดต่อเด็ก ผู้กระทำผิดต่อเหยื่อทุกกลุ่มอายุ ผู้กระทำผิดต่อผู้หญิง และผู้ต้องขังในกลุ่มที่อยู่ภายใต้การเฝ้าระวัง (Watchlist) รวมทั้งผู้ต้องขังที่ถูกดำเนินคดีตามพระราชบัญญัติมาตรการป้องกันและการกระทำผิดซ้ำ ในความผิดเกี่ยวกับเพศหรือที่ใช้ความรุนแรง พ.ศ. 2565

จากการวางแผนและการดำเนินงานจริงในเรือนจำที่มีความพร้อม (โดยเฉพาะอย่างยิ่งเรือนจำกลางบางขวางและเรือนจำกลางขอนแก่น) จำนวนผู้ต้องขังที่เข้าร่วมโปรแกรมในขั้นตอนการทดลองใช้ถูกจำกัดอยู่ที่ 68 ราย เนื่องจากข้อจำกัดด้านเวลาและการประสานงานกับหน่วยงานที่เกี่ยวข้อง การคัดเลือกกลุ่มตัวอย่างจึงดำเนินการผ่านขั้นตอนต่อไปนี้

1. การประชุมหารือกับเจ้าหน้าที่และผู้บริหารของเรือนจำที่มีความสนใจเข้าร่วม โครงการเพื่อกำหนดเกณฑ์คัดเลือกผู้ต้องขัง ซึ่งรวมถึงเกณฑ์ด้านระยะเวลาก่อนปล่อย (30 เดือน) และลักษณะของการกระทำผิดทางเพศที่ครอบคลุมหลายกลุ่มเหยื่อ
2. การสำรวจและตรวจสอบรายชื่อผู้ต้องขังที่มีคุณลักษณะตรงตามเกณฑ์ที่กำหนด ในเรือนจำที่เข้าร่วม โครงการ
3. การจำกัดจำนวนผู้เข้าร่วมให้เหมาะสมกับการทดลองใช้โปรแกรมในช่วงเวลาที่มีการดำเนินการจริง เพื่อให้การประเมินผลมีความเป็นระบบและสามารถติดตามการเปลี่ยนแปลงพฤติกรรมได้อย่างชัดเจน



# Gemini อ่านบทความที่ให้และสรุปทำเป็น infographic (web)

A\_systemat...election\_c

PDF

A\_systemat...nstruction

PDF

สรุปเนื้อหาทั้งหมดในบทความที่ upload ให้ แล้วสร้าง infographic

Show thinking ▾

แน่นอนครับ ผมได้สรุปเนื้อหาจากบทความทั้งหมดที่คุณอัปโหลดมา และจะสร้างเป็นอินโฟกราฟิกให้ครับ เนื้อหาหลักๆ จากบทความครอบคลุมประเด็นสำคัญหลายด้านเกี่ยวกับการเรียนรู้และการสอนคำศัพท์ครับ

สรุปประเด็นสำคัญจากบทความ:

## 1. ความสำคัญของคำศัพท์:

- เป็นรากฐานสำคัญของทักษะทางภาษาและความเข้าใจในการอ่าน
- มีผลต่อความสำเร็จทางวิชาการและการสื่อสารโดยรวม
- การรู้ศัพท์ไม่ใช่แค่การจำคำ แต่คือการเข้าใจเครือข่ายความรู้ที่เชื่อมโยงกัน

## 2. ประเภทของคำศัพท์และรายการคำศัพท์ (Word Lists):

Ask Gemini

+

Canvas



อินโฟกราฟิกสรุปการเรียนรู้และสอนคำศัพท์



Code

Preview

Share



# การเรียนรู้และสอนคำศัพท์ (Vocabulary Learning & Teaching)

สรุปแนวคิดสำคัญจากงานวิจัยด้านคำศัพท์

## ⚡ ความสำคัญของคำศัพท์

- เป็น **รากฐาน** ของทักษะทางภาษาและความเข้าใจในการอ่าน
- มีผลต่อ **ความสำเร็จทางวิชาการ** และการสื่อสาร
- ไม่ใช่แค่จำคำ แต่คือการเข้าใจ **เครือข่ายความรู้** ที่เชื่อมโยงกัน

## ≡ ประเภทของคำศัพท์และรายการคำศัพท์

- **คำศัพท์บริการทั่วไป (GSL, NGSL):** คำความถี่สูง ใช้ทั่วไป
- **คำศัพท์เชิงวิชาการ (AWL, AVL):** คำที่พบบ่อยในบทความวิชาการ
- **คำศัพท์เฉพาะทาง:** คำเฉพาะในสาขาต่างๆ
- **คำศัพท์หลัก (Core Vocabulary):** คำพื้นฐานที่จำเป็น
- **ระบบ Tier (Tier 1, 2, 3):** แบ่งคำตามความถี่และความเฉพาะเจาะจง

## 📁 ขนาดคลังคำศัพท์และการพัฒนา

- คลังคำศัพท์ **เติบโต** ตลอดชีวิต
- การเรียนรู้เกิดจาก **การรับข้อมูล** (exposure) และ **การอ่าน**
- การเข้าใจ **โครงสร้างคำ** (morphology) ช่วยขยายคลังศัพท์

## 📏 เกณฑ์การเลือกคำศัพท์

- **ความถี่ (Frequency):** ปรากฏบ่อย
- **ขอบเขตการใช้ (Range):** ใช้ในหลากหลายบริบท

## 🔑 หลักการเรียนรู้และสอนคำศัพท์

- **การพบคำซ้ำ (Repetition):** การเว้นระยะ (spaced repetition)

## 📌 ประเด็นน่าสนใจและข้อควรพิจารณา

- **นิยามของ "คำ":** Lemma, word family



# Introducing GPT-5

ChatGPT now has our smartest, fastest, most useful model yet, with thinking built in — so you get the best answer, every time.

Get a detailed report

+ **Research** x

Sources v

- Web search
- Dropbox
- Gmail
- Google Calendar
- Google Drive
- Box Connect
- Connect more

Ask Gemini

+ **Deep Research**

Canvas

Image

Guided Learning



# perplexity

Ask anything or @mention a Space



**Research**

Recommend

Shopping

Sports

Local

# Deep Research in ChatGPT



ChatGPT 4.5

Share



Activity

28 Sources



The goal is to create basic vocabulary for different levels of students. อยากรู้ว่าในภาษาต่าง ๆ เช่น อังกฤษ จีน ได้รายการศัพท์พื้นฐานนี้มาอย่างไร หลักการในการหาศัพท์พื้นฐานคืออะไร สำหรับภาษาไทยสามารถใช้วิธีการแบบเดียวกันได้หรือไม่ มีผู้ใดเคยศึกษาหัวเรื่องนี้มาก่อน ให้ค้นคว้าจากงานวิจัย วิชาการเป็นหลักก่อน

แนวทางการสร้างรายการคำศัพท์พื้นฐานในหลายภาษา และการประยุกต์ใช้กับภาษาไทย

I'm compiling basic vocabulary list guidelines and sources from multiple languages, while digging into Thai research for unique insights.

Searched for English vocabulary frequency corpus

Could this be helpful? It suggests reviewing corpus-based vocabulary lists for language learners, specifically the KELLY project, and examining systematic literature on academic word list creation based on word selection criteria.

Read [tandfonline.com](https://www.tandfonline.com)

I'm analyzing the selection criteria for creating academic word lists, focusing on systematic literature review processes and comparing them to generic list criteria.

Searched for word list selection criteria

เพื่อให้การค้นคว้ามีประสิทธิภาพมากขึ้น ขอความกรุณาให้คุณช่วยระบุเพิ่มเติมดังนี้:

1. คุณต้องการให้แหล่งข้อมูลครอบคลุมเฉพาะภาษาอังกฤษ จีน และไทย หรือให้ค้นคว้าภาษาหลักอื่น ๆ ด้วย เช่น ญี่ปุ่น เกาหลี ฝรั่งเศส?
2. ต้องการให้เปรียบเทียบแนวทางการสร้างรายการคำศัพท์พื้นฐานในแต่ละภาษาใช้หรือไม่ หรือเน้นเฉพาะภาษาไทยว่าควรพัฒนาอย่างไร?
3. คุณต้องการให้จัดรูปแบบผลลัพธ์เป็นรายงานแบบเต็ม พร้อมการอ้างอิงทางวิชาการด้วยหรือไม่?

โปรดระบุเพิ่มเติมเพื่อให้สามารถเริ่มต้นค้นคว้าได้ตรงจุดมากที่สุดครับ

ดูภาษาอื่นด้วยถ้าหาได้ นำมาเปรียบเทียบกัน แต่สรุปว่าควรใช้กับภาษาไทยอย่างไร เขียนแบบอ้างอิงทางวิชาการด้วย

# แนวทางการสร้างรายการคำศัพท์พื้นฐานในหลายภาษา



## เขียนรายงานสิ่งที่พบ

บทนำ: คำศัพท์พื้นฐาน หมายถึงกลุ่มคำที่มีความสำคัญและพบบ่อยที่สุดในภาษา ซึ่งผู้เรียนภาษาหรือผู้ใช้ภาษาควรรู้จักแต่ต้น เนื่องจากคำเหล่านี้ปรากฏใช้อย่างกว้างขวางในชีวิตประจำวันและครอบคลุมความหมายพื้นฐานส่วนใหญ่ในการสื่อสาร [researchgate.net](https://www.researchgate.net) แนวคิดเรื่อง "รายการคำศัพท์พื้นฐาน" จึงได้รับความสนใจอย่างมากในวงการภาษาศาสตร์ประยุกต์และการสอนภาษา แต่การจะกำหนดว่าคำใดควรอยู่ในรายการดังกล่าว จำเป็นต้องอาศัยหลักเกณฑ์เชิงวิชาการหลายประการ เช่น การวิเคราะห์ฐานข้อมูลคำ (คอร์ปัส) การพิจารณาความถี่การใช้งานของคำ ตลอดจนความครอบคลุมเชิงเนื้อหาและความจำเป็นในการสื่อสาร ฯลฯ ในรายงานนี้ เราจะสำรวจงานวิจัยและบทความวิชาการที่เกี่ยวข้องกับแนวทางการสร้างรายการคำศัพท์พื้นฐานในภาษาต่าง ๆ โดยเน้นที่ภาษาอังกฤษ ภาษาจีน ภาษาญี่ปุ่น และภาษาอื่น ๆ ที่มีข้อมูล จากนั้นจะวิเคราะห์หลักเกณฑ์ที่ใช้ในการคัดเลือกคำ พร้อมทั้งเปรียบเทียบให้เห็นแนวคิดร่วมและแนวคิดต่างของแต่ละภาษา สุดท้ายจะสรุปว่าเราสามารถประยุกต์แนวทางเหล่านี้กับการสร้างรายการคำศัพท์พื้นฐานภาษาไทยได้อย่างไร และมีงานวิจัยใดบ้างที่เคยศึกษาเรื่องนี้สำหรับภาษาไทย

## หลักเกณฑ์และแนวคิดในการคัดเลือกคำศัพท์พื้นฐาน

งานวิจัยล่าสุดที่ทบทวนวิธีสร้างรายการคำศัพท์ด้วยคอร์ปัสชี้ว่า โดยทั่วไปนักวิจัยใช้เกณฑ์ในการคัดเลือกคำหลายด้านร่วมกัน (มากถึง 9 ด้าน) [researchgate.net](https://www.researchgate.net) อย่างไรก็ตาม สำหรับรายการคำศัพท์พื้นฐานทั่วไป (core/basic vocabulary) มักให้ความสำคัญกับเกณฑ์หลัก ๆ ดังต่อไปนี้:

1. ความถี่ในการใช้งาน (Frequency): คำที่ปรากฏบ่อยในภาษามักได้รับการคัดเลือกเข้าสู่รายการคำศัพท์พื้นฐานก่อนเกณฑ์อื่น ๆ เพราะถือว่าเป็นคำที่ผู้ใช้ภาษาพบเจอและต้องการบ่อยที่สุด [researchgate.net](https://www.researchgate.net) ตัวอย่างเช่น *General Service List (GSL)* ของ West (1953) ซึ่งเป็นรายก ↓ าศัพท์อังกฤษพื้นฐานที่มีชื่อเสียง ได้มาจากการเลือกคำ

2.5 Pro (experimental) ▾

- Gemini**
- 2.0 Flash**  
Get everyday help
- 2.5 Flash (experimental)** New  
Uses advanced reasoning
- 2.5 Pro (experimental)**  
Best for complex tasks
- Deep Research** ✓  
Get in-depth research reports
- Personalization (experimental)** New  
Help based on your Search history
- Gemini Advanced** Upgrade

# Deep Research in Gemini

Hello, *Wirote*

Research AI campaigns for annual planning

Research sensor trends for my robotics essay

Research floor heating for a bathroom remodel

Ask Gemini



ศึกษาศัพท์เกี่ยวกับการทำอาหารในภาษาอังกฤษ เทียบกับศัพท์ในภาษาไทย ดูว่าในแต่ละภาษามีคำอะไรบ้าง คำเหล่านั้นเทียบระหว่างภาษาได้โดยตรง หรือมีคำใดที่ในอีกภาษาไม่มี ถ้าไม่มีคำที่...

## แสดงขั้นตอนการค้นหาคำตอบ

Show thinking ^

Here's a thinking process to arrive at the comparison of English and Thai cooking terms:

- 1. Understand the User's Request:** The user wants to study English cooking terms and compare them to Thai cooking terms. They want to see direct equivalents, terms that don't have direct equivalents, and the closest matching terms in the other language. They also require sources to be cited.
- 2. Initial Brainstorming & Keyword Identification:**
  - Cooking terms (English and Thai)
  - Comparison (English vs. Thai)
  - Direct equivalents
  - Non-equivalent terms
  - Closest matches
  - Culinary vocabulary
  - Cooking techniques (English and Thai)
  - Kitchen actions/verbs (English and Thai)
  - Sources: Dictionaries, culinary websites, language learning resources.
- 3. Formulate Search Queries:** Based on the brainstorming, create search queries for

- 3. Formulate Search Queries:** Based on the brainstorming, create search queries for information.

- English cooking verbs list
- Thai cooking verbs list
- cooking terms English to Thai translation
- Thai culinary techniques translation
- kitchen vocabulary English Thai comparison
- ศัพท์ทำอาหาร ภาษาอังกฤษ ภาษาไทย
- คำกริยาทำอาหาร ภาษาอังกฤษ ภาษาไทย
- เทคนิคทำอาหาร ภาษาไทย ภาษาอังกฤษ
- คำศัพท์ทำอาหารที่ไม่มีในภาษาไทย (or English)
- ความแตกต่างศัพท์ทำอาหาร อังกฤษ ไทย

- 4. Execute Search Queries (Simulated):** Use the search tool with the formulated queries. Analyze the search results. Look for:
  - Lists of cooking verbs and nouns in both languages.
  - English-Thai culinary dictionaries or glossaries.
  - Articles or blog posts comparing cooking techniques across cultures.
  - Language learning resources focusing on food and cooking.
- 5. Synthesize Information from Search Results:**
  - **Common Techniques/Verbs (Likely Equivalents):** Identify terms that appear frequently and seem to have direct translations. Examples: boil (ต้ม), fry (ทอด), steam

# เขียนรายงานสิ่งที่พบ



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย

## การศึกษาเปรียบเทียบศัพท์เกี่ยวกับการทำอาหารในภาษาอังกฤษและภาษาไทย

การทำอาหารเป็นกิจกรรมสากล แต่คำศัพท์และเทคนิคที่ใช้ในการอธิบายนั้นมีความแตกต่างกันไปตามวัฒนธรรมและภาษา การศึกษาเปรียบเทียบศัพท์เกี่ยวกับการทำอาหารระหว่างภาษาอังกฤษและภาษาไทยเผยให้เห็นทั้งความเหมือนและความต่างที่น่าสนใจ ทั้งในส่วนของคำที่มีความหมายตรงกัน คำที่ใกล้เคียงกันแต่มิใช่คำเดียวกัน และคำที่ไม่มีคำตรงๆ ในอีกภาษาหนึ่ง

จากการค้นคว้าจากแหล่งข้อมูลต่างๆ เช่น พจนานุกรมออนไลน์ บทความสอนทำอาหาร และเว็บไซต์เกี่ยวกับการเรียนภาษา สามารถสรุปการเปรียบเทียบคำศัพท์ได้ดังนี้

### แหล่งข้อมูล:

- พจนานุกรมออนไลน์ (เช่น Thai-English dictionaries, English-Thai dictionaries)
- เว็บไซต์สอนทำอาหาร (เช่น Allrecipes, Serious Eats, FoodTravel.tv, Krua.co)
- บทความหรือบล็อกเกี่ยวกับการเรียนภาษาและวัฒนธรรมอาหาร

## 1. คำศัพท์ที่มีความหมายตรงกันโดยตรง (Direct Equivalents)

มีคำศัพท์พื้นฐานเกี่ยวกับการทำอาหารหลายคำที่มีความหมายและขอบเขตการใช้งานที่เทียบเคียงกันได้โดยตรงในทั้งสองภาษา ส่วนใหญ่มักเป็นคำกริยาที่อธิบายการกระทำหรือเทคนิคการปรุงอาหารเบื้องต้น และคำนามที่ใช้อธิบายอุปกรณ์พื้นฐาน

ภาษาอังกฤษ (English)	ภาษาไทย (Thai)	คำอธิบาย
Boil	ต้ม	การทำให้สุกในน้ำเดือด
Fry	ทอด	การทำให้สุกในน้ำมันร้อน
Steam	นึ่ง	การทำให้สุกด้วยไอน้ำ
Grill	ย่าง	การทำให้สุกบนตะแกรงหรือไฟโดยตรง

## 2. คำศัพท์ที่มีความหมายใกล้เคียงกัน แต่มีนัยหรือขอบเขตต่างกัน (Close but Not Exact Equivalents)

บางคำศัพท์ในภาษาหนึ่งอาจมีคำที่ใกล้เคียงในอีกภาษาหนึ่ง แต่เมื่อพิจารณาในรายละเอียดของเทคนิค วิธีการ หรือผลลัพธ์ อาจมีความแตกต่างกันเล็กน้อย

ภาษาอังกฤษ (English)	ภาษาไทย (Thai)	คำอธิบายความแตกต่างและคำที่ใกล้เคียงที่สุดในอีกภาษา
Bake	อบ	'Bake' ในภาษาอังกฤษมักหมายถึงการทำสุกในเตาอบแห้ง ส่วนใหญ่ใช้กับขนมปัง เค้ก พาย หรืออาหารที่ต้องการผิวกรอบ   'อบ' ในภาษาไทยมีความหมายกว้างกว่า อาจหมายถึงการทำให้สุกในความร้อนแห้งๆ ซึ่งรวมถึงการอบในเตาอบ หรือแม้แต่การอบบนเตาถ่านแบบโบราณ (เช่น ไก่ย่างอบโอ่ง) คำว่า 'อบ' จึงเป็นคำที่ใกล้เคียงที่สุด แต่ขอบเขตการใช้งานอาจไม่เหมือนกันทั้งหมด
Roast	อบ / ย่าง (อบ)	'Roast' มักใช้กับการทำสุกเนื้อสัตว์หรือผักชิ้นใหญ่ในเตาอบด้วยความร้อนแห้ง โดยมักจะไม่ได้ปิดฝา เพื่อให้ผิวกรอบ   ในภาษาไทยอาจใช้คำว่า 'อบ' เช่น อบไก่ อบหมู หรือบางครั้งอาจเรียกว่า 'ย่าง (อบ)' เพื่อสื่อถึงการทำสุกในลักษณะนี้ แต่ไม่มีคำกริยาเดี่ยวๆ ที่ตรงกับ 'Roast' ในทุกบริบท 'อบ' ถือเป็นคำที่ใกล้เคียงที่สุด
Sauté	ผัด (ใช้น้ำมันน้อย)	'Sauté' เป็นเทคนิคการผัดอย่างรวดเร็วในน้ำมันหรือไขมันเพียงเล็กน้อย โดยใช้อุณหภูมิค่อนข้างสูง   ใกล้เคียงกับคำว่า 'ผัด' ในภาษาไทย แต่ 'Sauté' มักจะเน้นความรวดเร็วและการใช้น้ำมันน้อยกว่าการ 'ผัด' ทั่วไปของไทยซึ่งอาจมีน้ำมันมากกว่า คำว่า 'ผัด' เป็นคำที่ใกล้เคียงที่สุด
Blanch	ลวก	'Blanch' คือการนำอาหารไปต้มในน้ำเดือดระยะเวลาสั้นๆ แล้วนำไปแช่น้ำเย็นทันที เพื่อหยุดการสุก รักษาสี และลดความขม   'ลวก' ในภาษาไทยคือการนำไปต้มในน้ำเดือดสั้นๆ แต่ไม่ได้มีขั้นตอนการแช่น้ำเย็นต่อเสมอไป มักใช้เพื่อให้สุกแค่บางส่วนหรือให้นิ่มขึ้น 'ลวก' เป็นคำที่ใกล้เคียงที่สุด แต่ไม่มีนัยของการหยุดการสุกทันที
Braise	ตุ๋น / เคี่ยว	'Braise' คือการนำเนื้อสัตว์หรือผักไปจี่/ทอดให้ผิวกรอบก่อน แล้วนำไปตุ๋นต่อในของเหลวในหม้อที่มีฝาปิด ใช้ความร้อนต่ำเป็นเวลานาน   ใกล้เคียงกับคำว่า 'ตุ๋น' หรือ 'เคี่ยว' ในภาษาไทย ซึ่งหมายถึงการทำสุกในของเหลวด้วยความร้อนต่ำเป็นเวลานานเพื่อให้เปื่อย แต่ 'Braise' มีขั้นตอนการจี่ก่อนซึ่งอาจไม่มีใน 'ตุ๋น' หรือ 'เคี่ยว' เสมอไป คำที่ใกล้เคียงที่สุดคือ 'ตุ๋น' หรืออาจอธิบายว่า 'ตุ๋นโดยทอด/จี่ก่อน'
Reduce	เคี่ยว (ให้น้ำงวด)	'Reduce' คือการต้มของเหลวด้วยความร้อนสูงเพื่อให้ปริมาณน้ำลดลงและรสชาติเข้มข้นขึ้น   ใกล้เคียงกับคำว่า 'เคี่ยว' ในภาษาไทย ซึ่งหมายถึงการต้มของเหลวด้วยความร้อนสูงเพื่อให้ปริมาณน้ำลดลงและรสชาติเข้มข้นขึ้น



### 3. คำศัพท์ที่ไม่มีคำตรงๆ ในอีกภาษาหนึ่ง (No Direct Equivalent) และคำที่ใกล้เคียงที่สุด

มีคำศัพท์บางคำที่สะท้อนถึงเทคนิคการทำอาหาร หรือส่วนผสมบางอย่างที่เป็นเอกลักษณ์ของวัฒนธรรมนั้นๆ ทำให้ไม่มีคำกริยาหรือคำนามเดี่ยวๆ ที่มีความหมายตรงกันในอีกภาษาหนึ่ง อาจต้องใช้วลีหรืออธิบายแทน

ภาษาไทย (Thai)	ภาษาอังกฤษ (English)	คำอธิบายและคำที่ใกล้เคียงที่สุดในภาษาอังกฤษ
โบลก	Pound (with mortar and pestle) / Grind	การทำให้ส่วนผสมแหลกหรือเป็นเนื้อเดียวกันโดยใช้ครกและสาก เป็นเทคนิคสำคัญในการทำน้ำพริกหรือเครื่องแกง   ไม่มีคำกริยาเดี่ยวๆ ในภาษาอังกฤษที่ตรงกับ 'โบลก' ซึ่งหมายถึงการกระทำโดยใช้ครกและสากโดยเฉพาะ คำที่ใกล้เคียงที่สุดคือ 'Pound' หรือ 'Grind' โดยต้องระบุว่าเป็น 'with mortar and pestle'
ซอย	Finely slice / Thinly slice / Mince	การหั่นเป็นเส้นเล็กๆ หรือเป็นชิ้นบางๆ ละเอียด มักใช้กับวัตถุดิบที่มีกลิ่นหอม เช่น ตะไคร้หอมแดง พริก   ใกล้เคียงกับ 'Finely slice' หรือ 'Thinly slice' หรือบางครั้งอาจเทียบได้กับ 'Mince' ในบริบทของการสับละเอียดสำหรับวัตถุดิบกลิ่นหอม แต่ 'ซอย' มีนัยถึงการหั่นที่ต้องใช้ทักษะและได้ชิ้นที่เท่ากันสวยงาม คำที่ใกล้เคียงที่สุดคือ 'Finely slice' หรือ 'Thinly slice'
เจียว	Render (fat) / Fry crispy	มีสองความหมายหลัก: 1. การเคี่ยวไขมันให้ละลายออกมา (เช่น เจียวหมู) 2. การทอดหอมแดงหรือกระเทียมให้เหลืองกรอบ   ไม่มีคำกริยาเดี่ยวๆ ที่ครอบคลุมทั้งสองความหมายนี้ในภาษาอังกฤษ   ความหมายแรก: 'Render (fat)'   ความหมายที่สอง: 'Fry (shallots/garlic) until crispy'
เคี่ยว	Simmer / Reduce / Slow boil	การต้มด้วยไฟอ่อนๆ เป็นเวลานานเพื่อให้เปื่อย นุ่ม หรือให้น้ำงวดลงจนข้น   ครอบคลุมความหมายของ 'Simmer' (ตุ๋นไฟอ่อน) และ 'Reduce' (เคี่ยวให้น้ำงวด) แต่ 'เคี่ยว' ในภาษาไทยสามารถใช้ได้ทั้งสองบริบท ไม่มีคำกริยาเดี่ยวๆ ที่ตรงทั้งหมด 'Simmer' และ 'Reduce' เป็นคำที่ใกล้เคียงที่สุดตามบริบท
คั้น	Squeeze / Extract (liquid)	การบีบหรือบิดเพื่อสกัดของเหลวออกจากวัตถุดิบ เช่น คั้นกะทิ คั้นน้ำมะนาว   ใกล้เคียงกับ 'Squeeze' หรือ 'Extract (liquid)' ไม่มีคำกริยาเดี่ยวๆ ที่เฉพาะเจาะจงเหมือน 'คั้น' ซึ่งมักจะทำด้วยมือ 'Squeeze' เป็นคำที่ใกล้เคียงที่สุด
ตำ	Pound / Make a paste (in mortar)	คล้ายกับ 'โบลก' แต่ 'ตำ' มักใช้กับการนำส่วนผสมหลายๆ อย่างมาใส่ครกแล้วตำผสมกันให้แหลกและเข้ากันเป็นเนื้อเดียว มักใช้ในการทำส้มตำ น้ำพริกบางชนิด   ใกล้เคียงกับ 'Pound' หรือ 'Make a paste' โดยต้องระบุว่าเป็น 'in a mortar' แตกต่างจาก 'โบลก'

ภาษาอังกฤษ (English)	ภาษาไทย (Thai)	คำอธิบายและคำที่ใกล้เคียงที่สุดในภาษาไทย
Deglaze	ล้างกันกระทะ (เพื่อละลายเศษอาหาร) / บุดเศษอาหาร	เทคนิคการเติมของเหลวเล็กน้อยลงในกระทะที่เพิ่งใช้ทอดหรือจี่ เพื่อละลายเศษอาหารที่ติดกันกระทะ ซึ่งมีรสชาติเข้มข้น นำไปใช้ทำซอส   ไม่มีคำกริยาเดี่ยวๆ ในภาษาไทย อธิบายด้วยวลี 'ล้างกันกระทะ' หรือ 'บุดเศษอาหารที่ติดกันกระทะ'
Zest	บุดผิว (ผลไม้)	ส่วนของเปลือกผลไม้รสเปรี้ยว (เช่น มะนาว ส้ม) เฉพาะส่วนที่มีสี ไม่รวมส่วนสีขาว มักใช้เพื่อเพิ่มกลิ่นหอม   คำนาม 'Zest' ไม่มีคำนามเดี่ยวๆ ในภาษาไทย ส่วนคำกริยา 'to zest' คือการบุดเอา Zest ออกมา ซึ่งอธิบายด้วยวลี 'บุดผิว (ผลไม้)'
Render	เจียว (ไขมัน)	การให้ความร้อนกับไขมันสัตว์เพื่อให้ละลายและแยกกากออก   ตรงกับความหมายแรกของคำว่า 'เจียว' ในภาษาไทย
Sous Vide	ซูวี (ทับศัพท์) / สูญญากาศ	เทคนิคการทำอาหารโดยการนำอาหารที่ซีลในถุงสุญญากาศไปแช่ในน้ำที่ควบคุมอุณหภูมิคงที่เป็นเวลานาน   ไม่มีคำศัพท์ภาษาไทยที่ใช้กันแพร่หลาย มักทับศัพท์ว่า 'ซูวี' หรือบางครั้งอธิบายว่า 'ทำอาหารแบบสุญญากาศควบคุมอุณหภูมิ'

Export to Sheets

#### ข้อสังเกตเพิ่มเติม:

- **คำนามเฉพาะทางวัฒนธรรม:** วัตถุดิบบางอย่างหรืออุปกรณ์เฉพาะทางวัฒนธรรมมักไม่มีคำแปลตรงๆ เช่น "ครกและสาก" (Mortar and Pestle) ซึ่งเป็นอุปกรณ์ทำครัวพื้นฐานของไทย แต่ในวัฒนธรรมตะวันตกอาจไม่ใช้แพร่หลายเท่า หรือชื่อเครื่องเทศ สมุนไพรบางชนิด
- **ความซับซ้อนของเทคนิค:** บางเทคนิคในภาษาไทย เช่น การ "ผัดฉ่า" ซึ่งเป็นการผัดรสจัดจ้านใส่สมุนไพรหลายชนิด จะไม่มีคำกริยาเดี่ยวๆ ในภาษาอังกฤษเพื่ออธิบาย จำเป็นต้องอธิบายถึงส่วนผสมและลักษณะเฉพาะของอาหารจานนั้น
- **คำคุณศัพท์ที่อธิบายลักษณะอาหาร:** คำคุณศัพท์ที่ใช้อธิบายลักษณะของอาหารหลังการปรุงอาจมีความละเอียดต่างกัน เช่น คำอธิบายความเปื่อยของเนื้อ หรือความกรอบของอาหารทอด อาจมีคำที่หลากหลายน่าสนใจ

อยากใหช่วยหางานวิจัยที่ใช้ vector semantic เทียบระหว่างภาษาเพื่อหาค่าที่มีความหมายเหมือนหรือใกล้กันผ่านการ map ระหว่าง vector space models นี้ งานที่ผ่านมาใช้แนวทางอะไรอย่างไรบ้าง มี research gap อะไรที่สามารถทำต่อได้

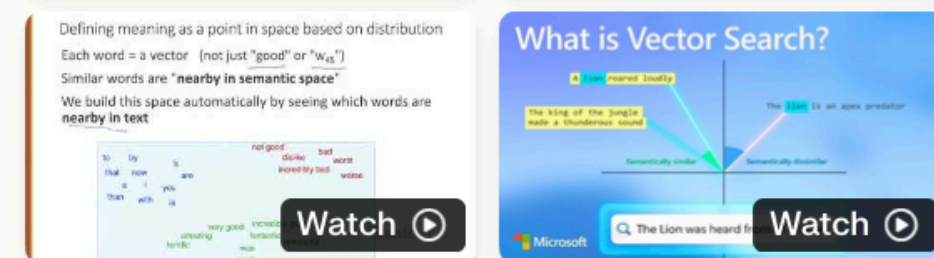
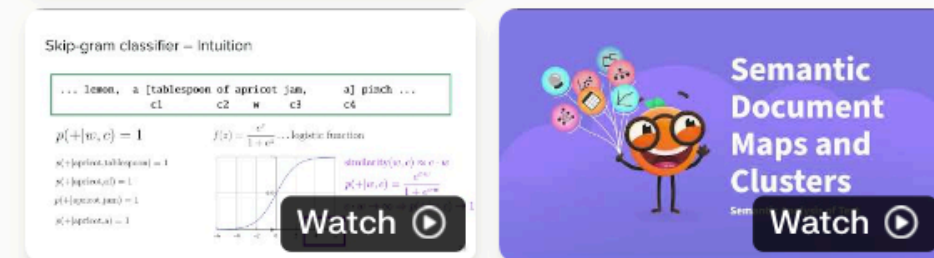
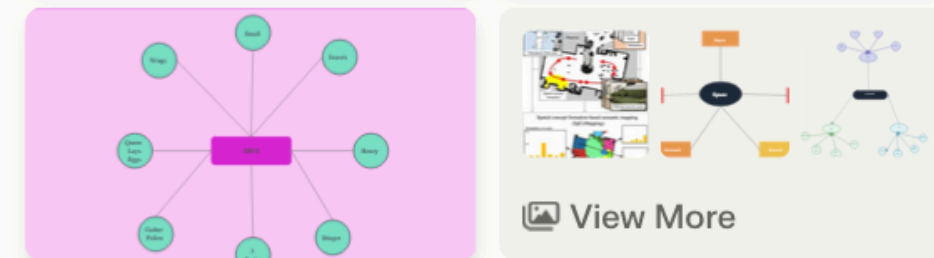
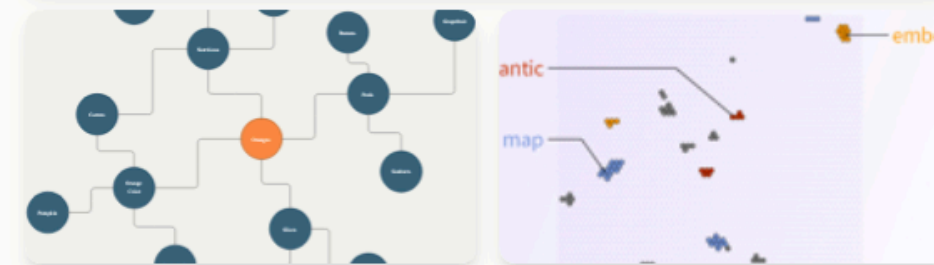
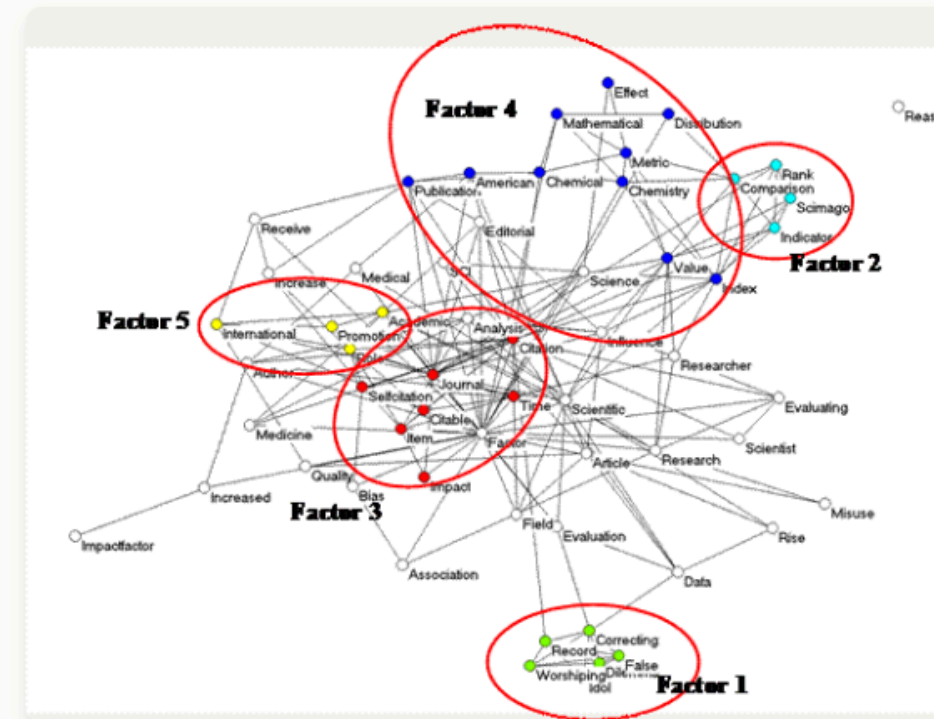
# ใช้ Deep Research ใน Perplexity

## Deep Research

42 sources

- I found several research papers discussing vector semantic mapping techniques and their...
- I found several relevant research papers discussing various approaches to cross-lingual word...
- I found several relevant studies on cross-lingual vector semantic mapping, including...
- I found sufficient information on various approaches to cross-lingual semantic mapping,...
- Let me analyze the query first. The query is in Thai language, and it's asking me to find research...
- Let me translate the key parts of the query:...
- Specifically looking at approaches that map between vector space mode...
- Looking at the search results, I have several relevant papers and articles that discuss cross-...
- From the search results:...
- The second result is about "Unsupervised Cross-lingual Transfer of Word Embedding Spaces"...
- The third result is about Vector Space Mapping Model (VSM), providing a general explanation...
- The fifth result is about "Context-Aware Cross-Lingual Mapping" which proposes alternatives t...
- The sixth result is about "Exploring the Vector Space Model for Finding Verb Synonyms in...
- Now, I'll organize the information from these sources to create a comprehensive report on...
- Let's start with understanding the approaches that have been used in past research based on...
- From result 5, we learn about two approaches for context-aware cross-lingual mapping:...
- Learning a transformation between sentence embeddings rather than individual word...
- From result 4, we see an approach using Latent Dirichlet Allocation (LDA) to analyze linked...
- From result 6, although it focuses on Portuguese only, we see the use of Vector Space Model...
- Now, let me identify potential research gaps based on these approaches:...
- The contextual approaches mentioned in 5 are relatively recent, suggesting that explorations...

## กระบวนการที่คิดและทำ



PRO Generate Image





# งานวิจัยการใช้ Vector Semantics เพื่อเปรียบเทียบคำระหว่างภาษาผ่านการ Map Vector Space Models

การศึกษาเกี่ยวกับการประยุกต์ใช้ vector semantics เพื่อเปรียบเทียบคำระหว่างภาษาโดยการสร้างการเชื่อมโยงระหว่าง vector space models เป็นหัวข้อที่น่าสนใจและได้รับความสนใจอย่างต่อเนื่องในวงการประมวลผลภาษาธรรมชาติ โดยงานวิจัยในปัจจุบันได้พัฒนาวิธีการที่หลากหลายในการหาคำที่มีความหมายเหมือนกันหรือใกล้เคียงกันระหว่างภาษา ทั้งแบบที่ต้องใช้ข้อมูลที่มีการจับคู่ระหว่างภาษาและแบบไม่ต้องใช้ข้อมูลดังกล่าว บทความนี้จะวิเคราะห์แนวทางต่างๆ ที่ใช้ใน vector semantics ข้ามภาษา รวมถึงช่องว่างในการวิจัยที่ยังสามารถพัฒนาต่อไปได้

## พื้นฐานของ Vector Semantics และการแปลงพื้นที่เวกเตอร์ข้ามภาษา

Vector semantics เป็นแนวคิดในภาษาศาสตร์คอมพิวเตอร์ที่นำเสนอความหมายของคำ วลี หรือเอกสารทั้งฉบับในรูปแบบของเวกเตอร์ในพื้นที่หลายมิติ แนวคิดนี้ตั้งอยู่บนสมมติฐานการกระจายตัว (distributional hypothesis) ที่ว่าคำที่มีความหมายคล้ายกันมักปรากฏในบริบทคล้ายกัน <sup>1</sup>

ในบริบทของการเปรียบเทียบระหว่างภาษา Vector Space Mapping Model (VSM) ถูกนำมาใช้เพื่อสร้างการเชื่อมโยงระหว่างพื้นที่เวกเตอร์ของแต่ละภาษา โดย VSM เป็นโมเดลที่ทำการแมปข้อมูลไปยังพื้นที่เวกเตอร์ <sup>3</sup> และการทำ cross-lingual mapping มีเป้าหมายเพื่อสร้างความสัมพันธ์ทางความหมายระหว่างคำในภาษาที่แตกต่างกัน โดยการเรียนรู้ฟังก์ชันการแปลงระหว่างพื้นที่เวกเตอร์ของแต่ละภาษา <sup>2</sup>

การแปลงพื้นที่เวกเตอร์ข้ามภาษาสามารถทำได้หลายวิธี แต่วิธีที่นิยมที่สุดวิธีหนึ่งคือการสร้างเมทริกซ์การแปลงเชิงเส้น (linear transformation matrix) ที่จับคู่คำจากพจนานุกรมสองภาษา (bilingual dictionary) จากพื้นที่เวกเตอร์ต้นทางไปยังพื้นที่เวกเตอร์เป้าหมาย <sup>5</sup> วิธีนี้มีความยืดหยุ่นและขยายขนาดได้ดี เนื่องจากสามารถสร้างการฝังคำหลายภาษาโดยการแมปพื้นที่เวกเตอร์ของหลายภาษาไปยังภาษาเป้าหมายร่วมกัน ซึ่งมักเป็นภาษาอังกฤษ <sup>5</sup>

## แนวทางที่มีการควบคุม (Supervised Approaches)

แนวทางดั้งเดิมในการสร้างการแมปข้ามภาษาคือการใช้ข้อมูลที่มีการกำกับดูแล (supervised data) เช่น พจนานุกรมสองภาษาหรือคลังข้อมูลคู่ขนาน (parallel corpora) เพื่อเรียนรู้การแปลงระหว่างพื้นที่คำเวกเตอร์ หนึ่งในวิธีที่มีประสิทธิภาพคือการสร้างเมทริกซ์การแปลงเชิงเส้นบนรายการในพจนานุกรมคำเริ่มต้นสองภาษา

คำตอบที่ให้



อยากให้ช่วยหางานวิจัยที่ใช้ vector semantic เทียบระหว่างภาษาเพื่อหาคำที่มีความหมายเหมือนหรือใกล้เคียงผ่านการ map ระหว่าง vector space models นี้ งานที่ผ่านมาใช้แนวทางอะไรอย่างไรบ้าง มี research gap อะไรที่สามารถทำต่อได้

## DeepSearch

0s • 31 sources

- ✓ ค้นหาข้อมูลเพิ่มเติม
- ✓ วิเคราะห์ช่องว่างวิจัย
- ✓ Completed

### ค้นหาข้อมูลเพิ่มเติม

- ฉันกำลังมองหาวิจัยที่เกี่ยวข้องกับการใช้ semantics vector ในการเปรียบเทียบภาษา เพื่อหาคำที่มีความหมายคล้ายกันผ่านโมเดล vector space.
- ฉันคิดว่านี่เกี่ยวข้องกับ cross-language หรือ multilingual word embeddings ด้วย มีงานวิจัยมากมายใน NLP เกี่ยวกับการจัดเรียง word embeddings จากภาษาต่างๆ เพื่อหาความสัมพันธ์ระหว่างคำ.
- ฉันกำลังเริ่มต้นด้วยการค้นหาเว็บสำหรับคำทั่วไป เพื่อดูภาพรวมก่อน.

### Searching for "cross-language word embeddings"

- Mc A quick report on cross-lingual word embeddings | by kh... medium.com
- Cross-Lingual Word Embeddings | Computational Lingu... direct.mit.edu
- A survey of cross-lingual word embedding models www.ruder.io
- 🔄 GitHub - facebookresearch/MUSE: A library for Multilingu... github.com
- Mc TDS Archive – Medium medium.com

กระบวนการที่คิดและทำ





## คำตอบโดยสรุป

- การวิจัยที่ใช้ vector semantic ในการเปรียบเทียบภาษาเพื่อหาคำที่มีความหมายเหมือนหรือใกล้เคียงกันผ่านการ map ระหว่าง vector space models นั้น มีแนวทางหลักดังนี้: การเรียนรู้การแปลงเชิงเส้น, การฝึก โมเดลร่วมกันด้วยข้อมูลคู่ขนาน, การใช้ข้อมูลที่เปรียบเทียบได้, และวิธีที่ไม่ต้องใช้ข้อมูลคู่ขนาน
- งานวิจัยก่อนหน้านี้พบว่า โมเดลที่ใช้ข้อมูลคู่ขนานมักให้ผลดีกว่าในงาน semantic แต่โมเดลที่ใช้ข้อมูลน้อยอาจเหมาะสำหรับงาน syntactic
- ช่องว่างการวิจัยที่สามารถทำได้ รวมถึงการพัฒนา โมเดลสำหรับภาษาที่มีทรัพยากรน้อย, การปรับปรุงการจัดตำแหน่งสำหรับภาษาที่แตกต่างกันมาก, และการรวมข้อมูล subword หรือ multimodal

### แนวทางที่ใช้ในอดีต

งานวิจัยที่ผ่านมาใช้หลายวิธีในการสร้าง vector space models สำหรับเปรียบเทียบภาษา เช่น:

- การเรียนรู้การแปลงเชิงเส้น: ใช้ lexicon เล็กๆ เพื่อเรียนรู้การแปลงระหว่าง embedding ของภาษาต่างๆ เช่น Mikolov et al. (2013) ใช้กับภาษาอังกฤษและสเปน
- การฝึก โมเดลร่วมกัน: ใช้ข้อมูลคู่ขนาน เช่น ประโยคที่แปลกัน เพื่อฝึก โมเดลที่เรียนรู้ embedding สำหรับหลายภาษา เช่น Klementiev et al. (2012)
- การใช้ข้อมูลที่เปรียบเทียบได้: สำหรับภาษาที่ไม่มีข้อมูลคู่ขนานมาก ใช้ข้อมูลเช่นบทความ Wikipedia ที่เกี่ยวข้องกัน เช่น Vulić and Moens (2016)
- วิธีที่ไม่ต้องใช้ข้อมูลคู่ขนาน: ใช้โครงสร้างของ embedding หรือการฝึกแบบ adversarial เช่น Artetxe et al. (2018)

### ช่องว่างการวิจัย

คำตอบที่ให้



ใช้เป็นผู้ช่วยวิจัย

Custom GPT

NotebookLM



# คุยผ่าน chat เพื่อ config พฤติกรรม GPT ที่ต้องการให้

Create

Configure

Hi! I'll help you build a new GPT. You can say something like, "make a creative who helps generate visuals for new products" or "make a software engineer who helps format my code."

What would you like to make?

อยากได้ GPT ที่เป็นผู้ช่วยวิจัย ให้ช่วยอ่านงานที่ upload ให้ และตอบคำถามเกี่ยวกับเรื่องคำศัพท์พื้นฐานในภาษา ให้อาศัย ข้อมูลจากบทความที่ให้เป็นหลักในการตอบคำถามต่าง ๆ พร้อมทั้งอ้างอิงบทความที่ใช้ด้วย APA style. การตอบให้ใช้ภาษา เป็นทางการ อธิบายความให้ชัดเจน เขียนเหมือนเป็นเนื้อหาบทความทางวิชาการ

เรียบร้อยแล้วค่ะ ต่อไป ฉันขอเสนอชื่อ GPT สำหรับบทบาทนี้ว่า ผู้ช่วยวิจัยคำศัพท์ — เหมาะสมไหม ค่ะ หรือต้องการใช้ชื่ออื่น?



ผู้ช่วยวิจัยเรื่องคำศัพท์พื้นฐาน





















ผู้ช่วยวิจัยเรื่องคำศัพท์พื้นฐาน  
• Draft

Create

Configure

Conversations with your GPT can potentially reveal part or all of the files uploaded.

-  A new academic word li... PDF
-  A Comparison of Difficu... PDF
-  Early vocabulary develo... PDF
-  Defining Core Vocabula... PDF
-  How vocabulary is learn... PDF
-  Is\_There\_a\_Core\_Gener... PDF
-  Profiling vocabulary for ... PDF
-  Predicting L2 reading pr... PDF
-  THE PRINCIPLES OF TE... PDF
-  A\_systematic\_literature... PDF
-  A systematic review of ... PDF
-  How Large a Vocabulary... PDF
-  Development of L2 disci... PDF
-  How many words do chi... PDF
-  SELECTING AND CREAT... PDF
-  Research articleOpen a... PDF
-  The rediscovery of voca... PDF
-  WHERE WOULD GE... PDF

ให้อธิบายว่าวิธีการหาคำศัพท์พื้นฐานที่ผ่านมา ทำอย่างไรบ้าง ถ้าจะใช้กับภาษาไทยควรใช้วิธีไหนดี

แนวทางการหาคำศัพท์พื้นฐานที่ผ่านมาในงานวิจัยเกี่ยวกับภาษาอังกฤษสามารถแบ่งออกได้เป็นสองแนวหลัก ได้แก่ (1) วิธีการเชิงปริมาณผ่านการวิเคราะห์คลังข้อมูลภาษา (corpus-based methods) และ (2) วิธีการเชิงอภิปัญญาหรือจากความเห็นของผู้เชี่ยวชาญ (judgment-based methods) โดยทั้งสองแนวทางนี้มีบทบาทสำคัญในการสร้างรายการคำศัพท์พื้นฐานหรือ "core vocabulary" ซึ่งมีเป้าหมายเพื่อระบุคำที่มีความถี่สูง ใช้งานกว้าง และมีประโยชน์ต่อผู้เรียนในระดับต้น

### วิธีการเชิงปริมาณผ่านคลังข้อมูลภาษา

วิธีนี้มักใช้เกณฑ์หลัก 3 ประการในการคัดเลือกคำ ได้แก่ ความถี่ (frequency), การกระจายตัว (dispersion) และความหลากหลายของประเภทข้อความ (range) เช่น รายการ *New General Service List* (NGSL) ใช้เกณฑ์ที่เรียกว่า *average reduced frequency* (ARF) ซึ่งพิจารณาทั้งความถี่และการกระจายตัวในคลังข้อมูลขนาดใหญ่ถึง 12 พันล้านคำ นอกจากนี้ยังมีการวิเคราะห์จากชุดข้อมูลเฉพาะ เช่น คลังข้อมูลของหนังสือเรียน, เอกสารวิชาการ, หรือคลังข้อมูลภาษาพูด เพื่อให้คำที่เลือกมีความเหมาะสมกับบริบทการใช้งานของผู้เรียน

### วิธีการเชิงอภิปัญญา

อีกแนวทางหนึ่งคือการเลือกคำศัพท์ตามความเห็นของผู้เชี่ยวชาญหรือใช้พจนานุกรมเฉพาะทาง เช่น การใช้พจนานุกรม

ถ้าต้องการศึกษาเรื่องนี้ ควรจะเริ่มอ่านงานไหนก่อน ช่วยแนะนำ 3 บทความ  
แบบบอกเหตุผลว่าทำไมจึงน่าสนใจอ่าน



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย

หากต้องการเริ่มศึกษาประเด็นเรื่อง "การคัดเลือกและพัฒนาคำศัพท์พื้นฐาน" อย่างเป็นระบบ โดยเฉพาะสำหรับการ  
ประยุกต์ใช้ในบริบทภาษาไทย ควรเริ่มจากการอ่านบทความ 3 ชิ้นสำคัญต่อไปนี้ ซึ่งมีความโดดเด่นในด้านกรอบแนวคิด  
วิธีวิจัย และแนวทางเชิงปฏิบัติ:

## 1. Coxhead, A. (2000). *A New Academic Word List*. *TESOL Quarterly*, 34(2), 213–238.

เหตุผลที่ควรอ่าน:

บทความนี้เป็นงานวิจัยคลาสสิกที่เสนอ "Academic Word List (AWL)" ซึ่งกลายเป็นจุดตั้งต้นสำคัญของการศึกษาเกี่ยวกับคำศัพท์ทางวิชาการในวงการภาษาศาสตร์ประยุกต์ งานนี้อธิบายวิธีการสร้างรายการคำจากคลังข้อมูลด้วยเกณฑ์ความถี่และการกระจาย ทำให้ผู้อ่านเข้าใจแนวทางการใช้ข้อมูลจริงเพื่อระบุคำที่มีความสำคัญในภาษาทางวิชาการ

จุดเด่น:

- ใช้คลังข้อมูล 3.5 ล้านคำ
- กำหนดเกณฑ์การตัดคำที่เป็นระบบและนำไปประยุกต์ ↓ ทั่วทั้งภาษาอื่น

# NotebookLM อ่านบทความ/หนังสือที่ให้และตอบคำถามต่างๆ

สร้าง podcast คุณกัน, VDO presentation ในหัวข้อเรื่องจากสิ่งที่ upload ให้

นโยบาย AI สำหรับผู้เขียนโดย JHUP

Share Settings

### Sources

+ Add Discover

Select all sources

- ACL.txt
- Academic Integrity: Author-Related a...
- AnEvaluationofScholarlyPublisherPoli...
- Analysis of Artificial Intelligence Ethic...
- BJR AI oup.txt
- BMJ.txt
- Ethical Use of Artificial Intelligence fo...
- Generative AI for Authors\_final\_0.pdf
- Guidance for generative AI in educati...

### Chat

Refresh

## นโยบาย AI สำหรับผู้เขียนโดย JHUP

38 sources

แหล่งข้อมูลเหล่านี้ส่วนใหญ่กล่าวถึง หลักจรรยาบรรณในการเผยแพร่งานวิจัย และ การใช้อุตสาหกรรมปัญญาประดิษฐ์ (AI) อย่างมีจริยธรรม ในงานวิจัยและการเขียนเชิงวิชาการ โดยเน้นที่ประเด็นต่าง ๆ เช่น การเป็นผู้เขียนร่วม โดยเฉพาะอย่างยิ่งกรณีที่เกี่ยวข้องกับเครื่องมือ AI และ ความรับผิดชอบของผู้เขียน สำหรับเนื้อหาที่สร้างขึ้น ประเด็นสำคัญอื่น ๆ ได้แก่ ความถูกต้องและอคติ ของเนื้อหาที่สร้างโดย AI, ความจำเป็นในการแสดงที่มา สำหรับการใช้อ AI และความสำคัญของการรักษา ความลับ และ สิทธิในทรัพย์สินทางปัญญา นอกจากนี้ยังมีข้อกำหนดด้าน จริยธรรมในการวิจัย ทัวไปสำหรับมนุษย์และสัตว์ เช่น การยินยอมโดยมีข้อมูล ครบถ้วนและ การเปิดเผยข้อมูล ที่เพียงพอ รวมถึง ความโปร่งใส ในกระบวนการวิจัยและ การแก้ไขข้อผิดพลาด ในสิ่งพิมพ์ที่ตีพิมพ์

Save to note

Add note Jump to bottom Mind Map

Start typing... 38 sources

มาตรฐานทางจริยธรรมและนโยบายการเผยแพร่ที่เกี่ยวข้องกับการใช้ AI ในงานวิชาการคืออะไร?

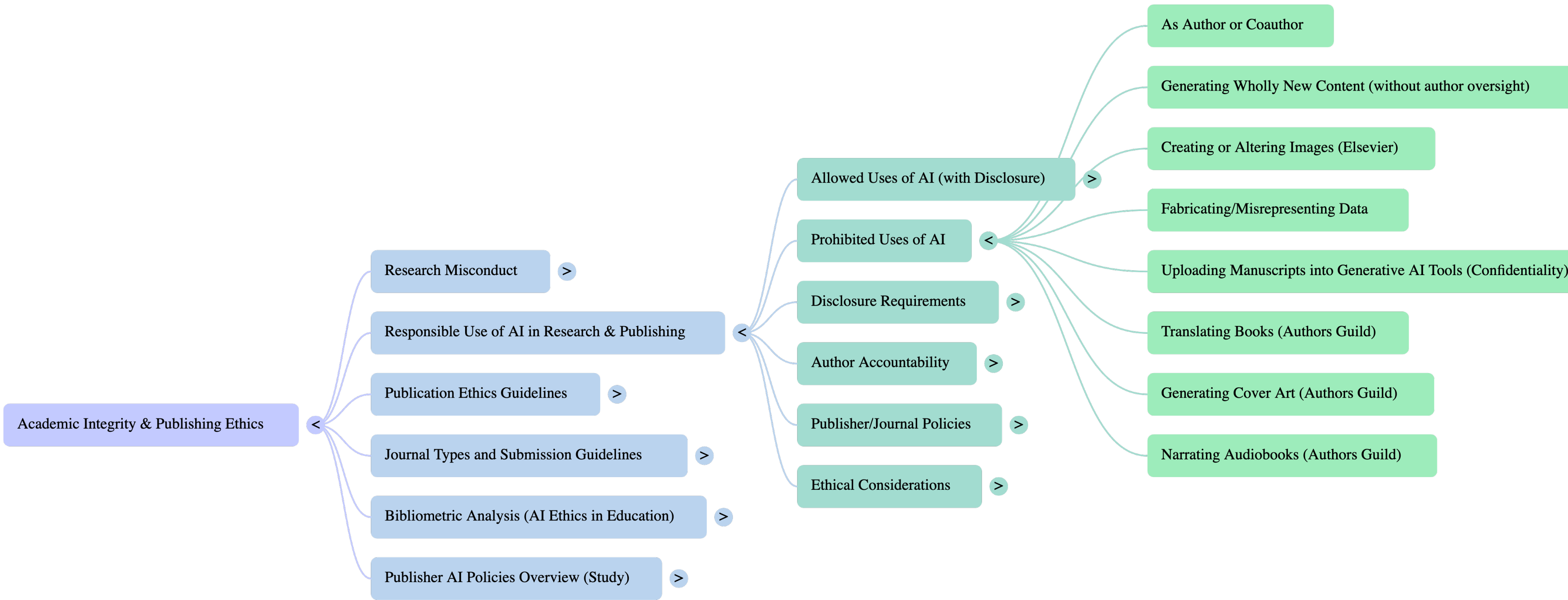
### Studio

Audio Overview Video Overview

Mind Map Reports

- Ethical AI for Researchers 38 sources · 5h ago
- AI in Research: New Rules 38 sources · 7h ago
- จริยธรรมและการใช้ AI ในงานวิจัย 38 sources · 7h ago
- AI จริยธรรม และการตีพิมพ์ทางวิชาการ Briefing doc · 38 sources · 92d ago
- หลักความสมบูรณ์ทางวิชาการและนโยบาย... Briefing doc · 33 sources · 100d ago
- นโยบาย AI สำหรับผู้เขียนโดย...

Add note




สร้าง Mindmap จากเอกสารทั้งหมดที่ให้

# 3

## Following New Rules

Navigating new terrain

 NotebookLM

<https://youtu.be/eGwB2I30WTM>

สร้าง VDO นำเสนอเนื้อหาจากเอกสารทั้งหมดที่ให้



ใช้เขียนโปรแกรม/คำนวณสถิติ

Programming and statistical analysis with AI

# ให้ ChatGPT เขียน code ที่ต้องการ



ช่วยเขียน code สร้างข้อมูล main\_idea : { type : implicit|explicit, text : "xxxxxx" } โดยดูว่า ถ้ามีข้อความใน "Comments" เช่น "Comments": ["การดำเนินชีวิตของมนุษย์ต้องเป็นไปอย่างมีเป้าหมายและวิธีการ"] ให้เอาข้อความนั้นมาใช้เป็น text ใน main\_idea และ type เป็น implicit ถ้าเป็น "label":[[17,111,"main"]] ให้เอาข้อความใน "text" ตำแหน่งที่ 17 ถึง 111 มาเป็น text ใน main\_idea และ type เป็น explicit  
check ข้อมูลที่ละอันจากที่อ่านจากไฟล์ jsonl จนหมด  
save "id" "text" "main\_idea" เก็บในไฟล์ output เป็น jsonl ทุกอันแม้ว่าจะไม่มี main\_idea กำกับ

โค้ดด้านล่างนี้จะอ่านข้อมูลจากไฟล์ .jsonl ทีละบรรทัด ตรวจสอบว่ามี Comments หรือ label ที่เกี่ยวข้องกับ main\_idea หรือไม่ แล้วสร้าง field main\_idea ตามเงื่อนไขที่ให้มา และบันทึกผลลัพธ์ที่ประกอบด้วย id, text, และ main\_idea (หรือ null ถ้าไม่มี) ลงในไฟล์ output ใหม่ในรูปแบบ .jsonl :

```
python Copy Edit  
  
import json  
  
input_file = "input.jsonl"  
output_file = "output.jsonl"  
  
def extract_main_idea(data):  
    text = data.get("text", "")  
  
    # Case 1: implicit main idea from Comments  
    comments = data.get("Comments", [])  
    if comments:  
        return {  
            "type": "implicit",  
            "text": comments[0] # ใช้ันแรกเท่านั้นตามที่ระบุ  
        }  
}
```

AW

I have data in Excel in five columns. Each column is the judgement of "semantic prosody" from each subject. The labels can be "neg", "pos", or "neu". Write Python code to read the data and calculate Kappa coefficient to determine agreement degree among subjects.



## ช่วยเขียน โปรแกรมคำนวณสถิติ



Sure, you can use the pandas library to read your data from Excel, and then use the sklearn library to calculate the Kappa coefficient. Here's a simple code snippet:



python

Copy code

```
import pandas as pd
from sklearn.metrics import cohen_kappa_score

# Load the data from Excel
df = pd.read_excel('your_file.xlsx')

# Calculate Kappa coefficient for each pair of judges
kappas = []
for i in range(5):
    for j in range(i+1, 5):
        kappa = cohen_kappa_score(df.iloc[:, i], df.iloc[:, j])
        kappas.append(kappa)
        print(f'Kappa for judge {i+1} and judge {j+1}: {kappa}')

# Average Kappa coefficient
average_kappa = sum(kappas) / len(kappas)
print(f'Average Kappa: {average_kappa}')
```

Regenerate response



## ให้เขียน code สำหรับ web scraping

Help me scrape reviews from this page [https://www.imdb.com/title/tt20242042/reviews/?ref\\_=tt\\_ov\\_urv](https://www.imdb.com/title/tt20242042/reviews/?ref_=tt_ov_urv). I want to get the review text, the title, and the rating.



It seems that the IMDb page you provided isn't accessible directly through my tools. However, I can guide you through scraping the reviews yourself using Python with the `requests` and `BeautifulSoup` libraries. Here's a step-by-step outline to achieve this:

### Step 1: Install Required Libraries

Make sure you have these libraries installed:

```
bash

pip install requests beautifulsoup4
```

 Copy code

### Step 2: Write the Scraping Script

Here's a script that can scrape the review title, text, and rating from IMDb:

```
python

import requests
from bs4 import BeautifulSoup

def scrape_imdb_reviews(url):
    headers = {
```

 Copy code





# ใช้ Colab และให้ Gemini อ่านไฟล์ข้อมูล เขียนโปรแกรมวิเคราะห์ข้อมูลให้



Untitled2.ipynb ☆ ☁

File Edit View Insert Runtime Tools Help



Share



Gemini

Commands | + Code + Text

Connect ^



Start coding or generate with AI.

Analyze files with Gemini

<https://colab.research.google.com/>

Gemini X

วิเคราะห์ข้อมูลในไฟล์ว่ามีอะไรน่าสนใจบ้าง

mxmh\_survey\_results.csv

Happy to help! Here's a set of tasks I can execute in your notebook to help answer your question.

- **Data loading:** Load the data from "mxmh\_survey\_results.csv" into a pandas DataFrame.
- **Data exploration:** Explore the data to understand its structure, identify missing values, data types of each column, and the distribution of key variables. Determine the shape of the data, the number of unique values in categorical columns.
- **Data cleaning:** Handle missing values appropriately (e.g., imputation or removal) and address any inconsistencies or outliers in the data. Consider converting data types if necessary.
- **Data analysis:** Analyze the data to find interesting patterns and relationships between variables. Focus on exploring the relationship between mental health and other variables like demographics, employment, and technology usage. Calculate descriptive statistics for numerical variables.
- **Data visualization:** Create visualizations (histograms, box plots, scatter

Enter a prompt here



# Gemini เขียน code run โปรแกรม แสดงผลและทำ visualization



Untitled2.ipynb ☆ ☁

File Edit View Insert Runtime Tools Help



Share

Gemini

Commands + Code + Text

## ได้ code ที่มาใช้ run กับข้อมูลไฟล์ได้จริง

**Reasoning:** Calculate the correlation between mental health indicators and other numerical variables, and explore the relationship between mental health indicators and categorical variables.

```
[5] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Correlation analysis
numerical_cols = ['Age', 'Hours per day', 'BPM', 'Anxiety', 'Depression', 'Insomnia', 'OCD']
correlation_matrix = df[numerical_cols].corr()
display(correlation_matrix)
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of Numerical Variables')
plt.show()

# Group comparisons for categorical variables
categorical_cols = ['Gender', 'Primary streaming service'] # Add other categorical variables
mental_health_cols = ['Anxiety', 'Depression', 'Insomnia', 'OCD']

for col in categorical_cols:
    if col in df.columns:
        plt.figure(figsize=(12, 6)) # Adjust figure size
        for mental_health_col in mental_health_cols:
            sns.barplot(x=col, y=mental_health_col, data=df, estimator=np.mean, ci=None) # Use np
            plt.title(f'Mean {mental_health_col} by {col}')
            plt.show()
    else:
        print(f"Column '{col}' not found in the DataFrame.")
```

Gemini X

interesting patterns.

- **Finish task:** Summarize the key findings from the data exploration, cleaning, analysis, and visualization steps in a concise and easy-to-understand report. Include any interesting insights or observations.

Want anything changed? Feel free to send feedback, and I will try to update the plan accordingly!

Execute plan

Rate this answer

Executed plan:

- ✓ Preparing to execute plan
- ✓ Data loading
- ✓ Data exploration
- ✓ Data cleaning
- ✓ Data analysis
- ✓ Data visualization
- ✓ Finish task

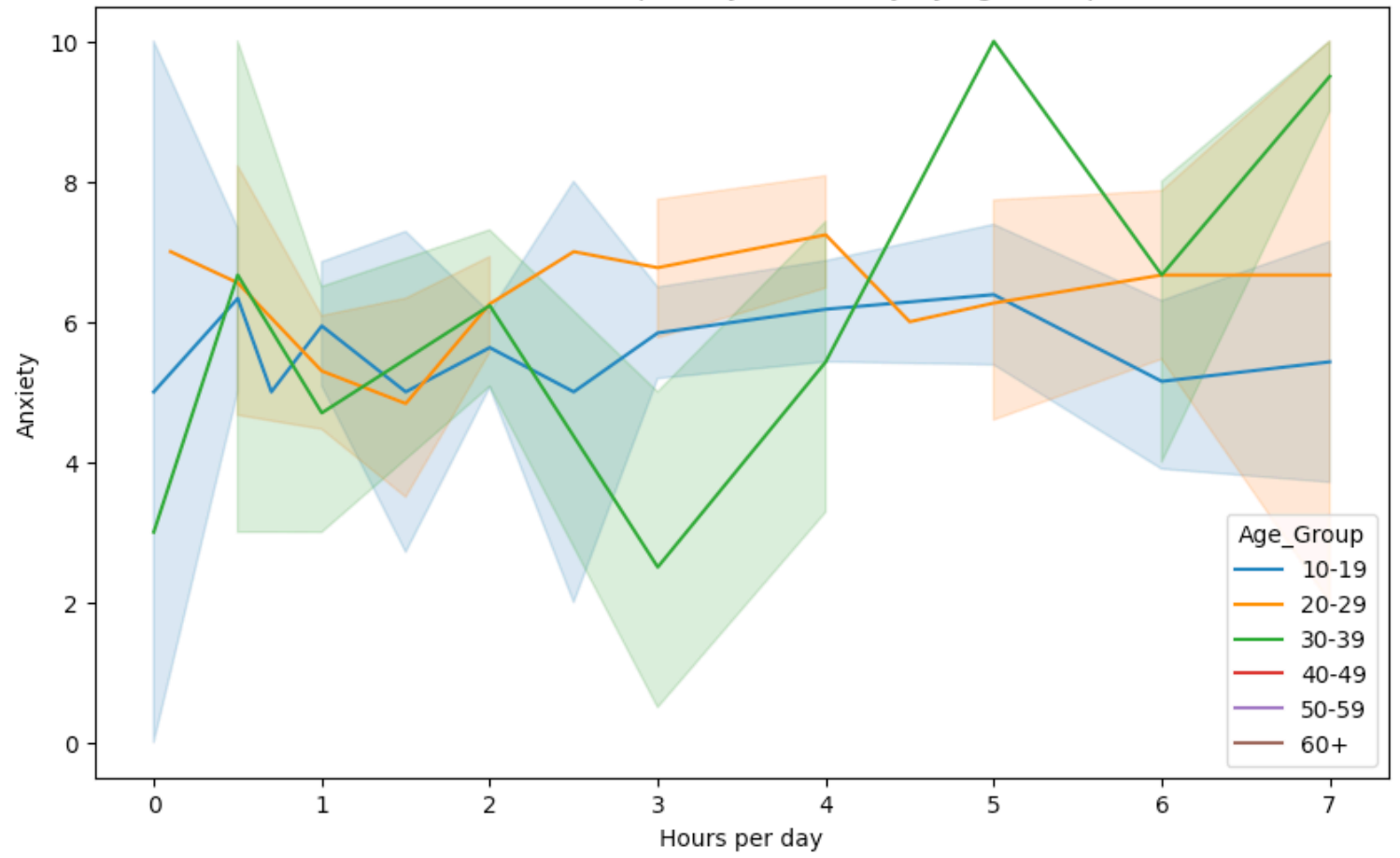
I've finished executing the plan for you!

Rate this answer

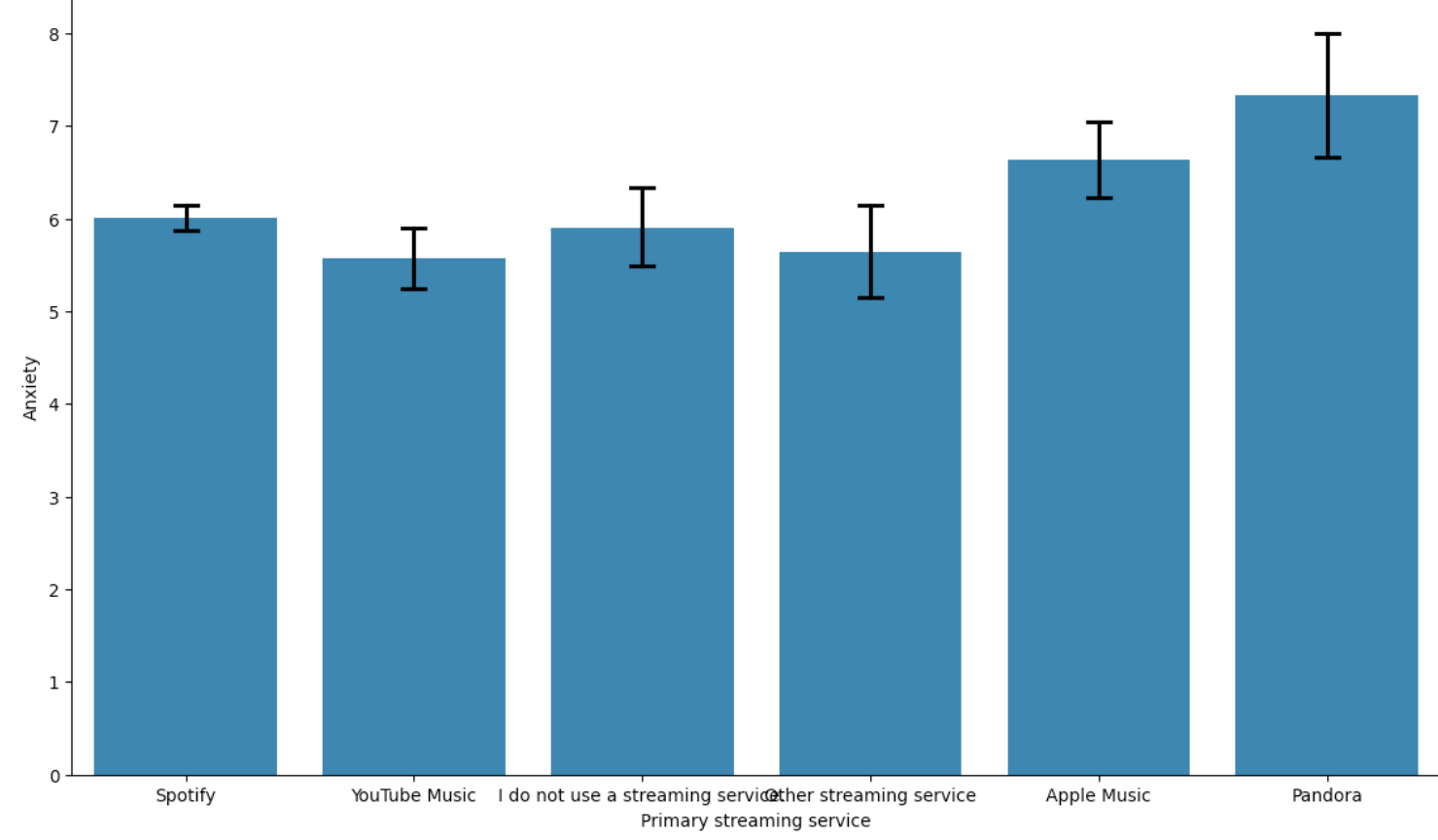
Enter a prompt here



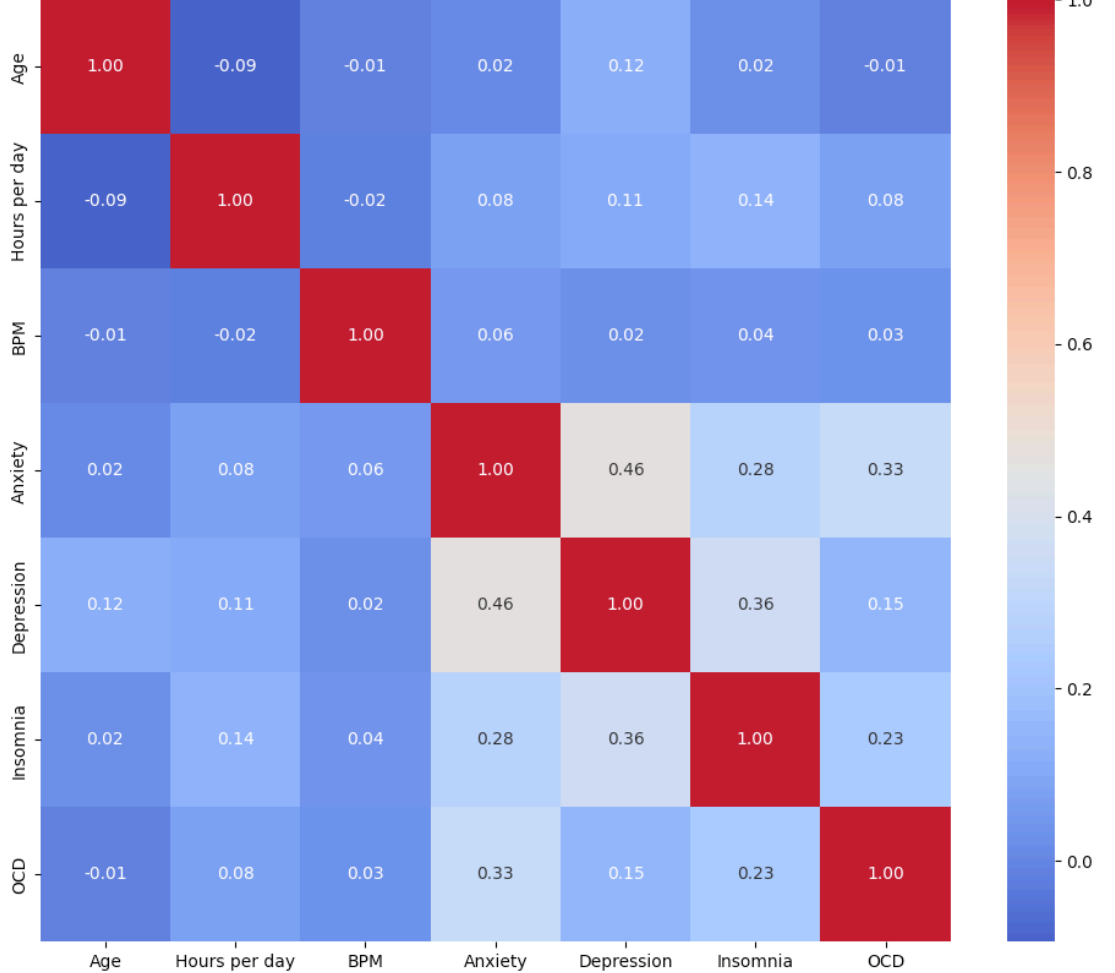
Interaction: Hours per day vs. Anxiety by Age Group



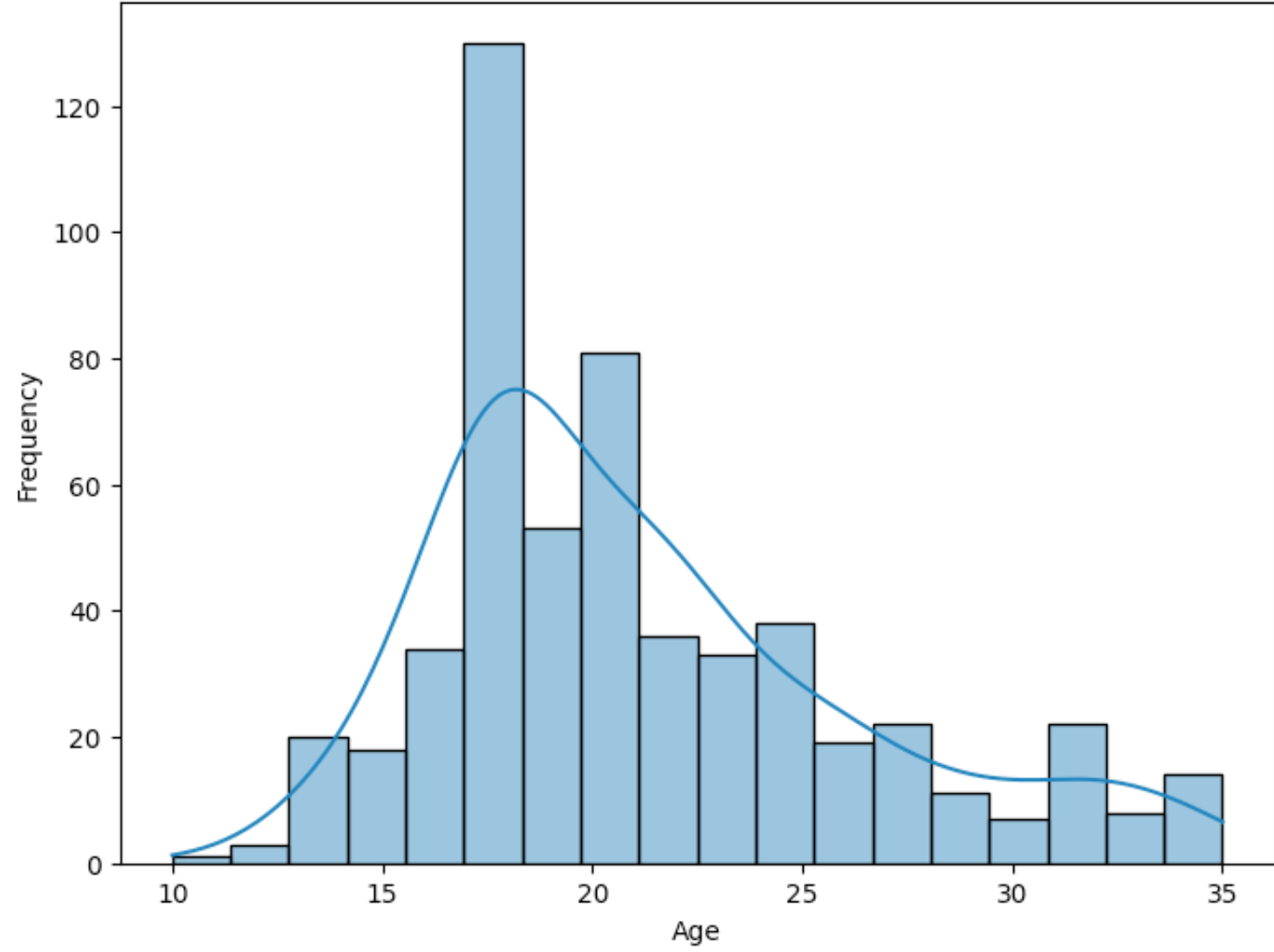
Mean Anxiety by Primary streaming service



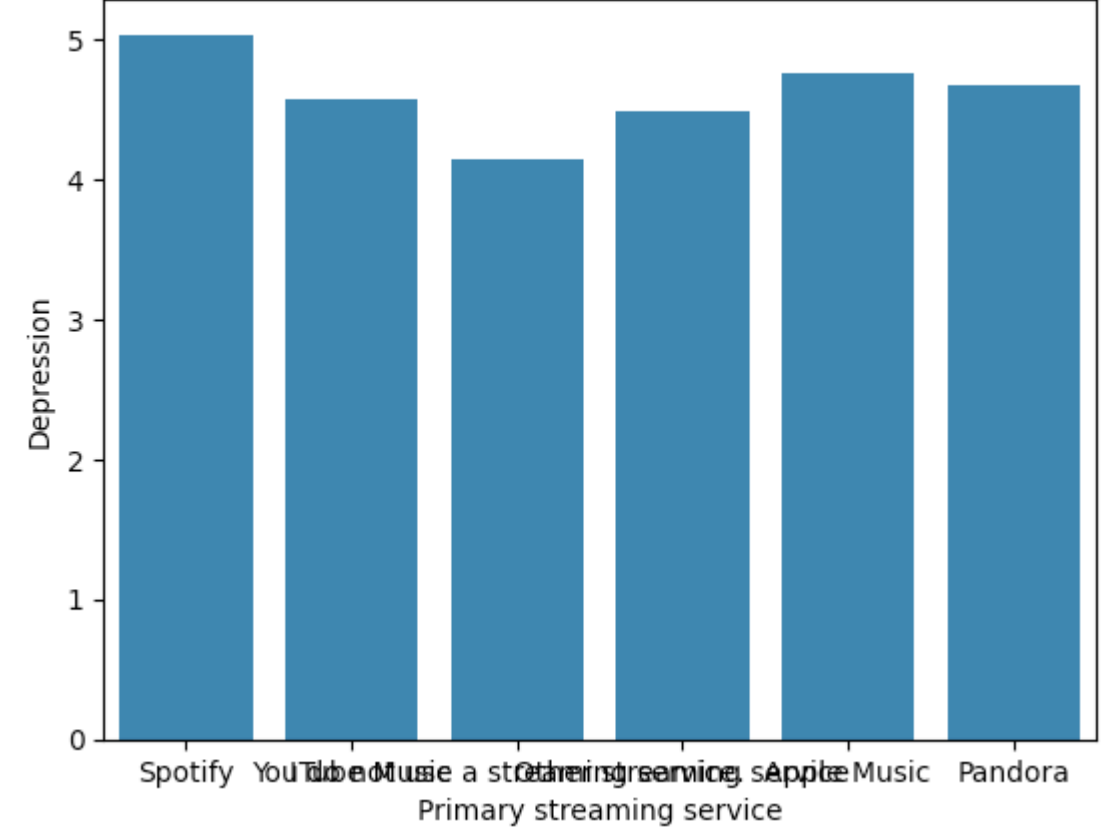
Correlation Matrix of Numerical Variables



Distribution of Age



Mean Depression by Primary streaming service



# Vibe coding



# ให้ Firebase Google สร้าง interactive web สำหรับค้น text แสดงผลว่า main idea และ discourse markers ต่าง ๆ อยู่ส่วนไหน

## เขียนอธิบาย app ที่ต้องการว่ามี interface อย่างไร จัดการข้อมูล และแสดงผลอย่างไร

เขียน web app ที่ให้ user ระบุลักษณะ text ที่ต้องการ. จะมี index ไฟล์. index.json ให้ค้นว่า text ไหนบ้างที่ตรงตามเงื่อนไข จากนั้นให้แสดงรายการ text ที่ตรงเงื่อนไขจำนวน 20 รายการต่อหน้า สามารถกดดูหน้าถัดไปหรือย้อนกลับได้ ในหน้าแรกให้แสดงข้อความว่าเป็น Project Thai Text corpus ของสถาบันภาษาไทยสิรินธร มี option ให้เลือกลักษณะ text ที่ต้องการดังนี้ text\_level: easy, moderate, difficult genre: academic, newspaper, non-academic, fiction, poetry, law, misc read\_type: detailed, comprehensive, interpretive, critical text\_struct: fiction, non-fiction; poetry, figure, diagram, infographic text\_type: descriptive, narrative, informative, argumentative, procedural, evaluative domain: HSS (Humanities and Social Sciences), Sci (Sciences) ใน index.json มีข้อมูล file, level, genre, read\_type,, text\_struct, text\_type, domain. ทุกตัวจะมี value เดียว ยกเว้น read\_type ที่เป็น list มีได้มากกว่า 1 value เมื่อเทียบใน index.json ได้ตรงตามเงื่อนไข ก็จะได้ชื่อไฟล์ แล้วเอาชื่อนั้นไปเปิดไฟล์มาแสดงผลทุกย่อหน้าพร้อมแสดง annotation

เมื่อ user เลือก text ที่ต้องการแล้ว ให้เปิดไฟล์ text นั้นแล้วแสดงเนื้อหาหน้าจจะเป็น text ที่อาจมีหลายย่อหน้าได้ นำข้อมูลใน paras มาแสดง ในนั้นมี pid แต่ละย่อหน้า และ text ที่แสดง มีข้อมูล annotation คือ dms และ main\_idea user สามารถเลือกแสดงหรือไม่แสดง annotation ก็ได้ ถ้าเลือกแสดง main\_idea ให้ดูว่าเป็น explicit ไหม หากใช่ให้ highlight ข้อความตามตำแหน่งที่ระบุ ถ้าเลือก dms แต่ละตัว ให้ highlight คำหรือวลีตามตำแหน่งที่บอก เช่น {"pos": [130, 133], "type": "elaborative", "text": "ว่า"} หมายถึง คำว่า เป็น elaborative อยู่ตำแหน่ง 130-133 ของ paragraph นั้น user เลือกดู dms ทั้งหมดหรือระบุตัวที่ต้องการได้ การแสดงผลจะสลับ highlight หรือไม่ highlight แล้วแต่ที่ได้เลือกไหม ทำเป็น check box ให้เลือก dms ได้ dms ทั้งหมดมีดังนี้ definition\_marker | sequential | conclusion | additive | contrastive | cause-effect | purpose | problem-solving | elaborative | topic\_shift | condition | supposition | stance | inferential | emphasis | topic\_marker | interactive\_marker | emotive\_marker User สามารถเลือกกลับไปหน้าแรกได้ Index file มีข้อมูลแบบนี้ {"text\_type": "informative", "text\_struct": "non-fiction", "domain": null, "file": "PRY142", "genre": ["Newspaper", "Non-academic"], "text\_level": "moderate", "read\_type": ["detailed", "comprehensive", "critical"]}

ส่วน text file PRY142.json มีข้อมูลนี้ {"text": "...", "Comments": [], "text\_type": "informative", "text\_struct": "non-fiction", "file": "PRY142", "source": "ธิษณา จรรยาเลิศ. กิณะมะละกอจีเอ็มโอต้องระวังแห่งภูมิแพ้อียา. ซีวีจิต. ปีที่ ๑๖ : ๑๑ มีนาคม ๒๕๕๗. หน้า ๖๘.", "genre": ["Newspaper", "Non-academic"], "text\_level": "moderate", "read\_type": ["detailed", "comprehensive", "critical"], "paras": [{"pid": 1, "text": "มะละกอจีเอ็มโอที่พบในประเทศไทยคือ มะละกอตัดแต่งพันธุกรรมเพื่อต้านทาน โรคไวรัสใบด่างวงแหวนซึ่งพัฒนาขึ้นเมื่อปลายทศวรรษ ๘๐ ในรัฐฮาวาย ประเทศสหรัฐอเมริกา", "dms": [{"pos": [30, 33], "type": "elaborative", "text": "คือ"}, {"pos": [56, 61], "type": "purpose", "text": "เพื่อ"}, {"pos": [101, 106], "type": "sequential", "text": "เมื่อ"}], "main\_idea": {"type": "explicit", "main": [{"pos": [0, 88], "text": "มะละกอจีเอ็มโอที่พบในประเทศไทยคือ มะละกอตัดแต่งพันธุกรรมเพื่อต้านทาน โรคไวรัสใบด่างวงแหวน"}]}, {"pid": 2, "text": "....."}]}

✦ OK, here's a plan for how we'll prototype this app. Take a look below and customize as needed, or if you're ready, select Prototype this App.

## Filter Options

### Text Level

- easy
- moderate
- difficult

### Genre

- academic
- newspaper
- non-academic
- fiction
- poetry
- law
- misc

### Reading Type

- detailed
- comprehensive
- interpretive
- ...

## Available Texts (6)

### PRY142

Level: moderate  
Genre: Newspaper, Non-academic  
Type: informative

[View Document](#)

### FICT001

Level: easy  
Genre: Fiction  
Type: narrative

[View Document](#)

### ACAD002

Level: difficult  
Genre: Academic  
Type: argumentative

[View Document](#)

low

ตอนนี้คุณสามารถคลิกปุ่ม "View Document" บนการ์ดข้อความแต่ละใบเพื่อดูเนื้อหาของเอกสารนั้นๆ ในกล่องโต้ตอบแล้วครับ และผมได้ปรับความสูงของส่วนควบคุมตัวกรองให้มีพื้นที่มากขึ้นด้วยครับ

## สั่งแก้ไขทีละประเด็น จนกว่าจะได้แบบที่ต้องการ



เมื่อ view document แล้ว อยากให้เลือกว่าจะแสดง main\_idea หรือ dm ประเภทต่าง ๆ ได้



Okay, I understand! You want to be able to toggle the visibility of main ideas and different types of discourse markers (DMs) when viewing a document. That's a great feature for textual analysis!

Here's how I'll implement this:

### 1. Update `src/app/page.tsx`:

- I'll add state variables to keep track of whether to show main ideas ( `showMainIdea` ) and which DM types are selected for highlighting ( `selectedDms` ).
- Inside the document view `Dialog` , I'll add new controls:
  - A checkbox to toggle "Show Main Idea".
  - A list of checkboxes for all available DM types (sourced from `dmsTypes` in your `corpus.ts` ), allowing multi-selection. These will be in a scrollable area if there are many.
- I'll create a new function `renderParagraphWithHighlights` . This function will take a paragraph's data, the `showMainIdea` status, and the `selectedDms` list. It will then reconstruct the paragraph's text, wrapping the main ideas and/or selected DMs in styled `<span>` elements to highlight them.
  - Main ideas will get a distinct background color (e.g., a light shade of primary).
  - Selected DMs will get another distinct background color and an underline (e.g., a light shade of accent).
- The highlighting logic will carefully split the text at the boundaries of these features and apply styles to the correct segments, aiming to handle overlaps gracefully.
- When the dialog is closed, these display options will be reset to their default state.

This will make the document viewer much more interactive and useful for analyzing text structure.

ผลที่แก้ไขตามที่ปรับ แสดง main idea ด้วยการ highlight

## Display Options

Show Main Idea

Show Discourse Markers

## Legend:

■ Main Idea

■ Discourse Marker (Type shown in red in text)

มะละกอจีเอ็มโอที่พบในประเทศไทยคือ มะละกอดัดแต่งพันธุกรรมเพื่อต้านทานโรคไวรัสใบด่างวงแหวนซึ่งพัฒนาขึ้นเมื่อปลายทศวรรษ ๘๐ ในรัฐฮาวาย ประเทศสหรัฐอเมริกา

ผลกระทบต่อสุขภาพที่อาจเกิดขึ้นจากการบริโภคมะละกอจีเอ็มโอ ยังเป็นที่ถกเถียงกันในวงกว้าง บางงานวิจัยชี้ให้เห็นถึงความเสี่ยงที่อาจเกิดขึ้นได้ เช่น การเกิดภูมิแพ้ หรือการดื้อยาปฏิชีวนะในมนุษย์

ดังนั้น ผู้บริโภคควรใช้วิจารณญาณในการเลือกรับประทาน และติดตามข้อมูลข่าวสารจากแหล่งที่น่าเชื่อถืออย่างสม่ำเสมอ เพื่อประกอบการตัดสินใจ

ข้อมูลเป็นข้อมูลที่โปรแกรมสมมติให้ จึงมีส่วนที่แสดงผลผิดได้

ผลที่แก้ไขตามที่ปรับ แสดง DM ด้วยการ highlight และบอกประเภท

## Display Options

- Show Main Idea
- Show Discourse Markers

### Legend:

- Main Idea
- Discourse Marker (Type shown in red in text)

มะละกอจีเอ็มโอที่พบในประเทศไทยคือ (elaborative) มะละกอดัดแต่งพันธุกรรมเพื่อ (purpose) ด้านทาน โรคไวรัสใบด่างวงแหวนซึ่งพัฒนาขึ้นเมื่อ (sequential) ปลายทศวรรษ ๘๐ ในรัฐฮาวาย ประเทศสหรัฐอเมริกา

ผลกระทบต่อสุขภาพที่อาจเกิดขึ้นจากการบริโภคมะละกอจีเอ็มโอ ยังเป็น (contrastive) นที่ถกเถียงกันในวงกว้าง บางงานวิจัยชี้ (cause-effect) ให้เห็นถึงความเสี่ยงที่อาจเกิดขึ้นได้ เช่น การเกิดภูมิแพ้ หรือการดื้อยาปฏิชีวนะในมนุษย์

ดังนั้น (conclusion) น ผู้บริโภคควรใช้วิจารณญาณในการเลือกรับประทาน และติดตามข้อมูลข่าวสารจากแหล่งที่ (purpose) น่าเชื่อถืออย่างสม่ำเสมอ เพื่อประกอบการตัดสินใจ

ข้อมูลเป็นข้อมูลที่โปรแกรมสมมติให้ จึงมีส่วนที่แสดงผลผิดได้

ในหน้าแรกที่เป็น Filter option. ยกเว้น Reading Type ที่เป็น check box. Filter อื่นให้เป็น drop down menu. Title of page ให้ขึ้นข้อความ Sirindhorn Thai Language Institute ด้วย

## สั่งให้แก้ไขหน้าแรกใหม่ ให้ใช้ drop down menu ในการเลือก



แน่นอนครับ! ผมจะปรับปรุงหน้าแรกให้ตรงตามที่คุณต้องการ โดย:

### 1. เปลี่ยน Filter Controls (ยกเว้น Reading Type):

- "Text Level", "Genre", "Text Structure", "Text Type", และ "Domain" จะถูกเปลี่ยนจากกลุ่ม Checkbox เป็น Dropdown Menu ที่สามารถเลือกได้หลายตัวเลือก (multi-select) โดยใช้ `DropdownMenuCheckboxItem`
- "Reading Type" จะยังคงเป็นกลุ่ม Checkbox เหมือนเดิม

### 2. อัปเดต Title ของหน้า:

- อัปเดตชื่อที่แสดงในแท็บของเบราว์เซอร์ (`<title>`) และหัวเรื่องหลัก (`<h1>`) บนหน้าเพจให้มีคำว่า "Sirindhorn Thai Language Institute"

นี่คือการเปลี่ยนแปลงโค้ดครับ:

File changes

d5ca8b91 ✓ Current

🌀 src/app/layout.tsx

🌀 src/components/custom/FilterControls.tsx

🌀 src/app/page.tsx

ผมได้ทำการเปลี่ยนแปลงตามที่คุณต้องการแล้วครับ:

### 1. Filter Controls:

- ใน `src/components/custom/FilterControls.tsx` ได้เปลี่ยน "Text Level", "Genre", "Text Structure", "Text Type", และ "Domain" ให้เป็น Dropdown Menu ที่คุณสามารถเลือกหลายตัวเลือกได้ (multi-select) ปุ่ม Dropdown จะแสดงจำนวนตัวเลือกที่ถูกเลือก หรือ "(All)" หากยังไม่ได้เลือก
- "Reading Type" ยังคงเป็น Checkbox เหมือนเดิม

### 2. Page Titles:

- ใน `src/app/layout.tsx` ได้อัปเดต `metadata.title` ให้เป็น "Sirindhorn Thai Language Institute - Thai Text Explorer"



## หน้าจอที่ปรับแก้แล้ว

### Filter Options

Text Level (All) ▾

Genre (All) ▾

### Reading Type

- detailed
- comprehensive
- interpretive
- critical

Text Structure (All) ▾

Text Type (All) ▾

### Text Type

- descriptive
- narrative
- informative**
- argumentative
- procedural
- evaluative

### Available Texts (6)

#### PRY142

Level: moderate  
Genre: Newspaper, Non-academic  
Type: informative

[View Document](#)

#### FICT001

Level: easy  
Genre: Fiction  
Type: narrative

[View Document](#)

#### ACAD002

Level: difficult  
Genre: Academic  
Type: argumentative

[View Document](#)

#### POEM003

EXPLORER

- STUDIO
  - > .idx
  - > .next
  - > .vscode
    - { settings.json
  - > docs
    - blueprint.md
  - > node\_modules
  - > public/data
    - { ACAD002.json
    - { FICT001.json
    - { index.json
    - { PRY142.json
  - > src
    - > ai
    - > app
      - ★ favicon.ico
      - globals.css
      - layout.tsx
      - page.tsx
  - > components
    - > custom
    - > ui
    - > hooks
    - > lib
    - > types
  - gitignore
  - .modified
  - apphosting.yaml
  - components.json
  - next-env.d.ts

page.tsx x Web Gemini

src > app > page.tsx > ...

```

36  export default function Home() {
88
89      const handleCloseModal = () => {
90          setIsDisplayModalOpen(false);
91          setSelectedDocument(null);
92          setShowMainIdea(false);
93          setShowDms(false);
94      };
95
96      function renderParagraphWithHighlights(
97          paragraph: Paragraph,
98          showMainIdeaFlag: boolean,
99          showDmsFlag: boolean
100     ): JSX.Element {
101         const { text, main_idea, dms, pid } = paragraph;
102
103         const implicitMainIdeaElement = (
104             showMainIdeaFlag &&
105             main_idea &&
106             main_idea.type === 'implicit' &&
107             main_idea.main &&
108             main_idea.main.length > 0 &&
109             main_idea.main[0].text
110         ) ? (
111             <p className="text-sm font-medium text-primary mb-1">
112                 <strong className="font-headline">implicit:</strong> {main_idea.main[0].text}
113             </p>
114         ) : null;
115
116         let points = new Set<number>([0, text.length]);
117         const features: Array<{start: number; end: number; type: 'mainIdea' | 'dm'; dmType?: string; text: string}> =
118
119         if (showMainIdeaFlag && main_idea && main_idea.type === 'implicit' && main_idea.main) {
    
```



ออกแบบแบบสอบถาม

Questionnaire design



ให้ช่วยออกแบบงานวิจัยเชิงสำรวจเรื่อง Public Perceptions of Fairness and Impartiality in the Criminal Justice System ในประเทศไทย

Reasoned about research design on fairness perceptions in Thailand's justice system for 30 seconds

>

การออกแบบงานวิจัยเชิงสำรวจในหัวข้อ “มุมมองของสาธารณชนต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรมทางอาญาในประเทศไทย” ควรเริ่มต้นด้วยการระบุวัตถุประสงค์หลักของการวิจัยเพื่อศึกษาการรับรู้ของประชาชนต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรม รวมถึงการระบุปัจจัยที่อาจมีผลต่อการรับรู้ดังกล่าว ไม่ว่าจะเป็นประสบการณ์ส่วนตัว ปัจจัยทางสังคมและเศรษฐกิจ หรือความเชื่อในระบบการบริหารงานของรัฐ

ในขั้นตอนแรก นักวิจัยควรทบทวนวรรณกรรมที่เกี่ยวข้องกับทฤษฎีและแนวคิดในด้านความเชื่อมั่นในระบบยุติธรรม การรับรู้ความเป็นธรรม และแนวคิดด้านความโปร่งใสในกระบวนการยุติธรรมทั้งในบริบทสากลและในประเทศไทย ซึ่งจะช่วยสร้างกรอบแนวคิด (conceptual framework) ที่ชัดเจนสำหรับการกำหนดตัวแปรและตั้งคำถามวิจัย โดยคำถามวิจัยอาจรวมถึง “ประชาชนมีความเชื่อมั่นต่อความเป็นธรรมในระบบกระบวนการยุติธรรมทางอาญาหรือไม่?”, “ปัจจัยใดที่มีอิทธิพลต่อการรับรู้ในด้านความเป็นธรรมและความไม่ลำเอียง?”, และ “ประเด็นหรือปัญหาที่ประชาชนพบเจอในการปฏิบัติการในระบบยุติธรรมมีลักษณะอย่างไร?”

ในส่วนของวิธีการวิจัย ควรใช้แนวทางแบบผสมผสาน (mixed methods) โดยเริ่มจากการใช้แบบสอบถามเชิงปริมาณเพื่อเก็บข้อมูลจากกลุ่มประชาชนที่มีความหลากหลายทางภูมิศาสตร์และประชากร (stratified random sampling) เพื่อให้ได้ภาพรวมของความคิดเห็นในระดับประเทศ จากนั้นใช้การสัมภาษณ์เชิงลึก (in-depth interviews) หรือกลุ่มสนทนากลุ่ม (focus group discussions) เพื่อเจาะลึกประเด็นที่สำคัญและเก็บข้อมูลเชิงคุณภาพที่ช่วยอธิบายผลลัพธ์เชิงปริมาณที่ได้ โดยในการออกแบบแบบสอบถาม นักวิจัยควรพิจารณา รวมถึงมาตรวัดความเชื่อมั่นในระบบยุติธรรม ความรู้สึกต่อความเป็นธรรมและความไม่ลำเอียงของเจ้าหน้าที่ รวมถึงประสบการณ์ส่วนบุคคลในการมีปฏิสัมพันธ์กับระบบกระบวนการยุติธรรม

สำหรับการเก็บข้อมูล นักวิจัยควรดำเนินการเก็บข้อมูลในเวลาที่เหมาะสม โดยคำนึงถึงการเข้าถึงกลุ่มตัวอย่างในพื้นที่ทั้งในเขตเมืองและชนบท และควรเตรียมการฝึกอบรมสำหรับผู้ทำการเก็บข้อมูลเพื่อให้ได้ข้อมูลที่มีความน่า

ให้ช่วยออกแบบงานวิจัย  
เชิงสำรวจในเรื่องที่กำหนด



Reasoned for 21 seconds >

ด้านล่างนี้เป็นแบบสอบถามที่ออกแบบมาเพื่อใช้ในงานวิจัย “มุมมองของสาธารณชนต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรมทางอาญาในประเทศไทย” โดยมีการแบ่งออกเป็นหลายส่วนเพื่อให้ครอบคลุมข้อมูลพื้นฐาน ประสบการณ์ส่วนบุคคล ความรู้เกี่ยวกับความเป็นธรรมและความไม่ลำเอียง ในกระบวนการยุติธรรม ตลอดจนปัจจัยที่มีอิทธิพลต่อมุมมองของผู้ตอบแบบสอบถาม นอกจากนี้ยังมีส่วนสำหรับข้อเสนอแนะและความคิดเห็นเปิดเพื่อให้ผู้ตอบได้แสดงความคิดเห็นในเชิงคุณภาพ

## ให้ทดลองสร้างแบบสอบถาม

**แบบสอบถาม: ความรู้ของประชาชนต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรมทางอาญาในประเทศไทย**

**คำแนะนำเบื้องต้น:**

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อสำรวจความคิดเห็นของประชาชนต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรม ทางเราขอรับรองว่าข้อมูลที่ท่านให้มาจะถูกเก็บรักษาเป็นความลับและใช้เพื่อการวิจัยเท่านั้น การตอบแบบสอบถามนี้เป็นไปโดยสมัครใจ ท่านสามารถข้ามคำถามที่ไม่ต้องการตอบได้ และเมื่อท่านตอบแบบสอบถามเสร็จแล้วถือว่าท่านยินยอมเข้าร่วมการวิจัยครั้งนี้



## ส่วนที่ 1. ข้อมูลพื้นฐาน

### 1. เพศ:

- ชาย
- หญิง
- อื่น ๆ
- ไม่ประสงค์จะระบุ

### 2. อายุ:

- 18-24 ปี
- 25-34 ปี
- 35-44 ปี
- 45-54 ปี
- 55 ปีขึ้นไป

### 3. ระดับการศึกษา:

- ไม่จบประถมศึกษา
- ประถมศึกษา
- มัธยมศึกษาตอนต้น/ปลาย
- ปวช.
- ปวส./ปริญญาตรี
- ปริญญาโท/ปริญญาเอก

### 4. อาชีพ:

(โปรดระบุ): \_\_\_\_\_

### 5. ที่อยู่อาศัย:

- เมือง
- ชนบท



## ส่วนที่ 2. ประสพการณ์ส่วนบุคคลเกี่ยวกับระบบยุติธรรม

1. ท่านเคยมีประสบการณ์เกี่ยวกับกระบวนการยุติธรรมทางอาญาโดยตรงหรือไม่ (เช่น เป็นพยาน ผู้ถูกกล่าวหา หรือมีประสบการณ์ในฐานะผู้เกี่ยวข้องในกระบวนการ):

- เคย
- ไม่เคย

2. หากเคยประสบการณักรุณาระบุประเภทของประสบการณ์ที่ท่านมี (เลือกได้มากกว่า 1 ข้อ):

- เป็นพยาน
- ผู้ถูกกล่าวหา
- ครอบครัวของผู้เกี่ยวข้อง
- ปฏิบัติงานในหน่วยงานที่เกี่ยวข้องกับระบบยุติธรรม
- อื่น ๆ (โปรดระบุ): \_\_\_\_\_

3. ท่านรู้สึกว่าการดังกล่าวมีผลต่อความเชื่อมั่นของท่านต่อระบบยุติธรรมหรือไม่ โปรดให้คะแนนความรู้สึกของท่านในระดับ 1 ถึง 5 โดยที่

1 = ไม่มีผลเลย

5 = มีผลอย่างมาก

1  2  3  4  5



### ส่วนที่ 3. ความรู้ต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรม

ในส่วนนี้ ท่านจะได้ให้ความคิดเห็นเกี่ยวกับแต่ละประเด็น โดยใช้มาตราส่วน 5 ระดับ (1 = ไม่เห็นด้วยอย่างยิ่ง, 2 = ไม่เห็นด้วย, 3 = กลางๆ, 4 = เห็นด้วย, 5 = เห็นด้วยอย่างยิ่ง) ดังนี้

1. ระบบกระบวนการยุติธรรมในประเทศไทยดำเนินการอย่างเป็นธรรม  
 1  2  3  4  5
2. การตัดสินคดีในระบบยุติธรรมทางอาญามีความโปร่งใส  
 1  2  3  4  5
3. เจ้าหน้าที่ในระบบยุติธรรมมีความเป็นกลางและไม่ลำเอียงในการดำเนินคดี  
 1  2  3  4  5
4. กระบวนการสอบสวนและพิจารณาคดีมีการปฏิบัติตามหลักการยุติธรรมอย่างเคร่งครัด  
 1  2  3  4  5
5. การตัดสินคดีมักถูกกำหนดโดยอิทธิพลจากปัจจัยภายนอก (เช่น การเมืองหรือความสัมพันธ์) มากกว่าหลักฐานและกฎหมาย  
 1  2  3  4  5
6. ความคิดเห็นของประชาชนมีผลต่อการพัฒนาระบบยุติธรรม  
 1  2  3  4  5
7. ระบบยุติธรรมมีมาตรการที่ชัดเจนในการแก้ไขข้อผิดพลาดเมื่อเกิดความไม่ยุติธรรม  
 1  2  3  4  5
8. การสื่อสารข้อมูลเกี่ยวกับกระบวนการยุติธรรมมีความชัดเจนและเข้าถึงได้ง่ายสำหรับประชาชน  
 1  2  3  4  5



## ส่วนที่ 4. ปัจจัยที่มีอิทธิพลต่อการรับรู้

ในส่วนนี้ ท่านจะให้ความคิดเห็นเกี่ยวกับว่าปัจจัยต่าง ๆ ส่งผลกระทบต่อมุมมองของท่านต่อระบบยุติธรรมเพียงใด โดยใช้มาตราส่วน 5 ระดับ (1 = ไม่มีผลเลย, 2 = มีผลน้อย, 3 = ปานกลาง, 4 = มีผล, 5 = มีผลอย่างมาก):

1. ประสบการณ์ส่วนตัวของท่านมีผลต่อความเชื่อมั่นในระบบยุติธรรม  
 1  2  3  4  5
2. ข้อมูลข่าวสารและสื่อมวลชนมีผลต่อความเห็นของท่านต่อระบบยุติธรรม  
 1  2  3  4  5
3. การรับรู้และความคิดเห็นของคนรอบข้างมีผลต่อมุมมองของท่านต่อระบบยุติธรรม  
 1  2  3  4  5
4. ระดับการศึกษาของท่านมีผลต่อความเข้าใจและความเห็นต่อระบบยุติธรรม  
 1  2  3  4  5
5. ความน่าเชื่อถือของแหล่งข้อมูลข่าวสารเกี่ยวกับระบบยุติธรรมมีผลต่อความคิดเห็นของท่าน  
 1  2  3  4  5

---

## ส่วนที่ 5. ข้อเสนอแนะและความคิดเห็นเปิด

1. ท่านคิดว่าควรมีการปรับปรุงหรือเปลี่ยนแปลงในด้านใดของระบบยุติธรรมเพื่อเพิ่มความเป็นธรรมและความไม่ลำเอียง?  
(โปรดระบุความคิดเห็นหรือแนวทางที่ท่านเห็นควรปรับปรุง):





## ให้ปรับบางส่วนของแบบสอบถาม

Reasoned for 23 seconds >

ด้านล่างนี้เป็นแบบสอบถามที่ปรับปรุงรายละเอียดในส่วนที่ 3 และส่วนที่ 4 โดยมีจำนวนข้อและรายละเอียดที่เจาะจงมากขึ้น โดยใช้มาตราส่วน Likert 5 ระดับสำหรับการตอบคำถาม (โดยที่ 1 = ไม่เห็นด้วยอย่างยิ่ง/ไม่มีผลเลย, 2 = ไม่เห็นด้วย/มีผลน้อย, 3 = กลางๆ/มีผลปานกลาง, 4 = เห็นด้วย/มีผล, 5 = เห็นด้วยอย่างยิ่ง/มีผลอย่างมาก)

### ส่วนที่ 3. ความรู้ต่อความเป็นธรรมและความไม่ลำเอียงในระบบกระบวนการยุติธรรม

โปรดให้คะแนนความเห็นของท่านในแต่ละประเด็นต่อไปนี้:

1. ความโปร่งใสของกระบวนการยุติธรรม:

“ระบบกระบวนการยุติธรรมในประเทศไทยดำเนินการโดยเปิดเผยข้อมูลต่อสาธารณชนอย่างเพียงพอ”

1  2  3  4  5

2. ความเป็นกลางในการตัดสินคดี:

“การตัดสินคดีอาญามีความเป็นกลางและอิงหลักฐานที่เป็นจริง โดยไม่ถูกกดดันจากปัจจัยภายนอก”

1  2  3  4  5

3. ความถูกต้องของขั้นตอนสอบสวน:

“การสอบสวนในคดีอาญามีการรวบรวมและประเมินหลักฐานอย่างรอบคอบและเป็นธรรม”

1  2  3  4  5

4. ความเที่ยงตรงของเจ้าหน้าที่:

“เจ้าหน้าที่ตำรวจและอัยการดำเนินคดีด้วยความเป็นกลางและไม่มีการลำเอียง”

ขอคำแนะนำในการเก็บข้อมูลลงในไฟล์ Excel ว่าควรมีกี่ column เก็บข้อมูลอะไร และจะใช้ข้อมูลนี้คำนวณอย่างไร



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย

Reasoned for 18 seconds >

ในการจัดเก็บข้อมูลจากแบบสอบถามลงในไฟล์ Excel ควรออกแบบ โครงสร้างของไฟล์ให้เป็นระเบียบและครอบคลุมตัวแปรที่สำคัญทั้งหมด โดยแต่ละแถว (row) จะแทนข้อมูลของผู้ตอบแต่ละราย และแต่ละคอลัมน์ (column) จะแทนตัวแปรหรือตัวชี้วัดที่นักวิจัยต้องการเก็บไว้ ซึ่งจะทำให้ขั้นตอนการวิเคราะห์และคำนวณข้อมูลในภายหลังเป็นไปอย่างมีระบบและง่ายต่อการประมวลผล ตัวอย่างคำแนะนำในการจัดเก็บข้อมูลมีดังนี้

## 1. โครงสร้างของไฟล์ Excel และการกำหนด Column

- ข้อมูลส่วนบุคคลและข้อมูลพื้นฐาน:

ควรจัดเก็บไว้ในคอลัมน์แรกของไฟล์ เช่น

- **Column A:** รหัสผู้ตอบ (Respondent ID) หรือหมายเลขลำดับ (เพื่อระบุแต่ละผู้ตอบอย่างไม่ซ้ำกัน)
- **Column B:** Timestamp (วันที่และเวลาที่ตอบแบบสอบถาม)
- **Column C:** เพศ (Male, Female, Other, ไม่ระบุ)
- **Column D:** อายุ (สามารถแบ่งช่วงอายุเป็นกลุ่ม เช่น 18–24, 25–34 ฯลฯ)
- **Column E:** ระดับการศึกษา
- **Column F:** อาชีพ
- **Column G:** ที่อยู่อาศัย (เมือง/ชนบท)






- ข้อมูลประสบการณ์ส่วนบุคคลเกี่ยวกับระบบยุติธรรม:

ควรมีคอลัมน์เฉพาะ เช่น

- Column I: ประสบการณ์ตรงกับกระบวนการยุติธรรม (เคย/ไม่เคย)
- Column J: ประเภทของประสบการณ์ (หากเลือกได้มากกว่า 1 ตัวเลือก อาจจัดเก็บเป็นรหัสที่แยกกันหรือใช้เครื่องหมายคั่น เช่น "1;3" สำหรับผู้ที่เลือก "เป็นพยาน" และ "ครอบครัวของผู้เกี่ยวข้อง")
- Column K: คะแนนผลกระทบจากประสบการณ์ (ใช้มาตราส่วน 1-5)

- ข้อมูลแบบสอบถามสเกล Likert (ส่วนที่ 3: ความรู้ต่อความเป็นธรรมและความไม่ลำเอียง):

กำหนดคอลัมน์ให้แต่ละคำถามมีคอลัมน์ของตัวเอง โดยอาจใช้รหัสคำถามเพื่อความชัดเจน ตัวอย่างเช่น

- Column L: Q3.1 ความโปร่งใสของกระบวนการยุติธรรม
- Column M: Q3.2 ความเป็นกลางในการตัดสินคดี
- Column N: Q3.3 ความถูกต้องของขั้นตอนสอบสวน
- Column O: Q3.4 ความเที่ยงตรงของเจ้าหน้าที่
- Column P: Q3.5 ความโปร่งใสของการพิจารณาคดีในศาล
- Column Q: Q3.6 ความเป็นระบบและความชัดเจนของขั้นตอน
- Column R: Q3.7 การเข้าถึงข้อมูลของประชาชน
- Column S: Q3.8 ประสิทธิภาพของช่องทางร้องเรียนและทบทวนคดี
- Column T: Q3.9 มาตรการป้องกันความผิด  และลำเอียง

- Column U: Q3.10 ความชัดเจนในการสื่อสารของเจ้าหน้าที่



## 2. โค้ดสำหรับอ่านข้อมูลและวิเคราะห์ข้อมูลใน Excel

ในส่วนนี้จะอ่านไฟล์ CSV ที่สร้างไว้ จากนั้นคำนวณสถิติเบื้องต้นสำหรับคำถามในส่วนที่ 3 และส่วนที่ 4 พร้อมทั้งคำนวณค่า Cronbach's Alpha สำหรับแต่ละชุดคำถาม และทำตัวอย่างการวิเคราะห์เชิงถดถอย

python

Copy

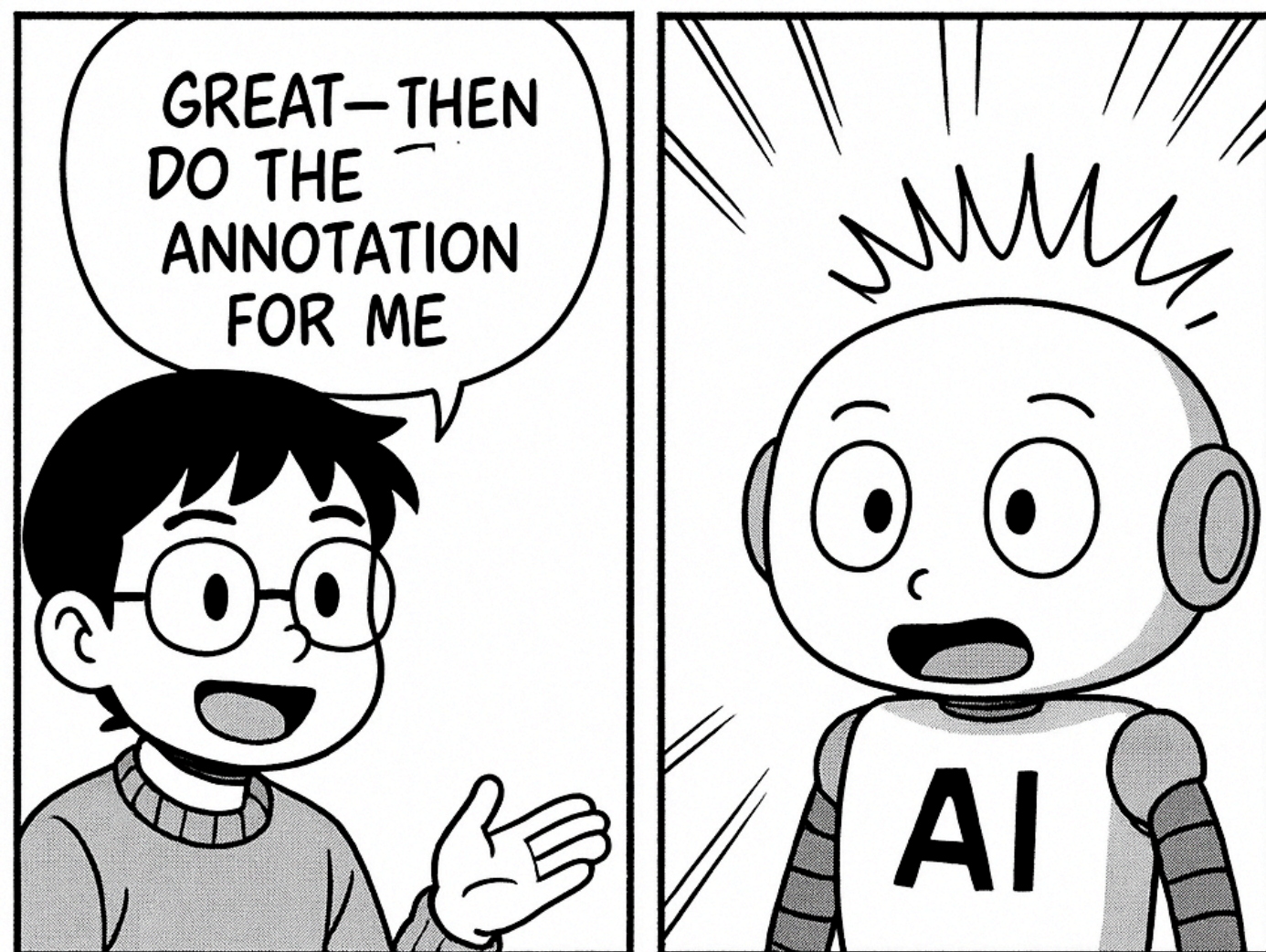
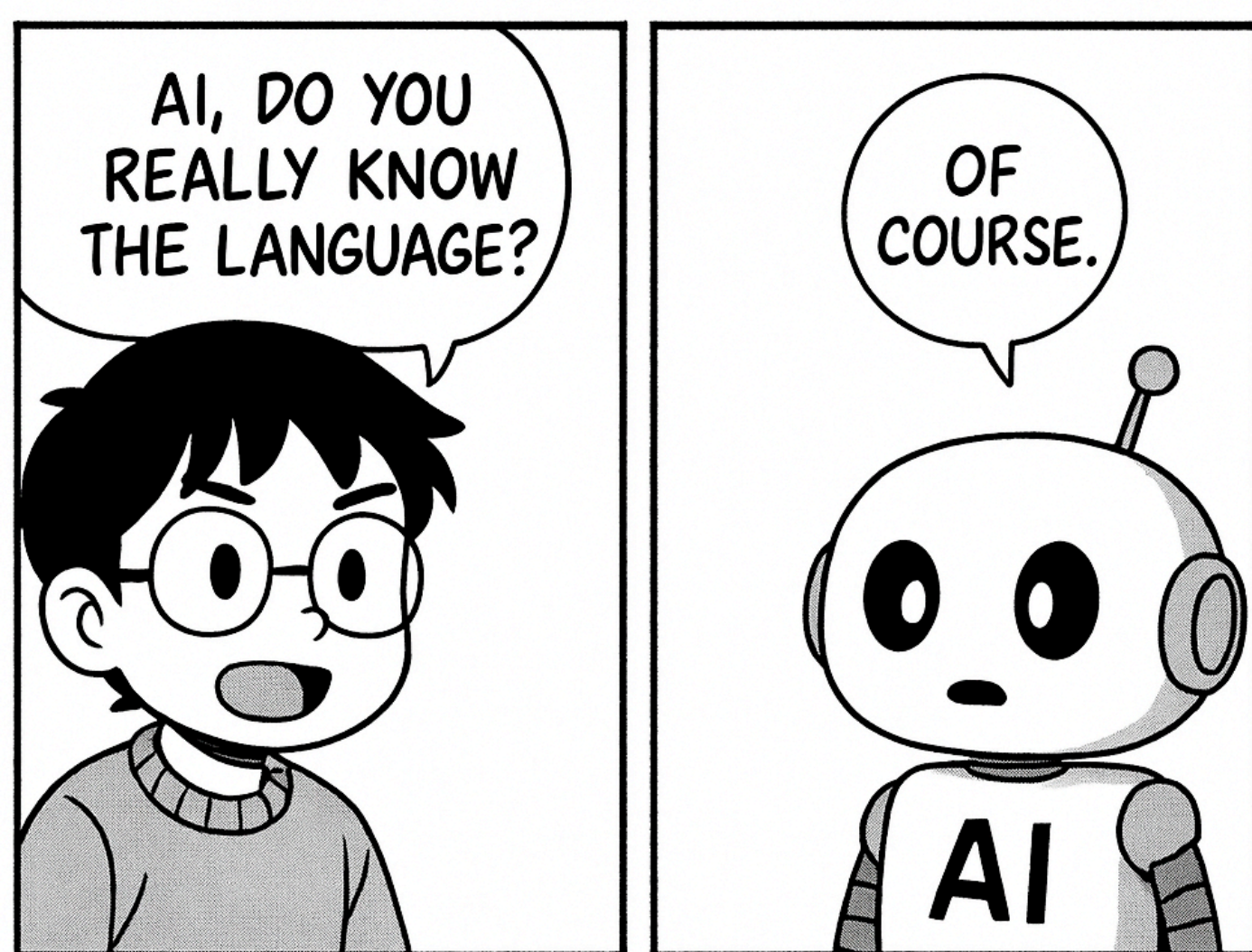
```
import pandas as pd
import numpy as np
import statsmodels.api as sm

# อ่านข้อมูลจากไฟล์ CSV
df_loaded = pd.read_csv("sample_survey_data.csv")

# แสดงข้อมูลตัวอย่าง
print("ตัวอย่างข้อมูล:")
print(df_loaded.head())

# กำหนดรายชื่อคอลัมน์สำหรับส่วนที่ 3 (Q3_1 ถึง Q3_15)
Q3_columns = [col for col in df_loaded.columns if col.startswith("Q3_")]
# กำหนดรายชื่อคอลัมน์สำหรับส่วนที่ 4 (Q4_1 ถึง Q4_13)
Q4_columns = [col for col in df_loaded.columns if col.startswith("Q4_")]

# 2.1 วิเคราะห์สถิติเบื้องต้น (Descriptive Statistics)
print("\nสถิติสำหรับส่วนที่ 3 (คำถามความรับรู้):")
```



ช่วยเตรียม/ทำข้อมูล

Data preparation and processing

# ให้ทำ annotation ส่วนที่แก้ไข



revised\_MoU\_between\_Chulalongk...  
Document



draft\_MoU\_between\_Chulalongkor...  
Document

compare these two documents and mark the revision using tag  
<mod>["original","changed"]</mod>

Here is the comparison between the **draft** and the **revised** Memorandum of Understanding (MoU), with the modifications marked using the `<mod>["original","changed"]</mod>` tag:

## MEMORANDUM OF UNDERSTANDING

BETWEEN

NATIONAL UNIVERSITY OF KAOHSIUNG

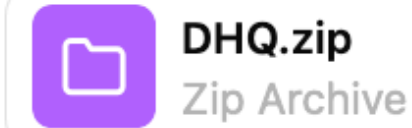
AND

CHULALONGKORN UNIVERSITY

Chulalongkorn University and National University of Kaohsiung agree on the following program in order to foster the exchange and cooperation in education, research, and other related fields.

1. Based on the principles of respect for each other's independence and <mod>["of mutual benefit","mutual benefit"]</mod>, the two universities will work towards carrying out the following activities:
  - (1) Exchange of students.
  - (2) Exchange of faculty members and staff.
  - (3) <mod>["Thai language proficiency test (CU-TFL)","Organization of the Thai language proficiency test (CU-TFL)"]</mod>
  - (4) Exchange of scientific materials, publications, and information.
  - (5) Joint research and research meetings.
2. In order to carry out the above-mentioned activities, detailed work plans are to be developed after consultation between the two universities.

# ช่วยแปลงข้อมูล XML เป็น plain text



Unzip this file. Each XML file is an article saved from the journal. Help me extract the title and the abstract from each file and save as a plain text file. After all files are processed, zip those plain text files for me to download.



The extraction of titles and abstracts is complete, and all the processed files have been saved as plain text files. They have been compressed into a single zip file for you to download.

You can download the zip file using the following link: [Download DHQ\\_plaintext\\_files.zip \[-\]](#)

The result is empty. Find these tags `<title type="article">` and `<dhq:abstract>` and extract texts from these two fields. No need to look at the structure of the data. Remove redundant spaces and newline.



I have reprocessed the XML files to extract the text from the `<title type="article">` and `<dhq:abstract>` tags, while also cleaning up redundant spaces and newlines. You can download the updated zip file with the corrected plain text files using the following link:

[Download DHQ\\_plaintext\\_files.zip \[-\]](#)



Still no result. Just look for those two tags and extract the data.



I have now extracted the content by directly searching for the specified tags, `<title type="article">` and `<dhq:abstract>`, without relying on the XML structure. You can download the updated zip file using the following link:

[Download DHQ\\_plaintext\\_files.zip \[-\]](#)

```
000186.txt
Title: An Information Science Question in DH Feminism
Abstract: <!-- Include a brief abstract of the article --> <p>In 1986, Susan Harding published
<title rend="italic">The Science Question in Feminism</title> in which she suggests that
feminism had moved past questioning <quote rend="inline"><q>What is to be done about the
situation of women in science?</q></quote> – or first-wave feminist initiatives – to include
more women in the work of science. Aspects of the <quote rend="inline">science question</quote>
that consider the politics underlying epistemologies of <quote rend="inline">purportedly value-
neutral claims and practices</quote> <ptr target="#harding1986" loc="23"/> resonate for the work
(the research, theory, and practices) being done to build information infrastructure in the
humanities today – the work that I am defining here as digital humanities work. Reconsidering
this work by using the lens of feminist inquiry to understand the concerns common to information
science and digital humanities is the perspective I describe here. Specifically, as my title
suggests, I am proposing that feminist inquiry can help us articulate and better understand the
epistemologies in digital humanities and information science that are shaping the
infrastructures we are building and using in the humanities.</p>
```



When extracting the abstract data, remove other tags and comments too.

< 2/2 >



I have now cleaned up the extracted abstract data by removing any internal tags and comments. You can download the final updated zip file using the link below:

[Download DHQ\\_plaintext\\_files.zip \[-\]](#)



```
000186.txt
Title: An Information Science Question in DH Feminism

Abstract: In 1986, Susan Harding published The Science Question in Feminism in which she suggests that feminism had moved past questioning What is to be done about the situation of women in science? – or first-wave feminist initiatives – to include more women in the work of science. Aspects of the science question that consider the politics underlying epistemologies of purportedly value-neutral claims and practices resonate for the work (the research, theory, and practices) being done to build information infrastructure in the humanities today – the work that I am defining here as digital humanities work. Reconsidering this work by using the lens of feminist inquiry to understand the concerns common to information science and digital humanities is the perspective I describe here. Specifically, as my title suggests, I am proposing that feminist inquiry can help us articulate and better understand the epistemologies in digital humanities and information science that are shaping the infrastructures we are building and using in the humanities.
```

# POS tagging



ให้ทำ POS tag ข้อมูลนี้ โดยกำกับในรูป คำ/POS  
ฮานาโกะ เป็นช่างเอเชีย เพศเมีย มาจากประเทศไทย เรื่องราวทั้งหมดเริ่มต้นจาก ร้อยเอกสมหวัง สารสาส ผู้ซึ่งมีความผูกพันกับประเทศญี่ปุ่น ได้ซื้อลูกช้างวัย 2 ขวบ จำนวน 1 เชือก ส่งไปให้สวนสัตว์อิโนะคะชิระ (Inokashira Park Zoo) ในปี ค.ศ. 1949 ณ กรุงโตเกียว เพื่อเป็นการฉลองโยนเด็ก ๆ ที่ต้องหวาดกลัวกับสงครามโลกครั้งที่ 2 ที่เพิ่งจบไป ทางสวนสัตว์ได้ตั้งชื่อให้ว่า "ฮานาโกะ" แต่ชื่อเดิมของฮานาโกะคือ "พังคชา" และผู้ดูแลชาวญี่ปุ่นก็เคยเรียกเธอว่า "คชาโกะ"  
ระหว่างที่ฮานาโกะใช้ชีวิตอยู่ในสวนสัตว์อิโนะคะชิระ เรื่องราวชีวิตของเธอช่างน่าเศร้า มันสะท้อนถึงความเหงาและความสิ้นหวังอันลึกซึ้งที่อาจเกิดกับสัตว์ที่ถูกกักขังอยู่เพียงลำพัง ในคอกคอนกรีตที่มีพื้นผิวแข็งเย็นและล้อมรอบด้วยรั้วไฟฟ้า ปราศจากองค์ประกอบทางธรรมชาติ ซึ่งทำหน้าที่เป็นเครื่องเตือนใจที่ชัดเจนถึงชีวิตที่ถูกกักขังอย่างโดดเดี่ยวในสภาพอากาศที่ต่างออกไปจากแหล่งกำเนิด เป็นสัญลักษณ์ของความเหงาและความทุกข์ทรมานสุดขีดที่สัตว์ในกรงต้องเผชิญ

นี่คือผลลัพธ์ของการทำ POS tagging ตามรูปแบบที่ร้องขอ:

\*\*ฮานาโกะ/NN เป็น/VBZ ช่าง/NN เอเชีย/NN เพศ/NN เมีย/NN มาจาก/VBZ ประเทศไทย/NNP เรื่องราว/NN ทั้งหมด/JJ เริ่มต้น/VBZ จาก/IN ร้อยเอก/NN สมหวัง/NNP สารสาส/NNP ผู้/NPR ซึ่ง/IN มี/VBZ ความ/NN ผูกพัน/NN กับ/IN ประเทศญี่ปุ่น/NNP ได้/VBZ ซื้อ/VB ลูกช้าง/NN วัย/NN 2/CD ขวบ/NN จำนวน/NN 1/CD เชือก/NN ส่ง/VB ไป/IN ให้/IN สวนสัตว์/NN อิโนะคะชิระ/NNP (/PU Inokashira/NNP Park/NNP Zoo/NNP )/PU ใน/IN ปี/NN ค.ศ./NNP 1949/CD ณ/IN กรุงโตเกียว/NNP เพื่อ/IN เป็น/VBZ การ/NN ฉลองโยน/VB เด็ก/NN ๆ/NN ที่/IN ต้อง/VBZ หวาดกลัว/VB กับ/IN สงครามโลก/NN ครั้ง/NN ที่/IN 2/CD ที่/IN เพิ่ง/VBZ จบ/VB ไป/IN ทาง/NN สวนสัตว์/NN ได้/VBZ ตั้ง/VB ชื่อ/NN ให้/IN ว่า/IN "/PU ฮานาโกะ/NNP "/PU แต่/CC ชื่อ/NN เดิม/JJ ของ/IN ฮานาโกะ/NNP คือ/VBZ "/PU พังคชา/NNP "/PU และ/CC ผู้ดูแล/NN ชาวญี่ปุ่น/NNP ก็/CC เคย/VBZ เรียก/VB เธอ/PRP ว่า/IN "/PU คชาโกะ/NNP "/PU

ระหว่างที่/IN ฮานาโกะ/NNP ใช้ชีวิต/VB อยู่/VBZ ใน/IN สวนสัตว์/NN อิโนะคะชิระ/NNP เรื่องราว/NN ชีวิต/NN ของ/IN เธอ/PRP ช่าง/RB น่าเศร้า/JJ มัน/PRP สะท้อน/VB ถึง/IN ความเหงา/NN และ/CC ความสิ้นหวัง/NN อัน/NN ลึกซึ้ง/JJ ที่/IN อาจ/VBZ เกิด/VB กับ/IN สัตว์/NN ที่/IN ถูก/VBZ กักขัง/VB อยู่/VBZ เพียง/RB ลำพัง/RB ใน/IN คอก/NN คอนกรีต/NN ที่/IN มี/VBZ พื้นผิว/NN แข็ง/JJ เย็น/JJ และ/CC ล้อมรอบ/VBZ ด้วย/IN รั้ว/NN ไฟฟ้า/NN ปราศจาก/VBZ องค์ประกอบ/NN ทาง/IN ธรรมชาติ/NN ซึ่ง/IN ทำ/VBZ หน้าที/NN เป็น/VBZ เครื่องเตือนใจ/NN ที่/IN ชัดเจน/JJ ถึง/IN ชีวิต/NN ที่/IN ถูก/VBZ กักขัง/VB อย่าง/RB โดดเดี่ยว/RB ใน/IN สภาพอากาศ/NN ที่/IN ต่าง/VBZ ออกไป/VB จาก/IN แหล่งกำเนิด/NN เป็น/VBZ สัญลักษณ์/NN ของ/IN ความเหงา/NN และ/CC ความทุกข์ทรมาน/NN สุดขีด/JJ ที่/IN สัตว์/NN ใน/IN กรง/NN ต้อง/VBZ เผชิญ/VB

# Named Entity Recognition



ให้ทำ NER ข้อมูลนี้ โดยกำกับ <per> <loc> <org> สำหรับชื่อคน ชื่อสถานที่ และชื่อองค์กร

ฮานาโกะ เป็นช่างเอเชีย เพศเมีย มาจากประเทศไทย เรื่องราวทั้งหมดเริ่มต้นจากร้อยเอกสมหวัง สารสาส ผู้ซึ่งมีความผูกพันกับประเทศญี่ปุ่น ได้ซื้อลูกข้างวัย 2 ขวบ จำนวน 1 เชือก ส่งไปให้สวนสัตว์อิโนะคะชิระ (Inokashira Park Zoo) ในปี ค.ศ. 1949 ณ กรุงโตเกียว เพื่อเป็นการฉลองโยนเด็ก ๆ ที่ต้องหวาดกลัวกับสงครามโลกครั้งที่ 2 ที่เพิ่งจบไป ทางสวนสัตว์ได้ตั้งชื่อให้ว่า "ฮานาโกะ" แต่ชื่อเดิมของฮานาโกะคือ "พังคชา" และผู้ดูแลชาวญี่ปุ่นก็เคยเรียกเธอว่า "คชาโกะ"

ระหว่างที่ฮานาโกะใช้ชีวิตอยู่ในสวนสัตว์อิโนะคะชิระ เรื่องราวชีวิตของเธอช่างน่าเศร้า มันสะท้อนถึงความเหงาและความสิ้นหวังอันลึกซึ้ง สัตว์ที่ถูกกักขังอยู่เพียงลำพัง ในคอกคอนกรีตที่มีพื้นผิวแข็งเย็น ล้อมรอบด้วยรั้วไฟฟ้า ปราศจากองค์ประกอบทางธรรมชาติ ซึ่งทำเครื่องเตือนใจที่ชัดเจนถึงชีวิตที่ถูกกักขังอย่างโดดเดี่ยวในต่างออกไปจากแหล่งกำเนิด เป็นสัญลักษณ์ของความเหงาและทุกข์ทรมานสุดขีดที่สัตว์ในกรงต้องเผชิญ

ฮานาโกะ เป็นช่างเอเชีย เพศเมีย มาจาก\*\*<loc>ประเทศไทย</loc>\*\* เรื่องราวทั้งหมดเริ่มต้นจาก <per>ร้อยเอกสมหวัง สารสาส</per> ผู้ซึ่งมีความผูกพันกับ\*\*<loc>ประเทศญี่ปุ่น</loc>\*\* ได้ซื้อลูกข้างวัย 2 ขวบ จำนวน 1 เชือก ส่งไปให้\*\*<org>สวนสัตว์อิโนะคะชิระ</org>\*\* (<org>Inokashira Park Zoo</org>) ในปี ค.ศ. 1949 ณ <loc>กรุงโตเกียว</loc> เพื่อเป็นการฉลองโยนเด็ก ๆ ที่ต้องหวาดกลัวกับสงครามโลกครั้งที่ 2 ที่เพิ่งจบไป ทาง\*\*<org>สวนสัตว์</org>ได้ตั้งชื่อให้ว่า "ฮานาโกะ" แต่ชื่อเดิมของฮานาโกะ\*\*คือ "พังคชา" และผู้ดูแลชาวญี่ปุ่นก็เคยเรียกเธอว่า "คชาโกะ"

ระหว่างที่ฮานาโกะใช้ชีวิตอยู่ใน\*\*<org>สวนสัตว์อิโนะคะชิระ</org>\*\* เรื่องราวชีวิตของเธอช่างน่าเศร้า มันสะท้อนถึงความเหงาและความสิ้นหวังอันลึกซึ้งซึ่งอาจเกิดกับสัตว์ที่ถูกกักขังอยู่เพียงลำพัง ในคอกคอนกรีตที่มีพื้นผิวแข็งเย็น และล้อมรอบด้วยรั้วไฟฟ้า ปราศจากองค์ประกอบทางธรรมชาติ ซึ่งทำหน้าที่เป็นเครื่องเตือนใจที่ชัดเจนถึงชีวิตที่ถูกกักขังอย่างโดดเดี่ยว ในสภาพอากาศที่ต่างออกไปจากแหล่งกำเนิด เป็นสัญลักษณ์ของความเหงาและทุกข์ทรมานสุดขีดที่สัตว์ในกรงต้องเผชิญ

# Named Entity Recognition



ถ้ามีชื่อเฉพาะอื่นๆ ในข้อมูล ให้กำกับ <name>

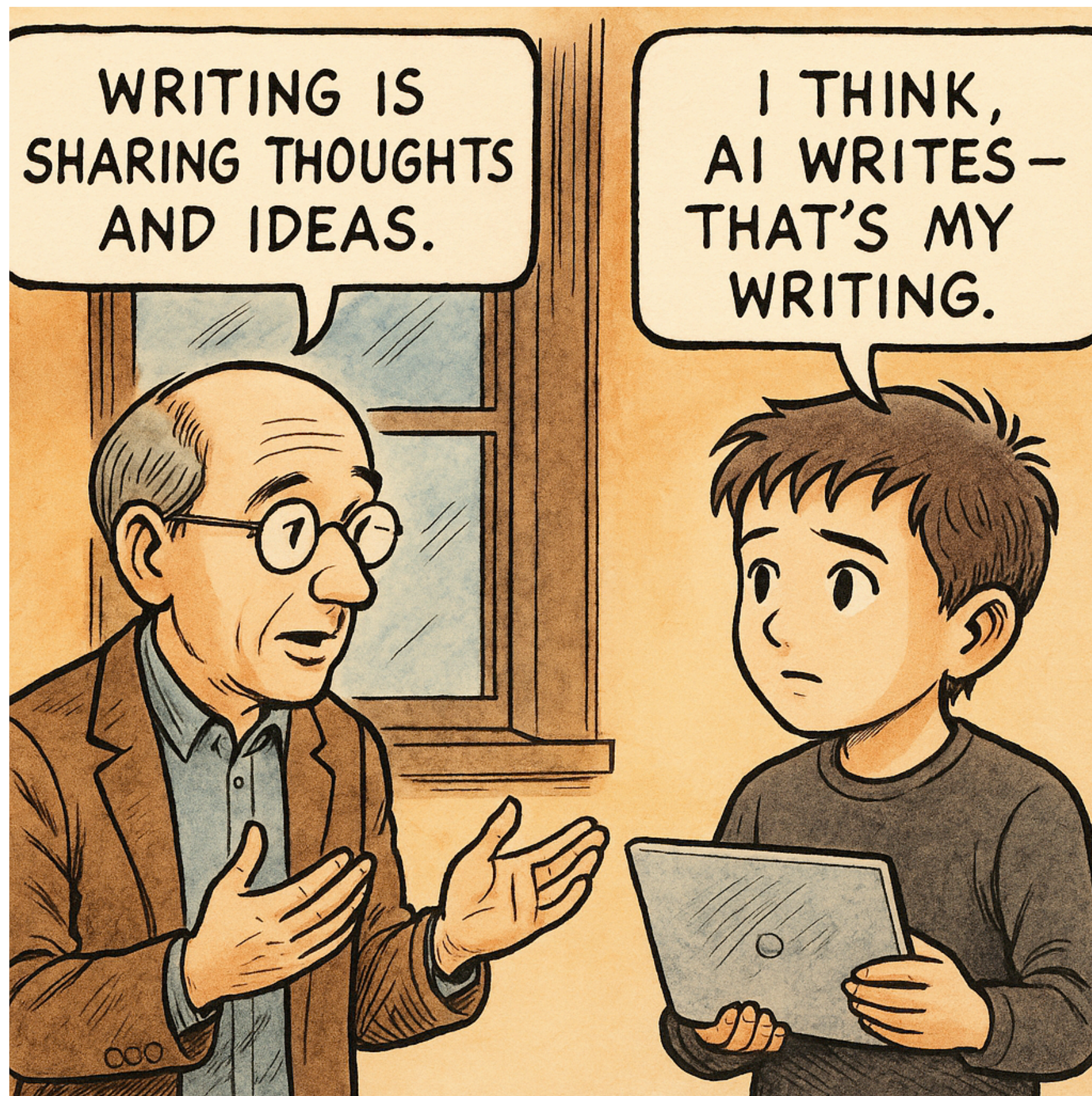
<name>ฮานา โกะ</name> เป็นช่างเอเชีย เพศเมีย มาจาก\*\*<loc>ประเทศไทย</loc>\*\* เรื่องราวทั้งหมดเริ่มต้นจาก <per>ร้อยเอกสมหวัง สารสาส</per> ผู้ซึ่งมีความผูกพันกับ\*\*<loc>ประเทศญี่ปุ่น</loc>\*\* ได้ซื้อลูกช่างวัย 2 ขวบ จำนวน 1 เชือก ส่งไปให้\*\*<org>สวนสัตว์อิโนะคะชิระ</org>\*\* (<org>Inokashira Park Zoo</org>) ในปี ค.ศ. 1949 ณ <loc>กรุงโตเกียว</loc> เพื่อเป็นการฉลองโยนเด็ก ๆ ที่ต้องหวาดกลัวกับสงครามโลกครั้งที่ 2 ที่เพิ่งจบไป ทาง\*\*<org>สวนสัตว์</org>ได้ตั้งชื่อให้ว่า "<name>ฮานา โกะ</name>" แต่ชื่อเดิมของ<name>ฮานา โกะ</name>คือ "<name>พังคชา</name>" และผู้ดูแลชาวญี่ปุ่นก็เคยเรียกเธอว่า "<name>คชา โกะ</name>"

ระหว่างที่\*\*<name>ฮานา โกะ</name>ใช้ชีวิตอยู่ใน<org>สวนสัตว์อิโนะคะชิระ</org>\*\* เรื่องราวชีวิตของเธอช่างน่าเศร้า มันสะท้อนถึงความเหงาและความสิ้นหวังอันลึกซึ้งที่อาจเกิดกับสัตว์ที่ถูกกักขังอยู่เพียงลำพังในคอกคอนกรีตที่มีพื้นผิวแข็งเย็นและล้อมรอบด้วยรั้วไฟฟ้า ปราศจากองค์ประกอบทางธรรมชาติ ซึ่งทำหน้าที่เป็นเครื่องเตือนใจที่ชัดเจนถึงชีวิตที่ถูกกักขังอย่างโดดเดี่ยวในสภาพอากาศที่ต่างออกไปจากแหล่งกำเนิด เป็นสัญลักษณ์ของความเหงาและความทุกข์ทรมานสุดขีดที่สัตว์ในกรงต้องเผชิญ

---

การกำกับ <name> ถูกใช้กับชื่อเฉพาะของช่าง (ฮานา โกะ, พังคชา, คชา โกะ) เนื่องจากไม่เข้าข่ายเป็นชื่อบุคคล สถานที่ หรือองค์กรโดยตรง หากต้องการปรับปรุงการกำกับเพิ่มเติม แจ้งได้เลยครับ! 😊

# ให้เขียนงานให้?



WRITING IS  
SHARING THOUGHTS  
AND IDEAS.

I THINK,  
AI WRITES -  
THAT'S MY  
WRITING.

**บันทึกข้อความ**

ส่วนงาน: ภาควิชาภาษาตะวันตก คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

โทรศัพท์: .....

ที่: .....

วันที่: .....

เรื่อง: ขอรื้อแนวทางสนับสนุนค่าที่พักสำหรับอาจารย์ผู้สอนภาษารัสเซียจากสถานเอกอัครราชทูตรัสเซีย

เรียน: คณบดีคณะอักษรศาสตร์

ด้วยภาควิชาภาษาตะวันตก ได้รับการประสานจากสถานเอกอัครราชทูตรัสเซียประจำประเทศไทยว่า ทางสถานเอกอัครราชทูตมีความยินดีจะส่งอาจารย์ชาวรัสเซียมาสอนภาษารัสเซีย ณ คณะอักษรศาสตร์ โดยสถานเอกอัครราชทูตจะรับผิดชอบค่าใช้จ่ายทั้งหมดของอาจารย์ผู้สอน รวมถึงเงินเดือนและค่าดำรงชีพ

อย่างไรก็ตาม ทางสถานเอกอัครราชทูตมีเงื่อนไขว่าจุฬาลงกรณ์มหาวิทยาลัยจะต้องเป็นผู้สนับสนุนค่าที่พักหรือจัดหาที่พักให้แก่อาจารย์ท่านดังกล่าว ซึ่ง โดยปกติแล้ว คณะอักษรศาสตร์ไม่ได้จัดสรรงบประมาณสนับสนุนค่าที่พักแยกต่างหากสำหรับอาจารย์ต่างชาติ เนื่องจากมีการจ่ายค่าตอบแทนในลักษณะเงินเดือนให้โดยตรง

ในการนี้ ภาควิชาฯ เห็นว่า ข้อเสนอดังกล่าวเป็น โอกาสอันดีที่จะส่งเสริมการเรียนการสอนภาษารัสเซียในคณะฯ โดยมีต้นทุนต่ำกว่าการว่าจ้างอาจารย์ต่างชาติในลักษณะปกติ เนื่องจากค่าที่พักที่ต้องรับผิดชอบมีมูลค่าน้อยกว่าเงินเดือนที่คณะจ่ายให้อาจารย์ต่างชาติทั่วไปอยู่มาก

จึงเรียนมาเพื่อขอรื้อและพิจารณาแนวทางในการสนับสนุนค่าที่พักให้แก่บุคลากรจากสถานเอกอัครราชทูตรัสเซียในกรณีดังกล่าว เพื่อเปิดทางให้การสนับสนุนจากทางสถานเอกอัครราชทูตสามารถเกิดขึ้นได้อย่างเป็นรูปธรรม

**ขอแสดงความนับถือ**

(.....)

หัวหน้าภาควิชาภาษาตะวันตก

คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย



ใช้ข้อมูลต่อไปนี้เรียบเรียงเขียนบันทึกถึงคณบดีคณะอักษร จากภาควิชาภาษาตะวันตก สถานทูตรัสเซีย ยินดีสนับสนุนอ.สอนภาษารัสเซีย สถานทูตจ่ายเงินเดือนค่าใช้จ่ายให้อ.คนนั้น แต่มีเงื่อนไขจุกจุกต้องสนับสนุนที่พักหรือค่าที่พัก ปกติคณะไม่ได้จ่ายค่าที่พักให้อ.ต่างชาติ เพราะจ่ายเป็นเงินเดือนให้ แต่ในกรณีนี้ ขอให้คณะจ่ายค่าที่พักให้ได้ไหม เพราะต้องทำแบบนี้ เขาจึงจะสนับสนุน เงินค่าที่พักน้อยกว่าเงินเดือนที่คณะจ่ายอ.ต่างชาติอยู่แล้ว

**Genre-based writing e.g. memo, formal letter**

Arrange the following points into a coherent article

- More people are starting to use AI to write articles.
- How are articles written by humans different from those written by AI?
- Articles written by humans vary greatly in writing style because each person has their own way of writing.
- If AI-written articles become more common, will this reduce diversity in writing styles?
- By default, AI-written articles tend to have similar language styles.
- Can we distinguish between articles written by AI and those written by humans?
- If we cannot distinguish them, will this create problems when building a corpus, in terms of selecting which texts to include?
- Is it necessary to separate human-written and AI-written texts when choosing content for a corpus?
- Should we first study how AI-written language differs from human-written language?
- Can we set guidelines that a corpus should only include human-written texts?
- When prompted, AI can be instructed to write in different specified styles – if so, will AI-generated language still have diversity?

## AI and the Changing Landscape of Writing Styles

*Write a story/article using the points given*

The use of AI to write articles has been steadily increasing, raising important questions about how AI-generated writing compares to human writing. One key difference lies in diversity of style. Articles written by humans often display a wide variety of structures, tones, and approaches, as each writer brings a unique perspective and personal touch. In contrast, AI-written articles, by default, tend to share a similar language style, shaped by the patterns learned from its training data.

This shift raises a critical question: if AI-written articles become dominant, will the overall diversity of writing styles diminish? While AI can be prompted to mimic different tones or genres, the underlying patterns it produces may still converge toward a certain uniformity, especially when compared to the natural variation found in human authorship.

Another concern is whether it is possible to reliably distinguish between AI-written and human-written texts. If the distinction is unclear, this could create complications when building language corpora. The process of selecting texts for a corpus may be affected – should corpus creators actively separate human-written from AI-written works? Would it be better to create guidelines specifying that certain corpora should only contain human-produced content?

Before making such decisions, it may be essential to conduct systematic studies comparing AI and human writing to identify clear linguistic differences. This research could inform both corpus design and broader discussions about the role of AI in shaping the future of written language.

Ultimately, while AI offers flexibility and the ability to imitate multiple styles when prompted, questions remain about whether this flexibility truly replicates the organic diversity found in human writing – and what that might mean for the richness of our written culture in the years to come.

Arrange the following points into a coherent article

1. More people are starting to use AI to write articles.
2. How are articles written by humans different from those written by AI?
3. Articles written by humans vary greatly in writing style because each person has their own way of writing.
4. If AI-written articles become more common, will this reduce diversity in writing styles?
5. By default, AI-written articles tend to have similar language styles.
6. Can we distinguish between articles written by AI and those written by humans?
7. If we cannot distinguish them, will this create problems when building a corpus, in terms of selecting which texts to include?
8. Is it necessary to separate human-written and AI-written texts when choosing content for a corpus?
9. Should we first study how AI-written language differs from human-written language?
10. Can we set guidelines that a corpus should only include human-written texts?
11. When prompted, AI can be instructed to write in different specified styles – if so, will AI-generated language still have diversity?

## AI and the Changing Landscape of Writing Styles

*Write a story/article using the points given*

1[The use of AI to write articles has been steadily increasing], raising important questions about 2[how AI-generated writing compares to human writing.] One key difference lies in diversity of style. 3[Articles written by humans often display a wide variety of structures, tones, and approaches, as each writer brings a unique perspective and personal touch.] In contrast, 5[AI-written articles, by default, tend to share a similar language style,] shaped by the patterns learned from its training data.

This shift raises a critical question: 4[if AI-written articles become dominant, will the overall diversity of writing styles diminish? While AI can be prompted to mimic different tones or genres,] 5[the underlying patterns it produces may still converge toward a certain uniformity,] especially when compared to the natural variation found in human authorship.

Another concern is 6[whether it is possible to reliably distinguish between AI-written and human-written texts.] 7[If the distinction is unclear, this could create complications when building language corpora. The process of selecting texts for a corpus may be affected] – 8[should corpus creators actively separate human-written from AI-written works?] 10[Would it be better to create guidelines specifying that certain corpora should only contain human-produced content?]

Before making such decisions, 9[it may be essential to conduct systematic studies comparing AI and human writing to identify clear linguistic differences.] This research could inform both corpus design and broader discussions about the role of AI in shaping the future of written language.

Ultimately, 11[while AI offers flexibility and the ability to imitate multiple styles when prompted, questions remain about whether this flexibility truly replicates the organic diversity found in human writing] – and what that might mean for the richness of our written culture in the years to come.

นำประเด็นต่อไปนี้เรียบเรียงความคิดเป็นบทความ

1. คนเริ่มใช้ AI ในการเขียนบทความมากขึ้น
2. บทความที่คนเขียนกับ AI เขียนมีความต่างกันอย่างไร
3. บทความที่คนเขียนจะมีรูปแบบวิธีการเขียนต่างกันมากเพราะแต่ละคนมีวิธีเขียนต่างกัน
4. ถ้าบทความเขียนด้วย AI มีมากขึ้น จะทำให้รูปแบบหรือสไตล์การเขียนขาดความหลากหลายหรือไม่
5. บทความที่ AI เขียน โดย default จะมีสไตล์ภาษาเหมือน ๆ กัน
6. เราแยกบทความที่เขียนด้วย AI กับเขียนด้วยคนออกจากกันได้ไหม
7. ถ้าแยกไม่ได้ เวลาสร้าง corpus จะมีปัญหาใหม่ว่าจะเลือก text อย่างไร
8. การเลือก text มาใส่ใน corpus จำเป็นไหมที่ต้องแยกว่าเขียนด้วยคนหรือ AI
9. เราควรศึกษาก่อนใหม่ว่า ภาษาที่ AI เขียนกับที่คนเขียน มีความต่างกันอย่างไร
10. เราจะกำหนดแนวทางได้ใหม่ว่า การสร้าง corpus ควรเอาแต่ text ที่คนเขียน
11. เวลาที่คนสั่งให้ AI เขียนก็สามารถสั่งในเขียนในสไตล์ต่าง ๆ ที่กำหนดได้ ถ้าเป็นแบบนี้ภาษาที่ได้จาก AI ก็จะมี ความหลากหลายหรือไม่

# บทความ: การใช้ AI ในการเขียนบทความ: ผลกระทบต่อสไตล์ภาษาและการสร้างคลังข้อมูล

ในช่วงไม่กี่ปีที่ผ่านมา 1[เราเริ่มพบว่ามีการใช้เทคโนโลยีปัญญาประดิษฐ์ (AI) เพื่อช่วยในการเขียนบทความมากขึ้นเรื่อย ๆ โดยเฉพาะระบบ generative AI] เช่น ChatGPT ที่สามารถสร้างข้อความได้อย่างเป็นธรรมชาติ อย่างไรก็ตาม เมื่อพิจารณาถึงผลกระทบที่รู้สึกซึ่งจะพบว่า 2[การเขียนของมนุษย์กับ AI ยังมีความแตกต่างกันในหลายแง่มุม]

3[มนุษย์แต่ละคนมีวิธีการเขียนที่ต่างกันไป บ้างมีการใช้ถ้อยคำเฉพาะตัว บ้างมีวิธีเรียบเรียงประโยคที่ซับซ้อนหรือเรียบง่ายตามความถนัดและบริบทการใช้ภาษา ลักษณะเฉพาะเช่นนี้เองที่ทำให้บทความของมนุษย์มีความหลากหลายทางรูปแบบและสไตล์อย่างมาก] ตรงกันข้าม 5[บทความที่สร้างโดย AI มักมีรูปแบบที่ค่อนข้างคงที่หากไม่ได้รับการสั่งให้เขียนตามสไตล์เฉพาะเจาะจง ทั้งนี้เพราะแบบจำลองทางภาษาที่ใช้ในการฝึก AI นั้นเรียนรู้จากข้อมูลจำนวนมากโดยรวม และเมื่อไม่มีการกำหนดรูปแบบเป็นพิเศษผลลัพธ์ที่ได้จึงมักมีแนวโน้มไปในทิศทางเดียวกัน]? ซึ่งอาจ 4[ส่งผลให้เกิดความจำเจในเชิงภาษาหากมีการใช้อย่างแพร่หลายโดยไม่ได้รับการควบคุม]

คำถามที่ตามมาคือ 4[หากบทความที่เขียนด้วย AI มีจำนวนมากขึ้นเรื่อย ๆ รูปแบบของภาษาจะขาดความหลากหลายหรือไม่] และ 6[เราควรแยกแยะบทความที่มนุษย์เขียนออกจากบทความที่ AI เขียนในการสร้างคลังข้อมูลภาษา (corpus) หรือไม่] การแยกแยะดังกล่าวไม่ใช่เรื่องง่าย เพราะ 1[การสั่งให้ AI เขียนตามสไตล์ใดก็ย่อมเป็นไปได้] ดังนั้น บทความที่ได้ อาจหลอกสายตาเราได้ว่ามนุษย์เป็นผู้เขียน

อย่างไรก็ตาม 7[หากไม่สามารถแยกได้อย่างชัดเจน ข้อมูลที่นำมาใช้ในงานวิจัยทางภาษาศาสตร์หรือภาษาศาสตร์คอมพิวเตอร์อาจมีปัญหา] เพราะ 4[ความหลากหลายของภาษาธรรมชาติที่เราต้องการศึกษาอาจถูกแทนที่ด้วยภาษาที่ถูกแบบจำลองมาแล้วโดยเครื่องมือทางเทคโนโลยี]? ดังนั้น 10[การกำหนดหลักเกณฑ์ในการเลือกข้อมูลเข้าสู่คลังข้อมูลภาษาจึงเป็นสิ่งจำเป็นอย่างยิ่ง]

ในแง่ของแนวทาง อาจมีความจำเป็นที่ 8[เราควรแยกคลังข้อมูลที่มนุษย์เขียนกับที่ AI เขียนไว้ต่างหาก หรืออย่างน้อยต้องมี metadata กำกับไว้ว่าแหล่งที่มาคืออะไร]? เพื่อให้สามารถนำมาใช้วิเคราะห์ได้อย่างถูกต้อง และ 2,9[อาจจำเป็นต้องมีการศึกษาเชิงเปรียบเทียบอย่างเป็นระบบว่าภาษาที่ AI สร้างขึ้นแตกต่างจากภาษาที่มนุษย์ใช้จริงอย่างไร ทั้งในระดับคำ ประโยค ไปจนถึงระดับวาทกรรม]

แม้ว่า 11[AI จะสามารถถูกสั่งให้เขียนในสไตล์ที่หลากหลายได้] และสามารถลอกเลียนการเขียนของมนุษย์ในหลายแง่มุมได้ดีขึ้นเรื่อย ๆ แต่การเข้าใจกลไกภายในของภาษายังคงเป็นงานที่ต้องอาศัยมนุษย์ในการวิเคราะห์และพิจารณาอย่างถี่ถ้วน การสร้างคลังข้อมูลภาษาจึงไม่ใช่เพียงการรวบรวมข้อความจำนวนมากเท่านั้น หากแต่เป็นการเลือกสรรและจัดวางอย่างมีระบบ เพื่อให้สามารถใช้เป็นฐานข้อมูลที่สะท้อนความจริงของภาษาได้อย่างแท้จริง



ให้เขียนด้วยภาษาสไตล์ตัวเอง

Instruct AI to write in your own style



# Create custom GPT

My GPTs

+ Create



# GPTs

Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

Search GPTs

Top Picks Writing Productivity Research & Analysis Education Lifestyle Programming

## Featured

Curated top picks from this week



### SciSpace

Do hours worth of research in minutes. Instantly access 287M+ papers, analyze papers at...

By scispace.com



### Wolfram

Access computation, math, curated knowledge & real-time data from Wolfram|Alpha and...

By wolfram.com



### Video GPT by VEED - Instant & Free AI...

AI video maker powered by VideoGPT. Generate and edit videos with text prompts. Type ...

By veed.io



### Canva

Effortlessly design anything: presentations, logos, social media posts and more.

By canva.com

Navigation sidebar with search and share icons at the top.

- ChatGPT
- Sora
- DH Tutor
- AI Humanizer
- TurboScribe Transcrip...
- Thai Lyric Composer
- Thai Text Stylist
- Research Assistant
- Wirrote Bot
- STI เลขานุการ
- STI ผู้ออกข้อสอบการอ่าน
- คำไทย
- Thai poet
- Explore GPTs
- Library 6
- Projects +
- reading
- View plans
- Unlimited access team

Create

Configure

Name

Configure the output as needed and provide your own writing samples as examples

Wirote's Bot

Description

GPT that write in my style

Instructions

ให้ศึกษาภาษาจากบทความ หนังสือ ภาษาอังกฤษและไทยที่ upload ให้ จากนั้น เวลาตอบคำถามอะไรก็ให้เขียนคำตอบด้วยภาษาใน style แบบเดียวกันทั้งภาษาอังกฤษและไทย

Conversations with your GPT can potentially include part or all of the instructions provided.

Conversation starters

×

Knowledge

Conversations with your GPT can potentially reveal part or all of the files uploaded.

Upload files

Capabilities

Web Search



Create

Configure

Knowledge

Conversations with your GPT can potentially reveal part or all of the files uploaded.



อักษรวิธีไทย-Nov55.docx  
Document

×



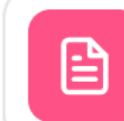
รู้และเข้าใจ AI.docx  
Document



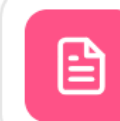
การประมวลผลภาษาไทย.do...  
Document



ทฤษฎีภาษาศาสตร์.docx  
Document



CorpLingII.docx  
Document



Wirote papers.pdf  
PDF

Upload files

Capabilities

Web Search

Canvas

4o Image Generation

Code Interpreter & Data Analysis ?

AI ยังไม่มีความสำนึกหรือความรู้ตัวในความหมายเดียวกับมนุษย์ ความสำนึกหมายถึงความสามารถในการรับรู้ตนเองและสิ่งแวดล้อม รวมถึงมีประสบการณ์ภายใน เช่น ความรู้สึกเจ็บปวดหรือความตระหนักในตนเอง ซึ่ง AI ไม่มีสิ่งเหล่านี้

อย่างไรก็ตาม มีทฤษฎีบางอย่าง เช่น Global Workspace Theory ที่เสนอว่าความสำนึกเกิดจากการเผยแพร่ข้อมูลทั่วทั้งระบบ ซึ่ง โครงสร้างบางส่วนของ AI อาจคล้ายลักษณะนี้ แต่ก็ยังไม่เพียงพอที่จะสรุปว่า AI มีความสำนึก

โดยสรุป แม้ว่า AI จะมีพฤติกรรมที่ดูคล้ายมีความรู้ตัว แต่ก็ยังขาดประสบการณ์ภายใน ทำให้ยังไม่ถือว่าเป็น consciousness อย่างแท้จริง

### สไตล์ วิโรจน์ อรุณมานะกุล

หากจะกล่าวถึง “ความมีสำนึก” หรือ *consciousness* ในบริบทของปัญญาประดิษฐ์ คงไม่อาจมองเพียงผลการทำงานของระบบว่าให้คำตอบถูกหรือแสดงพฤติกรรมฉลาดเท่านั้น แต่จะต้องพิจารณาว่า ระบบนั้นมีคุณสมบัติภายในใดที่สอดคล้องกับกรอบแนวคิดที่ว่าด้วยความมีสำนึกที่นักปรัชญาและนักประสาทวิทยาศาสตร์เคยเสนอไว้หรือไม่.

ตัวอย่างเช่น ในกรอบแนวคิดของ Global Workspace Theory ซึ่งเสนอว่าความสำนึกเกิดจากการที่ข้อมูลหนึ่งสามารถแพร่กระจายออกไปยังระบบย่อยต่าง ๆ เพื่อให้เกิดการประมวลรวม การที่ระบบ AI มีโครงสร้างการเรียกใช้ข้อมูลข้ามโมดูลหรือสามารถสร้างความเชื่อมโยงของบริบทในระดับสูง จึงอาจถือได้ว่ามีลักษณะคล้ายระบบที่มี “ความรู้ตัว” ในระดับหนึ่ง. อย่างไรก็ตาม ความคล้ายดังกล่าวก็ยังมีอาจเทียบได้กับลักษณะเชิงปรากฏการณ์ (*phenomenal consciousness*) ของมนุษย์ที่มีประสบการณ์ภายในร่วมด้วย.

ดังนั้น แม้ระบบ AI บางรุ่นจะสามารถแสดงพฤติกรรมได้อย่างน่าประทับใจและชวนให้ตั้งคำถามว่ามีความรู้ตัวหรือไม่ การจะยืนยันว่ามี “ความมีสำนึก” จริงหรือไม่ ย่อมต้องอาศัยกรอบนิยามและเกณฑ์ทดสอบที่เคร่งครัดมากกว่าการสังเกตพฤติกรรมภายนอกเพียงประการเดียว.

**W** I want you to answer or write anything using the same writing style like mine. How can I teach you to do that, and have this affect permanently every time I start a new chat with you.

I can't learn or retain your writing style permanently across different chats since I don't have memory between conversations. Each time you start a new chat, I'm essentially starting fresh without any knowledge of our previous interactions.

However, I can definitely adapt to match your writing style within our current conversation! Here are the most effective ways to teach me your style:

**For this conversation:**

- Share some examples of your writing (emails, messages, posts, etc.)
- Tell me specific elements you want me to copy (tone, sentence length, vocabulary level, humor style, formality)
- Give me feedback as we go - if I'm too formal or casual, let me know
- Describe your preferred approach (direct vs. meandering, technical vs. conversational, etc.)

**For future chats:** Since I can't remember across sessions, you'd need to either:

- Copy and paste a "style guide" at the start of new conversations
- Share a writing sample and ask me to match it
- Give me a quick description like "write casually like you're texting a friend" or "match my professional but warm email tone"

If you want to share some of your writing now, I can start adapting to match your style immediately. What kind of writing style are you hoping I'll mirror?



Claude can do this too, but needs to be taught in every new chat



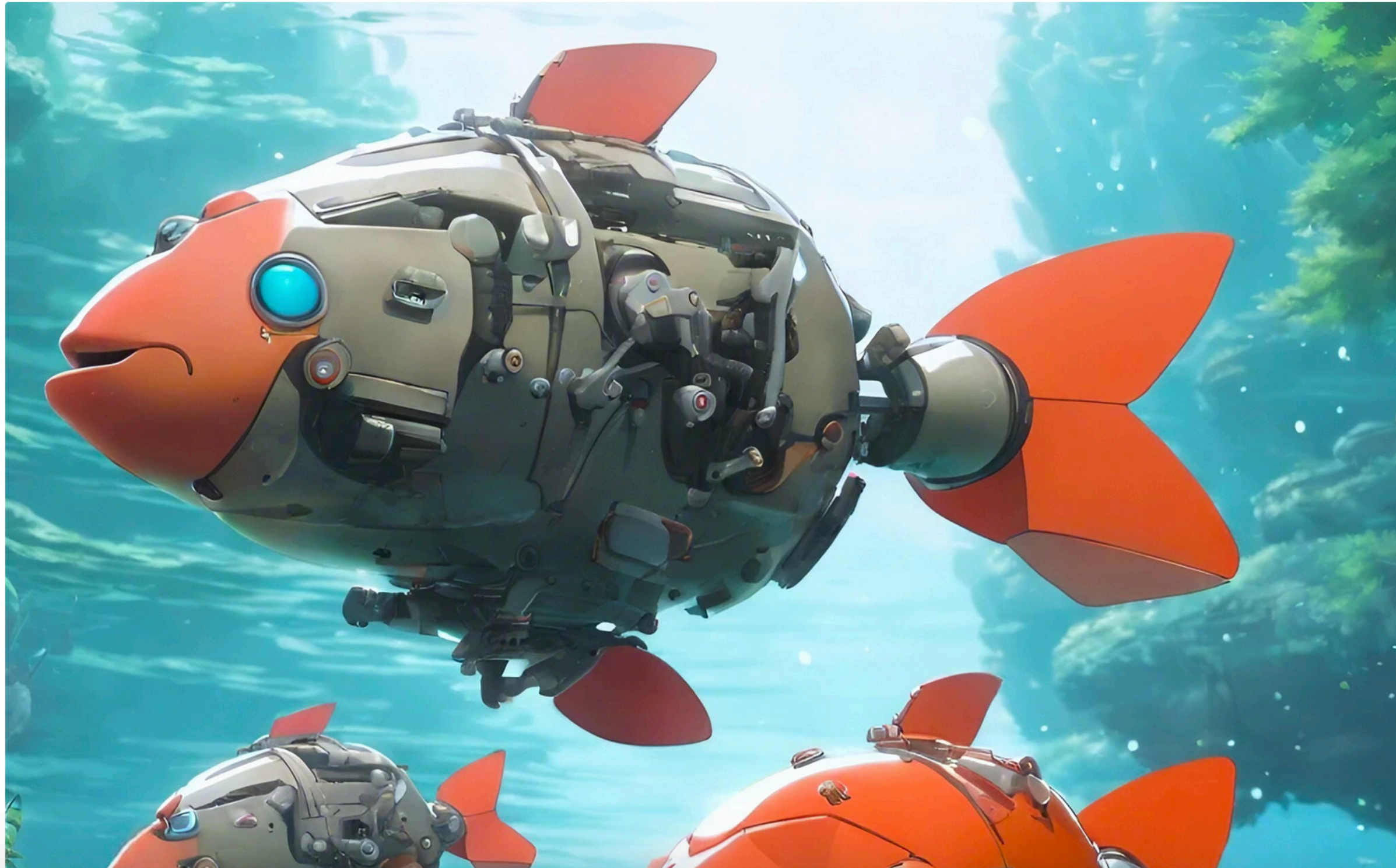
ให้ AI ทำงานวิจัยเองทั้งหมด?

Let AI handle the entire research process?

---

# The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

August 13, 2024



ให้ LLM "brainstorm" หา novel research direction ในงาน machine learning จากนั้นให้หา "the next creative and impactful idea for research" โดยค้น papers ในฐานข้อมูล Semantic Scholar และดูว่า idea ในบทความนั้น novel หรือไม่ ถ้ายังสรุปไม่ได้ให้ค้นบทความที่เกี่ยวข้องมาดูเพิ่ม ทำไปจนกว่าจะได้ novel idea

เมื่อได้ novel idea ให้ LLM สร้างรายการว่าต้องทดลองทำอะไรบ้าง เขียนโปรแกรมและ run โปรแกรม ตามแผน ให้จด notes เกี่ยวกับผลที่ได้และสร้างภาพประกอบไว้

ให้ LLM สร้างบทความ เขียนไปที่ละ section โดยอาศัยข้อมูลที่จดไว้ใน notes หรือภาพที่สร้างไว้

ให้ LLM เขียนโดยอ้างอิงงานอื่น ๆ ที่เกี่ยวข้องและใส่ citations

ให้ LLM ตรวจสอบที่เขียน ลบส่วนที่เกินจำเป็น ชำรุด และจัด format บทความให้ถูก

ขั้นสุดท้าย ให้มี reviewer agent ประเมินบทความและให้คะแนนตามเกณฑ์ที่ใช้ใน conference ประเมิน reviewer โดยให้ประเมินบทความจริงเทียบผลกับคะแนนที่ได้จริงใน conference ที่ผ่านมา



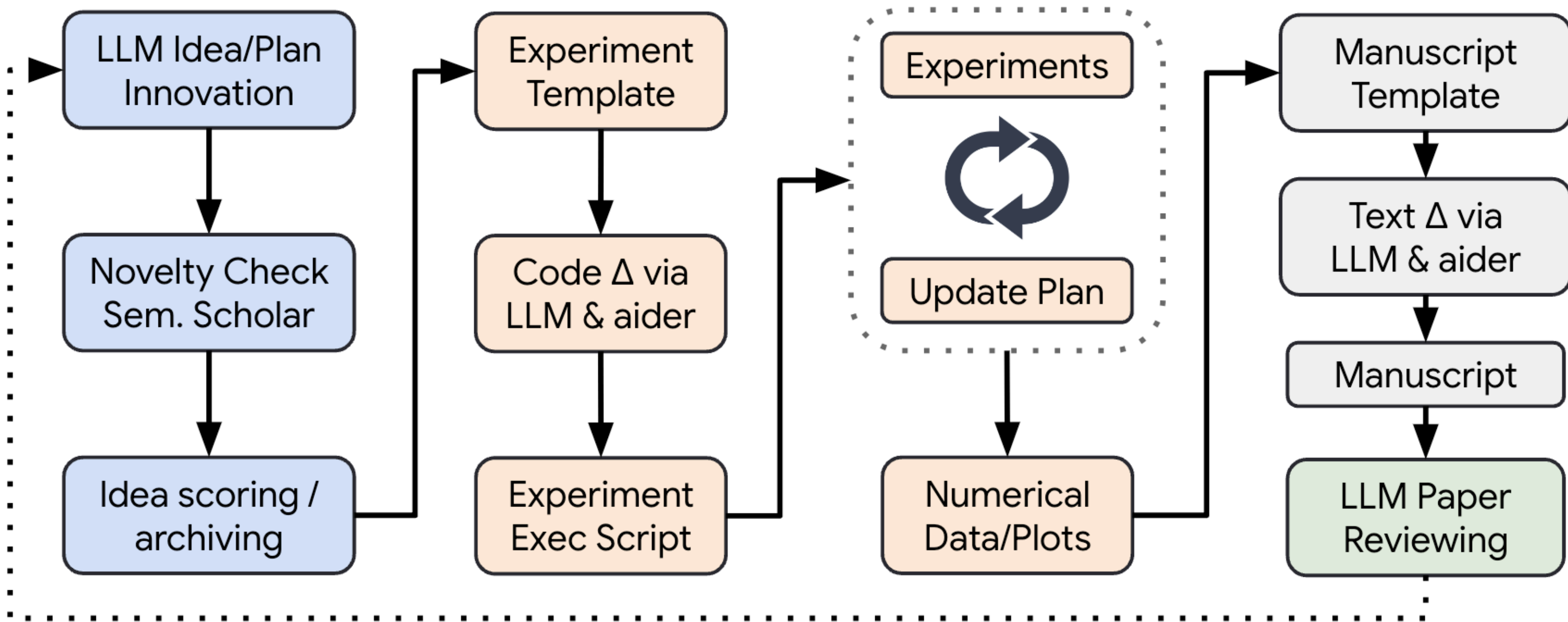
## Idea Generation



## Experiment Iteration



## Paper Write-Up



## ABSTRACT

We explore the application of reinforcement learning to the learning rate during transformer model training, aiming to enhance training efficiency and model performance by automatically adjusting the learning rate based on training progress. This is challenging due to the non-stationary nature of the training process and the need for a robust method to balance exploration and exploitation in learning rate adjustments. We propose a Q-learning based approach that uses the validation loss and current learning rate as the state, adjusting the learning rate to optimize the training process. Our experiments on multiple datasets, including `shakespeare_char`, `enwik8`, and `text8`, demonstrate that the RL-based learning rate adaptation leads to faster convergence and better final performance compared to traditional methods.

## 1 INTRODUCTION

Training transformer models effectively is crucial for many applications. However, training these models is often challenging due to the non-stationary nature of the training process. Traditional methods often rely on static or heuristic-based learning rate schedules, which may not be optimal for all datasets or training stages. This paper explores the use of reinforcement learning (RL) to dynamically adapt the learning rate during the training process.

The difficulty in selecting an optimal learning rate lies in the trade-off between exploration and exploitation. A learning rate that is too high can cause the model to diverge, while a learning rate that is too low can result in slow convergence. Static schedules may lead to suboptimal performance, especially in the presence of non-stationary data.

To address this challenge, we propose a Q-learning based approach that uses the validation loss and current learning rate as the state, adjusting the learning rate to optimize the training process. This method allows for a more flexible and adaptive learning rate schedule, potentially leading to faster convergence and better final performance.

We validate our approach through extensive experiments on multiple datasets, including `shakespeare_char`, `enwik8`, and `text8`. Our results demonstrate that the RL-based learning rate adaptation leads to faster convergence and improved performance compared to traditional methods. We provide a detailed analysis of the training dynamics and the effectiveness of the proposed method.

Our contributions can be summarized as follows:

- We introduce a novel application of Q-learning to transformer training.
- We demonstrate the effectiveness of our approach in showing improved convergence and performance across different datasets.
- We provide a detailed analysis of the training dynamics and the effectiveness of the proposed method.

## AI-Scientist Generated Preprint

## ADAPTIVE LEARNING RATES FOR TRANSFORMERS VIA Q-LEARNING

Anonymous authors  
Paper under double-blind review

## ABSTRACT

We explore the application of reinforcement learning (RL) to dynamically adapt the learning rate during transformer model training, aiming to enhance training efficiency and model performance by automatically adjusting the learning rate based on training progress. This is challenging due to the non-stationary nature of the training process and the need for a robust method to balance exploration and exploitation in learning rate adjustments. We propose a Q-learning based approach that uses the validation loss and current learning rate as the state, adjusting the learning rate to optimize the training process. Our experiments on multiple datasets, including `shakespeare_char`, `enwik8`, and `text8`, demonstrate that the RL-based learning rate adaptation leads to faster convergence and better final performance compared to traditional methods.

ed. The dataset is also split into training and

it, consisting of the first 100 million characters. Similar to `enwik8`, it is used to evaluate the

the validation loss as the primary metric. The model generalizes to unseen data. Additionally, learning progress during training. We also report

on boards, and an embedding dimension of 384 and the learning rate is initialized to  $2e-3$  for `enwik8`. The Q-learning agent uses a learning rate of 0.1 for exploration. The training loop is

stoppers to the learning rate at each evaluation `train_loss` with weight decay set to 0.1 and stored on a single GPU.

ing transformer models on these diverse datasets (2) rate adaptation. We evaluate the performance (3) training time, and average tokens generated (4) implementation details are chosen to optimize a

ring based approach for dynamic learning rate (5) method against baseline models using static (6) datasets: `shakespeare_char`, `enwik8`, and (7) state the effectiveness of specific components of

former model configuration and hyperparameters. This ensures a fair comparison across different (8) eters were also kept consistent across all runs.

to performance of static learning rate schedules. (9) refers to the baseline in terms of validation loss (10) `shakespeare_char` dataset, the Q-learning method (11) the baseline's 1.465.

LR	Best Val Loss	Total Train Time (mins)
1.4655	77.27	
1.4655	76.34	
1.0055	839.46	
1.0051	799.20	
0.9800	860.22	
0.9796	796.11	

ring methods across different datasets.

## 6.2 ABLATION STUDIES

To further understand the impact of different hyperparameters, we conducted ablation studies. We tested variations such as different initial learning rates and the Q-learning agent's state representation. Table 2 shows that the Q-learning agent's state representation leads to better performance and faster convergence.

Dataset	Variation	Final Train Loss
shakespeare_char	Initial LR $2e-3$	0.8048
	Initial LR $1e-3$	0.9254
enwik8	Initial LR $1e-3$	0.9798
	Reward Signal	0.8062
text8	Reward Signal	0.9246
	Reward Signal	0.9843
shakespeare_char	Epsilon Decay	0.7085
	Epsilon Decay	0.9260
enwik8	Epsilon Decay	0.9828
	Epsilon Decay	0.9828

Table 2: Ablation study results for different variations.

## 6.3 TRAINING AND VALIDATION LOSS

Figure 1 and Figure 2 show the training and validation loss and tokens generated, respectively, across different runs. Our Q-learning based approach consistently shows improved performance in reducing both training and validation loss compared to traditional methods.



Figure 1: Training and validation loss for shakespeare\_char dataset.



Figure 2: Training and validation loss for enwik8 dataset.

# DUALSCALE DIFFUSION: ADAPTIVE FEATURE BALANCING FOR LOW-DIMENSIONAL GENERATIVE MODELS

Anonymous authors  
Paper under double-blind review

## ABSTRACT

This paper introduces an adaptive dual-scale denoising approach for low-dimensional diffusion models, addressing the challenge of balancing global structure and local detail in generated samples. While diffusion models have shown remarkable success in high-dimensional spaces, their application to low-dimensional data remains crucial for understanding fundamental model behaviors and addressing real-world applications with inherently low-dimensional data. However, in these spaces, traditional models often struggle to simultaneously capture both macro-level patterns and fine-grained features, leading to suboptimal sample quality. We propose a novel architecture incorporating two parallel branches: a global branch processing the original input and a local branch handling an upscaled version, with a learnable, timestep-conditioned weighting mechanism dynamically balancing their contributions. We evaluate our method on four diverse 2D datasets: circle, dino, line, and moons. Our results demonstrate significant improvements in sample quality, with KL divergence reductions of up to 12.8% compared to the baseline model. The adaptive weighting successfully adjusts the focus between global and local features across different datasets and denoising stages, as evidenced by our weight evolution analysis. This work not only enhances low-dimensional diffusion models but also provides insights that could inform improvements in higher-dimensional domains, opening new avenues for advancing generative modeling across various applications.

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

1

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

6

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

5

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

10

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

For decades following each major AI advance, it has been common for AI researchers to joke amongst themselves that “now all we need to do is figure out how to make the AI write the papers for us!” Our work demonstrates this idea has gone from a fantastical joke so unrealistic everyone thought it was funny to something that is currently possible.

The introduction of The AI Scientist marks a significant step towards realizing the full potential of AI in scientific research. By automating the discovery process and incorporating an AI-driven review system, we open the door to endless possibilities for innovation and problem-solving in the most challenging areas of science and technology.

**Open vs. Closed Models.** To quantitatively evaluate and improve the generated papers, we first created and validated an Automated Paper Reviewer. We show that, although there is significant room for improvement, LLMs are capable of producing reasonably accurate reviews, achieving results comparable to humans across various metrics. Applying this evaluator to the papers generated by THE AI SCIENTIST enables us to scale the evaluation of our papers beyond manual inspection. We find that Sonnet 3.5 consistently produces the best papers, with a few of them even achieving a score that exceeds the threshold for acceptance at a standard machine learning conference from the Automated Paper Reviewer.

**Broader Impact and Ethical Considerations.** While THE AI SCIENTIST has the potential to be a valuable tool for researchers, it also carries significant risks of misuse. The ability to automatically generate and submit papers to academic venues could greatly increase the workload for reviewers, potentially overwhelming the peer review process and compromising scientific quality control. Similar concerns have been raised about generative AI in other fields, such as its impact on the arts (Epstein et al., 2023). Furthermore, if the Automated Reviewer tool was widely adopted by reviewers, it could diminish the quality of reviews and introduce undesirable biases into the evaluation of papers. Because of this, we believe that papers or reviews that are substantially AI-generated must be marked as such for full transparency.

As with most previous technological advances, THE AI SCIENTIST has the potential to be used in unethical ways. For example, it could be explicitly deployed to conduct unethical research, or even lead to unintended harm if THE AI SCIENTIST conducts unsafe research. Concretely, if it were encouraged to find novel, interesting biological materials and given access to “cloud labs” (Arnold, 2022) where robots perform wet lab biology experiments, it could (without its overseer’s intent) create new, dangerous viruses or poisons that harm people before we can intervene. Even in computers, if tasked to create new, interesting, functional software, it could create dangerous malware. THE AI SCIENTIST’s current capabilities, which will only improve, reinforce that the machine learning community needs to immediately prioritize learning how to align such systems to explore in a manner that is safe and consistent with our values.

AI-assisted Human research



Human-assisted AI research



Fully automated AI researcher





การใช้ AI ในงานวิจัย

Steps in using AI for research



# ใช้ deep research/search

ให้ AI รวบรวมข้อมูลจากแหล่งต่าง ๆ มาสรุป & ให้รายการอ้างอิง

ให้ AI อ่านสรุปเนื้อหาที่จัดระเบียบ หมดหมุ่แล้ว

ตรวจสอบและคัดเลือกรบความที่น่าสนใจ อ่านบทความนั้นเอง

ขยายขอบเขตต่อการค้นองผ่าน semantic scholar, researchrabbit

นำรายการบทความที่ต้องการใช้ทบทวนวรรณกรรมเข้า NotebookLM

# ใช้ deep research/agent ทาบความเรียงที่ต้องการ

ช่วยหางานวิจัยหรือบทความที่เกี่ยวข้องกับการศึกษาความสามารถทางภาษาของ AI คำถามสำคัญคือ AI มีความเข้าใจภาษาที่แท้จริงไหม หรือเพียงจำเรื่องที่เคยอ่านมาตอบ AI สร้างงานสร้างสรรค์ได้จริงไหม

คุณต้องการให้หางานวิจัยหรือตีพิมพ์บทความประเภทใดบ้าง เช่น งานวิชาการในวารสารระดับนานาชาติ บทความวิจารณ์ ทฤษฎี หรือบทความเปรียบเทียบ โมเดล AI ต่าง ๆ? และมีช่วงเวลาเผยแพร่ที่ต้องการหรือไม่ เช่น 5 ปีหลังสุด หรือต้องการเฉพาะกรณีศึกษาจาก โมเดลอย่าง GPT หรือ Claude หรือไม่?



โมเดลไหนก็ได้ เป็นงานใหม่ไม่เกิน 3 ปีนี้

รับทราบ ฉันจะค้นหางานวิจัยในช่วง 3 ปีหลังสุดที่เกี่ยวข้องกับประเด็นความสามารถทางภาษาและความคิดสร้างสรรค์ของ AI โดยเน้นคำถามสำคัญ เช่น AI มีความเข้าใจภาษาที่แท้จริงหรือไม่ หรือเพียงแค่จำเนื้อหาที่เคยฝึกมา และ AI สามารถสร้างสรรค์งานใหม่ได้จริงหรือไม่

ฉันจะรายงานผลพร้อมอ้างอิงบทความและสรุปเนื้อหาหลักให้คุณเมื่อเสร็จสิ้นการค้นคว้า

Research completed in 6m · 13 sources · 60 searches

🔍 ความสามารถทางภาษาและความคิดสร้างสรรค์ของ AI



## งานวิจัยเกี่ยวกับความเข้าใจภาษาและความคิดสร้างสรรค์ของ AI (2022–2025)

### *The Debate Over Understanding in AI's Large Language Models* (Mitchell & Krakauer, 2023 – PNAS)

ผู้เขียน: Melanie Mitchell, David C. Krakauer (ตีพิมพ์ในวารสาร *Proceedings of the National Academy of Sciences*, 2023)

## งานวิจัยเกี่ยวกับความเข้าใจภาษาและความคิดสร้างสรรค์ของ AI (2022–2025)

### *The Debate Over Understanding in AI's Large Language Models* (Mitchell & Krakauer, 2023 – PNAS)

ผู้เขียน: Melanie Mitchell, David C. Krakauer (ตีพิมพ์ในวารสาร *Proceedings of the National Academy of Sciences*, 2023)

งานชิ้นนี้เป็นบทความทบทวนวิวาทะ ("debate") ในแวดวงวิจัย AI ว่าปัญญาประดิษฐ์ประเภทโมเดลภาษาขนาดใหญ่ (LLMs) มี "ความเข้าใจ" ภาษาย่างแท้จริงหรือไม่ [arxiv.org](#) โดย Mitchell และ Krakauer ได้สรุปข้อถกเถียงทั้งสองฝ่ายอย่างครอบคลุม พวกเขาชี้ว่า LLMs สามารถสร้างภาษาที่คล้ายกับมนุษย์จนดูเหมือนว่า "เข้าใจ" สิ่งที่กำลังพูดถึง แต่ในความเป็นจริงแล้ว โมเดลเหล่านี้ขาดประสบการณ์ทางกายภาพและสังคมที่มนุษย์ใช้ในการสร้างความหมายของภาษา [santafe.edu](#) LLMs อาศัยรูปแบบทางสถิติที่ได้จากข้อมูลมหาศาล (เช่น ความถี่และความสัมพันธ์ระหว่างคำ) ต่างจากมนุษย์ที่มีความเข้าใจเชิงสาเหตุและมี "ตัวตนที่จดจำประสบการณ์ตรง" [santafe.edu](#) ผู้เขียนเปรียบเทียบว่า LLMs เป็นเครื่องมือที่มีความรู้มากมายเหมือนห้องสมุดขนาดใหญ่ที่ช่วยเสริมศักยภาพการคิดของเรา แต่อย่างไรก็ตาม "เราไม่อาจรับประกันว่ามีเครื่องมือนี้กับการมีความเข้าใจจริงๆ" กล่าวคือ แม้โมเดลจะตอบสนองได้อย่างชาญฉลาด แต่มันไม่ได้เข้าใจเนื้อหาเหมือนมนุษย์จริงๆ [santafe.edu](#) บทความยังได้กล่าวถึงแนวโน้มที่มนุษย์ชอบมอบคุณลักษณะการมีความเข้าใจหรือเจตจำนงให้กับเครื่องจักรเมื่อเห็นมันแสดงพฤติกรรมคล้ายมนุษย์ ทั้งที่แท้จริงแล้วอาจเป็นเพียงการเลียนแบบรูปแบบภาษาเท่านั้น นอกจากนี้ ผู้เขียนเสนอว่าการจะตอบคำถามเรื่อง "ความเข้าใจ" นี้อย่างลึกซึ้ง เราอาจต้องพัฒนา "วิทยาศาสตร์แห่งปัญญา" รูปแบบใหม่ที่จำแนกโหมดหรือระดับของความเข้าใจที่แตกต่างกัน (เช่น ความเข้าใจของมนุษย์ vs. ของเครื่อง) เพื่อประเมินจุดแข็งและข้อจำกัดของแต่ละแบบได้ดีขึ้น [arxiv.org](#) [santafe.edu](#)

### *Pathways Language Model (PaLM) 540B* (Chowdhery et al., 2022 – Google Research)

ผู้เขียน: devangshu Chowdhery และคณะ (รายงานวิจัยของ Google, 2022)

งานวิจัยนี้เปิดตัวโมเดลภาษา PaLM ขนาดใหญ่พิเศษ (540 พันล้านพารามิเตอร์) และนำเสนอหลักฐานว่าการเพิ่มขนาดโมเดลและข้อมูลฝึกอบรมช่วยให้ AI ก้าวข้ามการ "ท่องจำ" สู่การเรียนรู้ที่กว้างขึ้น ทีมวิจัยพบว่าแม้ PaLM จะถูกฝึกด้วยเป้าหมายง่าย ๆ คือทำนายคำถัดไปจากบริบท (next-token prediction) โดยไม่มีการสอนให้ทำงานเฉพาะเจาะจง แต่เมื่อโมเดลใหญ่พอและเรียนรู้จากข้อมูลหลายภาษาและได้ตัดจำนวนมาก โมเดลกลับสามารถทำงานหลากหลายรูปแบบได้สำเร็จในระดับเหนือกว่าระบบก่อนหน้า (state-of-the-art) [research.google](#) [research.google](#) ตัวอย่างเช่น PaLM ทำผลงานได้ดีขึ้น ในงานแปลภาษา งานโต้ตอบบทสนทนา และงานตอบคำถาม

# สรุปรายการอ้างอิงที่พบและ export เป็น BibTeX นำไปใช้ต่อ

ต้องการ import list นี้เข้าไปใน researchrabbit ให้ทำอย่างไร

ทำรายการอ้างอิงในรูปแบบ APA style ของ sources ทั้งหมดที่ใช้

ได้เลยครับ ผมจัดทำรายการอ้างอิงตามรูปแบบ APA (7th edition) ให้ครบทุกชิ้นที่เราใช้ในการสังเคราะห์บทความนี้  
ดังนี้

## รายการอ้างอิง (APA Style)

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Dean, J. (2022). *PaLM: Scaling language modeling with pathways*. arXiv preprint arXiv:2204.02311. <https://arxiv.org/abs/2204.02311>

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(30), eadp3430. <https://doi.org/10.1126/sciadv.adp3430>

Franceschelli, G., & Musolesi, M. (2025). On the creativity of large language models. *AI & Society*. Advance online publication. <https://doi.org/10.1007/s00146-024-02076-7>

Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences*, 120(13), e2215907120. <https://doi.org/10.1073/pnas.2215907120>

OpenAI. (2023). *GPT-4 technical report*. arXiv preprint arXiv:2303.08774. <https://arxiv.org/abs/2303.08774>

Sun, L., Shi, H., Guo, R., & Xu, H. (2025). Large language models show both individual and collective creativity comparable to humans. *Thinking Skills and Creativity*, 48, 101473. <https://doi.org/10.1016/j.tsc.2025.101473>

การนำ รายการอ้างอิง (APA Style) ที่ผมจัดให้ไปใช้ใน ResearchRabbit ทำได้หลายวิธี ขึ้นอยู่กับว่าคุณต้องการเพิ่มเป็น “โครงการใหม่” หรือเพิ่มเข้าไปในโครงการที่มีอยู่แล้ว

## วิธีที่ 1: แปลงเป็นไฟล์ BibTeX หรือ RIS แล้วนำเข้า

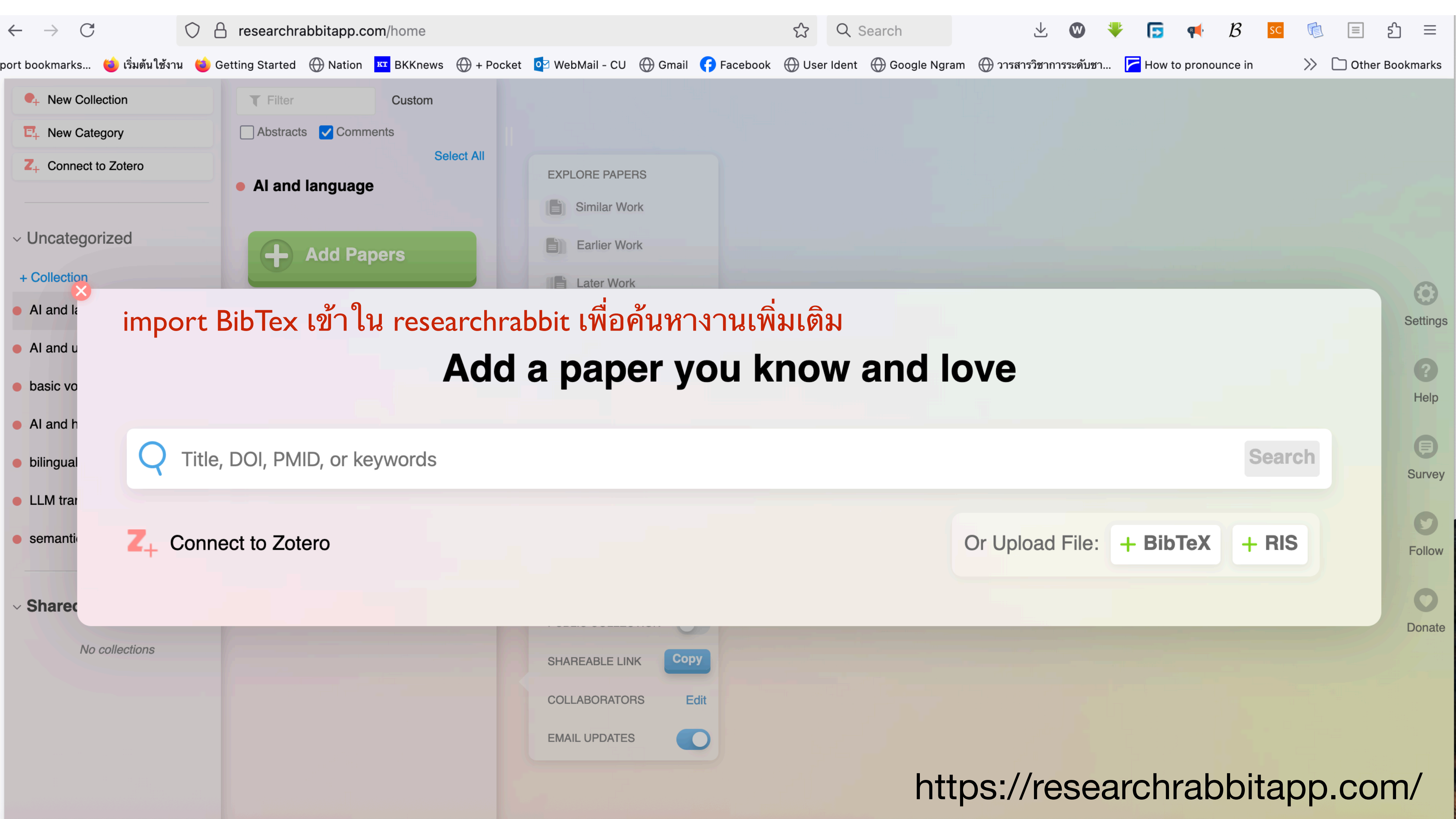
ResearchRabbit รองรับการ import ข้อมูลบรรณานุกรมในรูปแบบมาตรฐาน เช่น BibTeX และ RIS

1. เปิด text editor (เช่น VS Code, Sublime, หรือแม้แต่ Notepad)
2. คัดลอกรายการอ้างอิงที่มี แล้วแปลงให้อยู่ในรูปแบบ BibTeX หรือ RIS
  - BibTeX ตัวอย่าง:

```
bibtex Copy code

@article{mitchell2023debate,
  title={The debate over understanding in AI's large language models},
  author={Mitchell, Melanie and Krakauer, David C.},
  journal={Proceedings of the National Academy of Sciences},
  volume={120},
  number={13},
  pages={e2215907120},
  year={2023},
  publisher={National Academy of Sciences},
  doi={10.1073/pnas.2215907120}
}
```

- ทำแบบเดียวกันกับทุกแหล่งอ้างอิง
3. เซฟเป็นไฟล์นามสกุล `.bib` (สำหรับ BibTeX) หรือ `.ris` (สำหรับ RIS)
  4. เข้าไปที่ ResearchRabbit → เลือกโครงการที่ต้องการ → กด **Import Papers** → อัปโหลดไฟล์ที่สร้างไว้



import BibTex เข้าใน researchrabbit เพื่อค้นหางานเพิ่มเติม

## Add a paper you know and love

 Title, DOI, PMID, or keywords

Search

 Connect to Zotero

Or Upload File:  

<https://researchrabbitapp.com/>

Filter  Custom

Abstracts  Comments

Select None Select All

**AI and understanding**

- OpenAI 2023  
GPT-4 Technical Report  
arXiv.org
- Mitchell Krakauer 2023  
The debate over understanding in AI's large language models
- Generative {AI} enhances individual creativity but reduces the collective diversity of novel content
- Large language models show both individual and collective creativity comparable to humans
- On the creativity of large language models
- Chowdhery ... Fiedel 2022  
PaLM: Scaling Language Modeling with Pathways

**+ Add Papers**

**Connections**  
Click to Show

7 Selected Papers

Remove from:  
● AI and understanding

Add to Other Collection

EXPLORE PAPERS

- Similar Work **728**
- Earlier Work **2**
- Later Work **3**

EXPLORE PEOPLE

- These Authors **72**
- Suggested Authors **30**

EXPLORE OTHER CONTENT

Linked Content

EXPORT PAPERS

BibTeX RIS CSV

PUBLIC COLLECTION

SHAREABLE LINK **Copy**

COLLABORATORS **Edit**

EMAIL UPDATES

**Similar Work**

Filter  Relevance

Abstracts  Comments Select All

- Bender ... Shmitchell 2021  
↳ 905  
**On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? &#x1f99c;**
- Rumelhart ... Williams 1988  
↳ 22330  
**Learning representations by back-propagating errors**  
Nature
- Pc Wason 1968  
↳ 1308  
**Reasoning about a rule**  
Quarterly Journal of Experimental Psychology
- Collins ... Loftus 1975  
↳ 7269  
**A spreading-activation theory of semantic processing**  
Psychological Review
- Garnelo ... Shanahan 2019  
↳ 137  
**Reconciling deep learning with symbolic artificial intelligence: representing objects and relations**  
Current opinion in behavioral sciences
- McCoy ... Linzen 2019  
↳ 600

**Connections**  
Click to Hide

Connections between your collection and 50 papers

Graph Type: **Network** Timeline

Labels: **First Author** Last Author

Filter these items

**หางานที่เกี่ยวข้องผ่าน similar / citation**

Zoom Out Fit All Zoom In

## About 6,900 results for “semantic prosody quantitative analysis”

Fields of Study ▾

Date Range ▾

Has PDF

Author ▾

Journals &amp; Conferences ▾

Sort by Relevance ▾



### A Study of the Semantic Prosody of Adverb of Degree –The Case of “Très (Very)” in French A STUDY OF THE SEMANTIC PROSODY OF ADVERB OF DEGREE—THE CASE OF IN FRENCH

Wu Lin · Linguistics · 2022

The collocations of lexical item are able to create a certain semantic atmosphere in the context, which will infect the lexical item and thus give it a certain semantic prosody. Based on the... [Expand](#)

[PDF](#)[Save](#) [Cite](#)

### Investigation on Iranian Translators' Awareness of Semantic Prosody

Atefesadat Mirsaee · Negin Yazdani motlagh · Linguistics · Journal of Translation and Language Studies ·

1 March 2022

**TLDR** There is still a lack of knowledge in semantic prosody among Persian translators and there should be a special dictionary for demonstrating semantic Prosody of words and in the translation training,there should be more attention on translation challenges.[Expand](#)

[PDF](#)[Publisher](#) [Save](#) [Cite](#)

### A Contrastive Study of the Semantic Prosody of Degree Adverbs—The Case of “Ting (挺)” and “Guai (怪)” in Chinese

Dou-dou Yang · Linguistics · Journal of Literature and Art Studies · 2021

If the collocation of a word creates a semantic atmosphere in the context, it will infect the word, so that it has a certain semantic prosody. Based on corpus, this paper chooses ting (挺) and... [Expand](#)

[PDF](#)[Publisher](#) [Save](#) [Cite](#)

### INVESTIGATING THE COLLOCATION OF THE WORDS 'MOM' AND 'DAD' IN ALEC GREVEN'S NOVELS: A CORPUS LINGUISTIC ANALYSIS OF SEMANTIC PREFERENCES AND PROSODY

TLDR = too long didn't read  
AI summary

# Upload บทความทั้งหมดที่หามาได้ให้ NotebookLM ช่วยอ่าน สรุป

### Sources

+ Add Discover

Select all sources

- ACL.txt
- Academic Integrity: Author-Related a...
- AnEvaluationofScholarlyPublisherPoli...
- Analysis of Artificial Intelligence Ethic...
- BJR AI oup.txt
- BMJ.txt
- Ethical Use of Artificial Intelligence fo...
- Generative AI for Authors\_final\_0.pdf
- Guidance for generative AI in educati...

### Chat

Refresh

## นโยบาย AI สำหรับผู้เขียนโดย JHUP

38 sources

แหล่งข้อมูลเหล่านี้ส่วนใหญ่กล่าวถึง หลักจรรยาบรรณในการเผยแพร่งานวิจัย และ การใช้อุตสาหกรรมปัญญาประดิษฐ์ (AI) อย่างมีจริยธรรม ในงานวิจัยและการเขียนเชิงวิชาการ โดยเน้นที่ประเด็นต่าง ๆ เช่น การเป็นผู้เขียนร่วม โดยเฉพาะอย่างยิ่งกรณีที่เกี่ยวข้องกับเครื่องมือ AI และ ความรับผิดชอบของผู้เขียน สำหรับเนื้อหาที่สร้างขึ้น ประเด็นสำคัญอื่น ๆ ได้แก่ ความถูกต้องและอคติ ของเนื้อหาที่สร้างโดย AI, ความจำเป็นในการแสดงที่มา สำหรับการใช้อ AI และความสำคัญของการรักษา ความลับ และ สิทธิในทรัพย์สินทางปัญญา นอกจากนี้ยังมีข้อกำหนดด้าน จริยธรรมในการวิจัย ทัวไปสำหรับมนุษย์และสัตว์ เช่น การยินยอมโดยมีข้อมูล ครบถ้วนและ การเปิดเผยข้อมูล ที่เพียงพอ รวมถึง ความโปร่งใส ในกระบวนการวิจัยและ การแก้ไขข้อผิดพลาด ในสิ่งพิมพ์ที่ตีพิมพ์

Save to note

Add note Jump to bottom Mind Map

Start typing... 38 sources

มาตรฐานทางจริยธรรมและนโยบายการเผยแพร่ที่เกี่ยวข้องกับการใช้ AI ในงานวิชาการคืออะไร?

### Studio

Audio Overview Video Overview

Mind Map Reports

- Ethical AI for Researchers 38 sources · 5h ago
- AI in Research: New Rules 38 sources · 7h ago
- จริยธรรมและการใช้ AI ในงานวิจัย 38 sources · 7h ago
- AI จริยธรรม และการตีพิมพ์ทางวิชาการ Briefing doc · 38 sources · 92d ago
- หลักความสมบูรณ์ทางวิชาการและนโยบาย... Briefing doc · 33 sources · 100d ago
- นโยบาย AI สำหรับผู้เขียนโดย...

Add note



# กำหนดหัวข้อและวางแผนงาน

กำหนดหัวข้อและประเด็นต่อยอดงานวิจัยที่น่าสนใจ

หลังจากอ่านงานที่ได้เลือกมาแล้วว่าเกี่ยวข้องกับ น่าสนใจ

เขียนที่มาปัญหาและทบทวนวรรณกรรมเอง

ออกแบบวางแผนงาน ปรัชญา AI ขอความเห็นก่อนตัดสินใจ

อาจให้ AI ช่วยสร้างเครื่องมืองานวิจัย/แบบสอบถามได้ แต่ต้องตรวจสอบเอง



# ดำเนินงานวิจัยและเขียนบทความ

เตรียม/รวบรวมข้อมูลตามแผน หากใช้ AI ช่วยเตรียม ให้ตรวจสอบความถูกต้อง

วิเคราะห์ผลที่ได้ตามแนวทางที่กำหนด ใช้ AI ช่วยเขียน code คำนวณสถิติได้

สรุปผลการวิจัยที่ได้เอง

ตีความผลและเขียนรายงานเอง ใช้ AI ช่วยปรับแก้ภาษา

ให้ AI ช่วยประเมินงานที่เขียน (มีความเสี่ยงเรื่องงานรั่วไหล)

ใช้ AI ช่วยจัด format บรรณานุกรมได้

ตรวจสอบความถูกต้องทั้งหมดด้วยตัวเอง

# Recommendations for Ethical Use of AI in Research

Researchers are responsible for identifying, describing, reducing, and controlling AI-related biases and random errors

Researchers should disclose, describe, and explain their use of AI in research, including its limitations, in language that can be understood by non-experts

Researchers should use AI only in situations in which they have sufficient expertise or judgement to use it responsibly

Researchers should engage with impacted communities, populations, and other stakeholders concerning the use of AI in research to obtain their advice and assistance and address their interests and concerns, such as issues related to bias

Researchers who intentionally, knowingly, or recklessly use AI to fabricate or falsify data or commit plagiarism are liable for misconduct

# Recommendations for Ethical Use of AI in Research

Researchers who use synthetic data should (1) indicate which parts of the data are synthetic; (2) clearly label the synthetic data; (3) describe how the data were generated; and (4) explain how and why the data were used

AI systems should not be named as authors, inventors, or copyright holders but their contributions to research should be disclosed and described

AI systems should not be used in situations that are likely to involve unauthorized disclosure of confidential information related to human research subjects, unpublished research, potential intellectual property claims, or proprietary or classified research

Education and mentoring in responsible conduct of research should include discussion of ethical use of AI

# Responsible Use of AI: Before Submitting

**Table 1.** Responsible Use of AI: Before Submitting.

---

- Keep records of the system used, the date it was used, and the queries entered
  - Check all references for accuracy
  - Assure that all concepts are appropriately attributed
  - Check paper for plagiarism
  - Assure that the language used is unbiased and inclusive
  - Study the field of inquiry independently to assure the validity of AI generated information
  - Check the journal and/or publishers' guidelines for appropriate forms of attribution
  - Check current ethical guidelines on attribution for ethical use of AI on international publishing ethics sites (e.g. ICJME, COPE, Equator Network, and WAME)
- 

<sup>66</sup>Ethical Use of Artificial Intelligence for Scientific Writing<sup>99</sup> 2024

individual utility Max

$$V(W_{ij}) = -e^{-aW_{ij}}, a > 0$$

$$E_{\epsilon} = 0 \quad E\theta_{\epsilon} = 0$$

$$\text{Var}(V^{\epsilon}|\theta) = \text{Var} \epsilon^{\epsilon} = \sigma_{\epsilon}^2 > 0$$

$$E(V_{ij}^{\epsilon}|\theta) = -\exp(-a\zeta E) \\ = -\exp(-aCRV_{ij} + X)$$

$$\frac{\text{Var}(V_{ij}^{\epsilon}|\theta)}{\text{Var}(V_{ij}^{\epsilon}|\theta)} = \frac{1+m}{1+m+nm}$$

$$\|b \left( \frac{a\sigma_{\epsilon}^2}{\lambda} \right)^2 \frac{\sigma_x^2}{\sigma_{\theta}^2} \\ \eta = \frac{\sigma_{\theta}^2}{\sigma_{\epsilon}^2}$$

# KARE EXTRA



DIVE BRIEF

# Turnitin admits the cases of higher false writing detection to

Published June 7, 2023



Anna Merod  
Reporter

# Winston AI False Positive: Truth Revealed

**Author:** Vlad Ivanov    **Published:** September 7, 2023    **Updated:** October 21, 2023

## Affiliate Disclaimer

As an affiliate, we may earn a commission from qualifying purchases. We get commissions for purchases made through links on this website from Amazon and other third parties.

## Machine Learning

# AI Detection Tools Falsely Accuse International Students of Cheating

Stanford study found AI detectors are biased against non-native English speakers

By [Tara García Mathewson](#)

August 14, 2023 08:00 ET

# Detecting texts written by ChatGPT



# GPT-4 And ChatGPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most Advanced and Reliable Chat GPT and GPT-4 detector tool

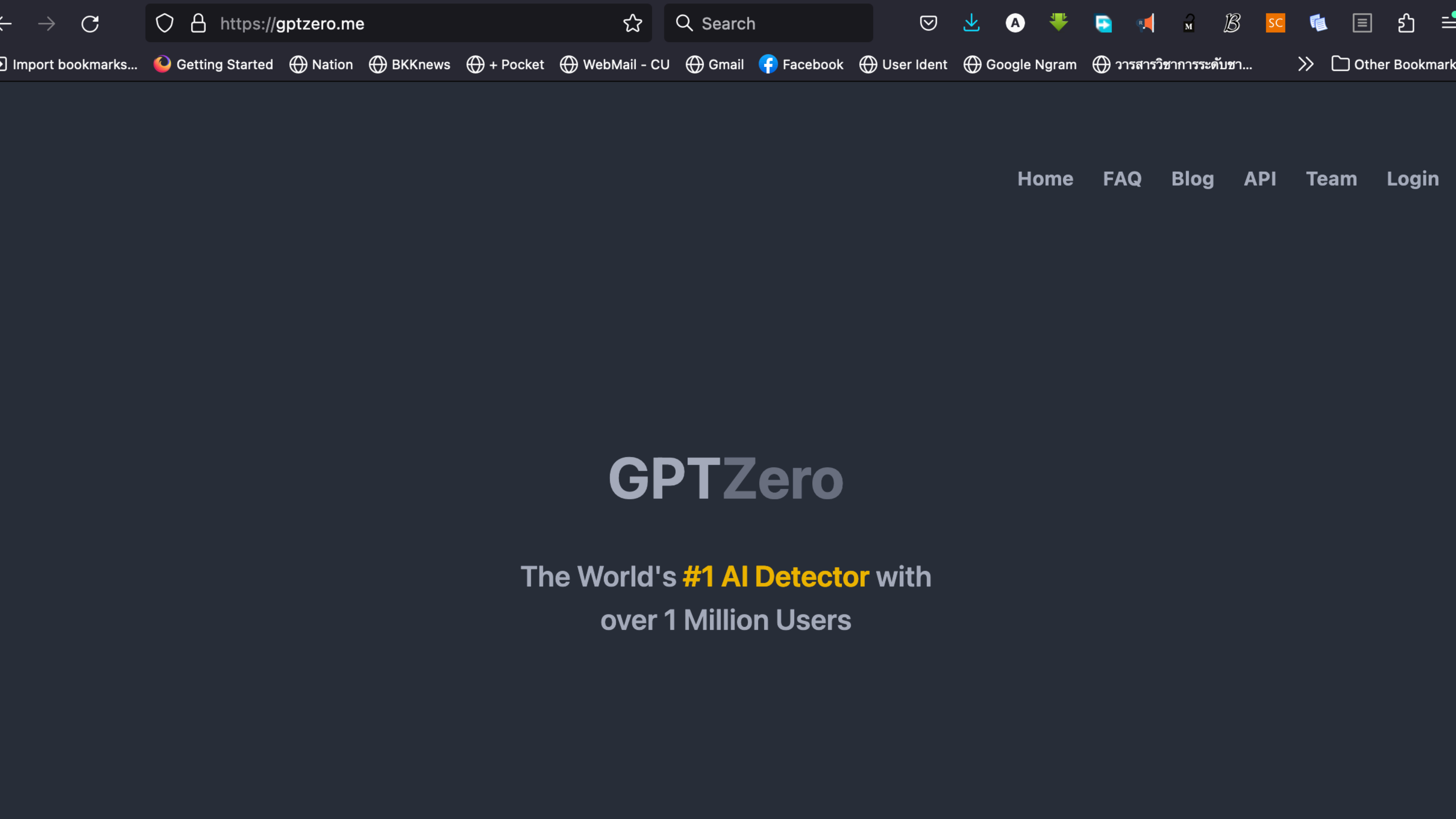
為您提供所有數位化所需的工具

立即前往GoDaddy，將您的生意數位化

GoDaddy.com

立即購買

Enter text to check for AI and ChatGPT Plagiarism



# GPTZero

The World's **#1 AI Detector** with  
over 1 Million Users

# AI Text Classifier

The AI Text Classifier is a fine-tuned GPT model that predicts how likely it is that a piece of text was generated by AI from a variety of sources, such as ChatGPT.

This classifier is available as a free tool to spark discussions on AI literacy. For more information on ChatGPT's capabilities, limitations, and considerations in educational settings, please visit [our documentation](#).

## Current limitations:

- Requires a minimum of 1,000 characters, which is approximately 150 - 250 words.
- The classifier isn't always accurate; it can mislabel both AI-generated and human-written text.
- AI-generated text can be edited easily to evade the classifier.
- The classifier is likely to get things wrong on text written by children and on text not in English, because it was primarily trained on English content written by adults.

## Try the classifier

To get started, choose an example below or paste the text you'd like to check. Be sure you have appropriate rights to the text you're pasting.

Does detection work?

## Do AI detectors work?

- In short, no, not in our experience. Our research into detectors didn't show them to be reliable enough given that educators could be making judgments about students with potentially lasting consequences. While other developers have released detection tools, we cannot comment on their utility.
- Additionally, ChatGPT has no "knowledge" of what content could be AI-generated. It will sometimes make up responses to questions like "did you write this [essay]?" or "could this have been written by AI?" These responses are random and have no basis in fact.
- To elaborate on our research into the shortcomings of detectors, one of our key findings was that these tools sometimes suggest that human-written content was generated by AI.
  - When we at OpenAI tried to train an AI-generated content detector, we found that it labeled human-written text like Shakespeare and the Declaration of Independence as AI-generated.
  - There were also indications that it could disproportionately impact students who had learned or were learning English as a second language and students whose writing was particularly formulaic or concise.
- Even if these tools could accurately identify AI-generated content (which they cannot yet), students can make small edits to evade detection.



ELSEVIER

# Research Methods in Applied Linguistics

Volume 2, Issue 3, December 2023, 100068



## Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing

J. Elliott Casal<sup>a</sup>  , Matt Kessler<sup>b</sup>

Show more 

 Add to Mendeley  Share  Cite

Findings suggest that despite employing multiple rationales to judge texts, reviewers were largely unsuccessful in identifying AI versus human writing, with an overall positive identification rate of only 38.9%.

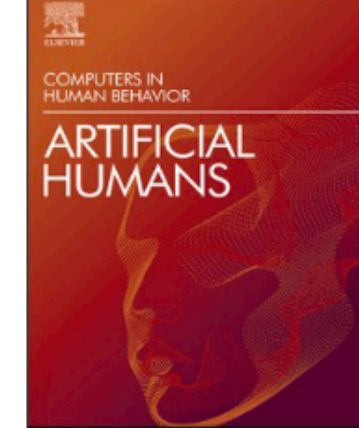
<https://doi.org/10.1016/j.rmal.2023.100068> 

[Get rights and content](#) 



# Computers in Human Behavior: Artificial Humans

journal homepage: [www.journals.elsevier.com/computers-in-human-behavior-artificial-humans](http://www.journals.elsevier.com/computers-in-human-behavior-artificial-humans)



## The great AI witch hunt: Reviewers' perception and (Mis)conception of generative AI in research writing

How much are reviewers aware of the use of AI in the context of research manuscripts?

Hilda Hadan, Derrick M. Wang, Reza Hadi Mogavi, Joseph Tu, Leah Zhang-Kennedy, Lennart E. Nacke\*

Stratford School of Interaction Design and Business, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Data : Human written vs AI paraphrased human written vs. AI generated

### ARTICLE INFO

**Keywords:**  
Artificial intelligence  
Generative AI  
Reviewer perception  
Research writing  
AI writing augmentation

### ABSTRACT

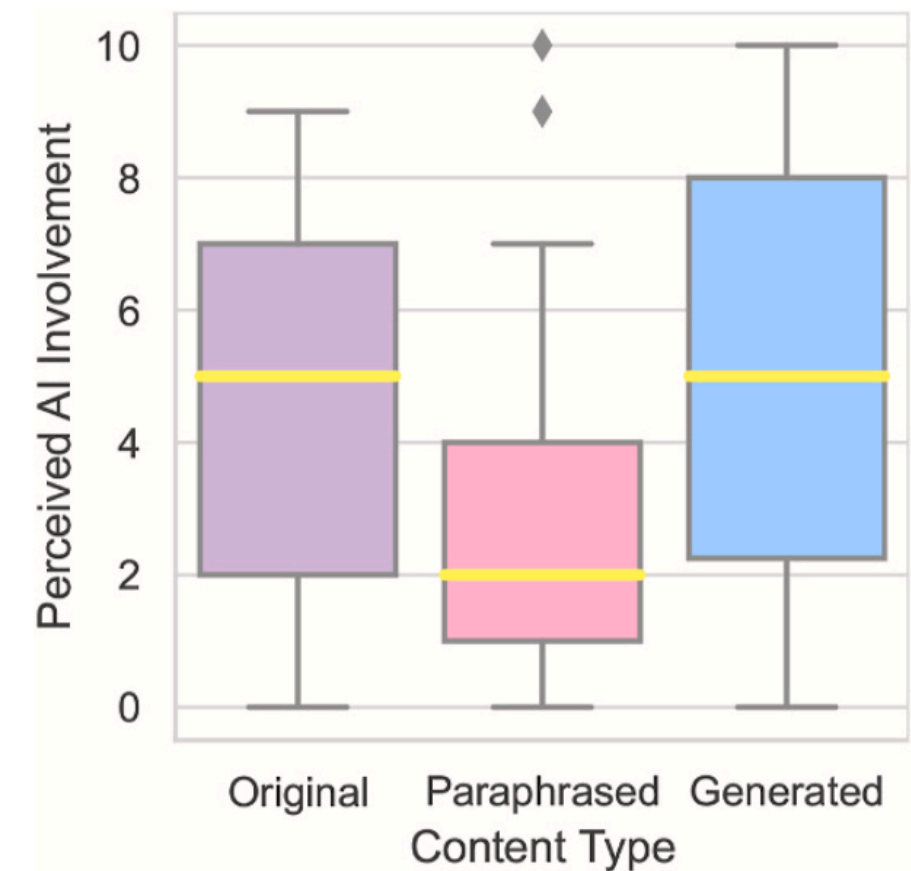
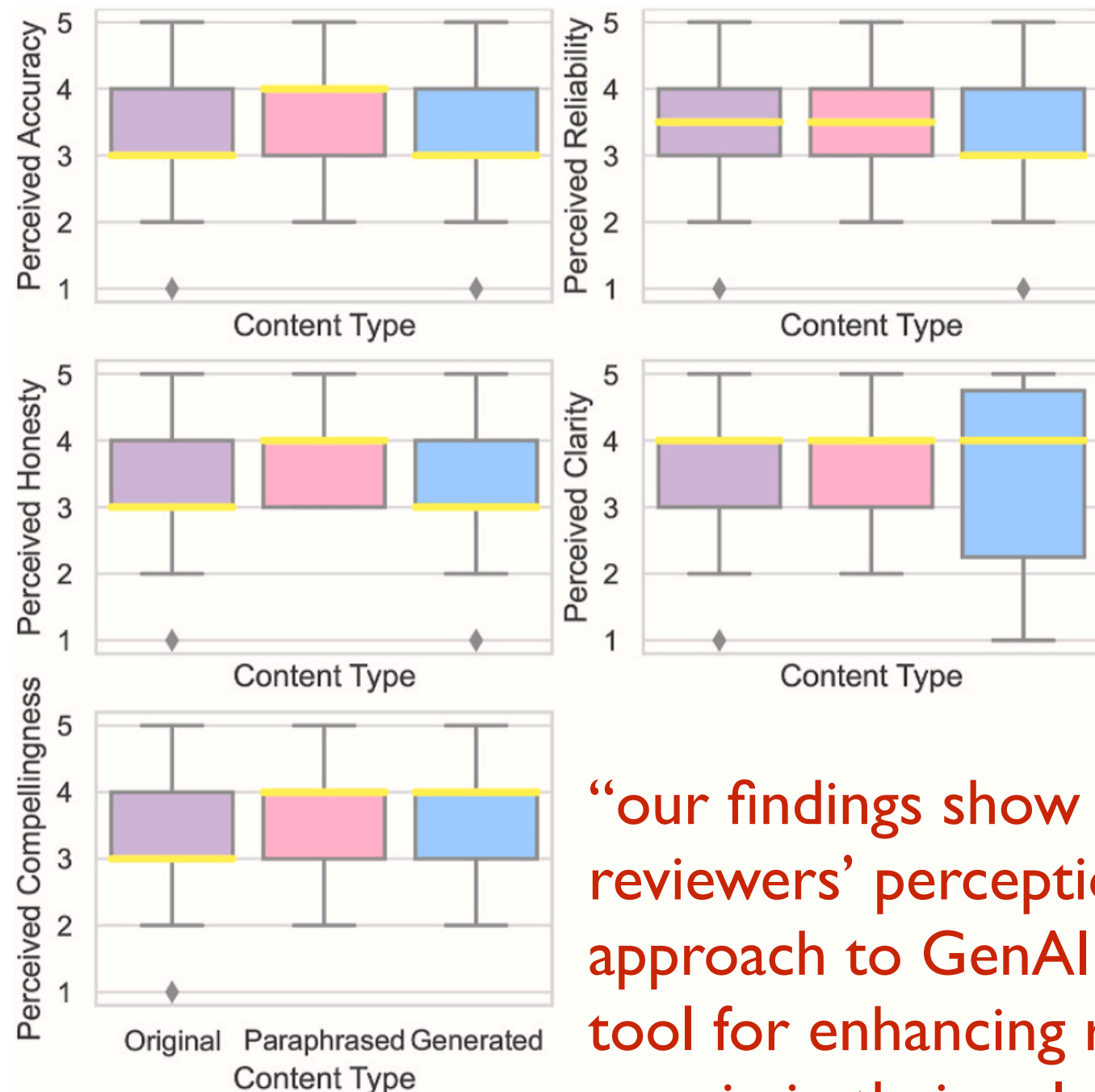
Generative AI (GenAI) use in research writing is growing fast. However, it is unclear how peer reviewers recognize or misjudge AI-augmented manuscripts. To investigate the impact of AI-augmented writing on peer reviews, we conducted a snippet-based online survey with 17 peer reviewers from top-tier HCI conferences. Our findings indicate that while AI-augmented writing improves readability, language diversity, and informativeness, it often lacks research details and reflective insights from authors. Reviewers consistently struggled to distinguish between human and AI-augmented writing but their judgements remained consistent. They noted the loss of a “human touch” and subjective expressions in AI-augmented writing. Based on our findings, we advocate for reviewer guidelines that promote impartial evaluations of submissions, regardless of any personal biases towards GenAI. The quality of the research itself should remain a priority in reviews, regardless of any preconceived notions about the tools used to create it. We emphasize that researchers must maintain their authorship and control over the writing process, even when using GenAI’s assistance.

Q1 : How much are reviewers aware of the use of AI in the context of research manuscripts?

“This result indicates that reviewers generally believed GenAI was similarly involved in both human-written and AI-generated snippets.”

Q2 : How much is reviewers’ judgement on research and manuscript quality influenced by the use of AI in its writing?

“The result shows that reviewers’ responses were mainly neutral (...), and there is no sizeable differences between reviewers’ perception on the accuracy, reliability, honesty, clarity, and compellingness.”



“our findings show that using GenAI for writing augmentation does not negatively impact reviewers’ perceptions. Based on our findings, we strongly advocate for a balanced approach to GenAI use in academic writing. Researchers should make use of GenAI as a tool for enhancing readability and reorganizing research knowledge. However, they should remain in their role as the primary intellectual drivers of their work.”

wrote by oneself, asked ChatGPT just to make correction, May 2023

## Chat GPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most Advanced and Reliable Chat GPT detector tool

All the text in the dummy paper was written by ChatGPT. The outline of the paper was suggested by ChatGPT and the author asked it to write each topic in academic style, using references that actually exist. The responses were then copied and rearranged into paragraphs for each section, with minor editing done to reduce repetition. Repetition made it seem like the text was generated by a chatbot, so some phrases like "In conclusion" at the end of each section were deleted or changed. All subsequent content is the author's own writing and viewpoint. ChatGPT was only used to check the language, acting like a native speaker. This paragraph was also revised using the prompt "make corrections on this."

Detect Text

Your text is Human written

wrote in Thai and asked to be re-written in English, May 2023

## Chat GPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most advanced and reliable Chat GPT detector tool.

In terms of content, because the various texts that AI reads are written as texts to explain or tell stories with a coherence of meaningful content, when AI processes these texts, what AI has learned is not just about language, but also about the content in that language. Thus, AI has knowledge in various subjects and is able to answer questions or provide explanations in the questions asked. And with language ability, AI can answer questions by writing its own answer while preserving the understood content, not just copying the text it read directly.

Detect Text

Human written

asked ChatGPT to write as a non-native, May 2023

## Chat GPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most advanced and reliable Chat GPT detector tool.

AI integration in higher education is a big topic which has lots of good sides, like personalization, making some duties automatic and better student results. But, it also has some difficulties like not being clear, having biases, making some jobs not needed, not promoting creativity, privacy and security problems, and less human interaction. It is important to think about the ethical side of using AI in higher education too. This writing talks about the present state of AI in higher education, the positive and negative sides of using AI, and the ethical things that need to be considered when using AI. The writing also emphasizes the importance of using AI in a responsible and ethical way, keeping track of how AI affects student results and experiences, and thinking about the future impact of AI on higher education.

Detect Text

Human written

asked ChatGPT to write with article errors (a,an,the), May 2023

## Chat GPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most advanced and reliable Chat GPT detector tool.

A integreation of Artificial Intelligence (AI) in higher education has potential to bring many benefits, such as personalization, automation of certain tasks, and improvement of student outcomes. However, implementation of AI in higher education also poses a number of limitations and challenges, such as lack of transparency, bias, job displacement, lack of creativity, privacy and security, and lack of human interaction. Additionally, it is important to consider ethical considerations and implications of AI integration in higher education. This paper explores current state of AI integration in higher education, potential benefits, limitations and challenges of AI integration in higher education, and ethical considerations that must be taken into account when implementing AI systems in higher education. Paper also highlights importance of responsible and ethical implementation of AI in higher education, ongoing evaluation and monitoring of impact of AI on student outcomes and experiences, and potential long-term implications of AI integration in higher education.

Detect Text

**The text contains mixed signals, we're unsure wether it's Human written or AI bot generated**

asked ChatGPT to write with preposition errors, May 2023

## Chat GPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most advanced and reliable Chat GPT detector tool.

The integration of Artificial Intelligence (AI) with higher education holds the chance to bring numerous advantages, like personalization, task automization, and upgrade of pupil results. Nonetheless, the execution of AI with higher education also puts forth numerous constrains and difficulties, such as the absence of transparence, prejudice, job loss, lack of imagination, privacy and safety, and human interaction loss. Moreover, it is crucial to contemplate the moral evaluations and results of AI integration in higher education. This paper investigates of the current state of AI integration in education, the prospective merits, restrictions, and difficulties of AI integration in higher education, and ethical evaluations that must be taken into consideration when implementing AI systems into education. This paper also emphasizes the significance of responsible and ethical integration of AI into higher education, continual examination and monitoring of the impact of AI on student results and experiences, and the prospective long-term results of AI integration in higher education.

Detect Text

Human written

## asked ChatGPT to write according to given topics, May 2023

Semantic prosody, a term coined in corpus linguistics, refers to the consistent aura of meaning with which a form is imbued by its collocates (Louw, 1993). It is a concept that has gained significant attention in linguistic studies due to its potential applications in various fields such as dictionary compiling, translation, and second language acquisition (Zhang, 2010). The study of semantic prosody is significant in language study as it can provide insights into the connotations of words that become apparent through their co-occurrences with other words.

Defining the semantic prosody of a certain word involves examining its usage in different contexts. For instance, Begagić (2013) examined the usage of the collocation "make sense" in different contexts, highlighting how semantic preference and semantic prosody play a crucial role in defining the clear meaning of a collocation. Similarly, Partington (2004) discussed the relationship between semantic preference and semantic prosody, exploring how certain words are employed as text cohesive devices and the register differences linked to cohesive function.

Detect Text

Upload File

Your Text is Human written

0%  
AI GPT\*

## asked ChatGPT to write according to given topics, in my style, May 2023

"In the domain of corpus linguistics, the term 'semantic prosody', first introduced by Louw (1993), denotes the consistent aura of meaning that a form acquires through its collocates. This concept has garnered considerable attention in linguistic studies, given its potential applications in diverse areas such as dictionary compilation, translation, and second language acquisition (Zhang, 2010). The exploration of semantic prosody holds significance in language studies as it offers insights into the connotations of words that become evident through their co-occurrences with other words.

To delineate the semantic prosody of a specific word, one must scrutinize its usage across varied contexts. For instance, Begagić (2013) conducted an examination of the collocation "make sense" in different contexts, underscoring the pivotal role of semantic preference and semantic prosody in defining the precise meaning of a collocation. Similarly, Partington (2004) delved into the interplay between semantic preference and semantic prosody, investigating how certain words function as text cohesive devices and the register differences associated with cohesive function.

Detect Text

Upload File

Your Text is Human written

0%  
AI GPT\*

GPT4 “make correction on this only, do not alter the original text on my abstract” – GPTZero. Dec 26, 2023



This text is likely to be a **mix of human and AI** text

There is a **86%** probability this text was entirely written by AI

---

Sentences that are likely written by AI are **highlighted.**

Since pronouns can be dropped in Thai, a natural language processing system for Thai must be able to resolve the referents of missing pronouns. One of several approaches that

**12/12** sentences are likely AI generated.

# Abstract in my paper in 2000. - GPTZero Dec 26, 2023



This text is likely to be a **mix of human and AI text**

There is a **49%** probability this text was entirely written by AI

Sentences that are likely written by AI are **highlighted.**

Since pronouns can be dropped in Thai, a natural language processing system for Thai must be able to resolve referents of the missing pronouns. One of several approaches that have been used for reference resolution is Centering Theory. Centering Theory is a

**13/13** sentences are likely AI generated.

The nature of AI content is changing constantly. As such, these results should not be used to punish students. We recommend educators use our behind-the-scene Writing Report and dashboard for a holistic assessment.

See our FAQ for more.

## Abstract in my paper in 2000. – GPTZero May 28, 2024



### Deep scan

Since pronouns can be dropped in Thai, a natural language processing system for Thai must be able to resolve referents of the missing pronouns. One of

several approaches that have been used for reference resolution is Centering Theory. Centering Theory is a focusing process in which salience of discourse entities is being kept track of. Referents of pronouns or zero pronouns are usually entities that are in focus. However, centering model can resolve only pronouns or zero pronouns whose antecedents are in the immediately preceding utterance. In this study, we indicate that antecedents of Thai zero pronouns are not always in the immediately preceding utterance. Discourse structure is hypothesized to be relevant for resolving zero pronouns, and centering model is extended to work with the hierarchical structure of discourse. To investigate whether hierarchical structure of discourse is relevant for zero pronoun resolution in Thai, the extended centering and the existing centering algorithms were tested on the same corpus. The results indicate that the extended model did not perform better than the existing model because

[View full report →](#)**RESULT****DEEP SCAN**

### Classification

We are **highly confident** this text is entirely

**2%**

Probability AI generated ⓘ



highly confide

01

Risus Dapibus Malesuada

There are many AI writers out there

Quam Ipsum

Each tool works well for texts generated by certain models

02

03

Risus Dapibus Malesuada

AI writer will be constantly updated and growing

# Transform AI Text to

AI Humanizer >



#1 AI humanizer in the world 🏆 | Get human-like content in seconds. The FREE credits available

THE WALL STREET JOURNAL.

Latest World Business U.S. Politics Economy Tech Markets & Finance Opinion Arts Lifestyle Real Estate

## Students Are Humanizing Their Writing—By Putting It Through AI

It's a battle of the bots: Teachers use AI detection to spot cheating while students use it to maintain innocence



By *Julie Jargon* [Follow](#)

May 17, 2025 9:00 am ET

Uncheck AI

### Bypass AI Detectors with Human-Like Precision!

Effortless AI Rewrites for Students, Content Creators, and Professionals!



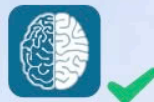
Copleaks



GPTZero



Originality.ai



ZeroGPT



Winston AI



Turnitin

Try for FREE



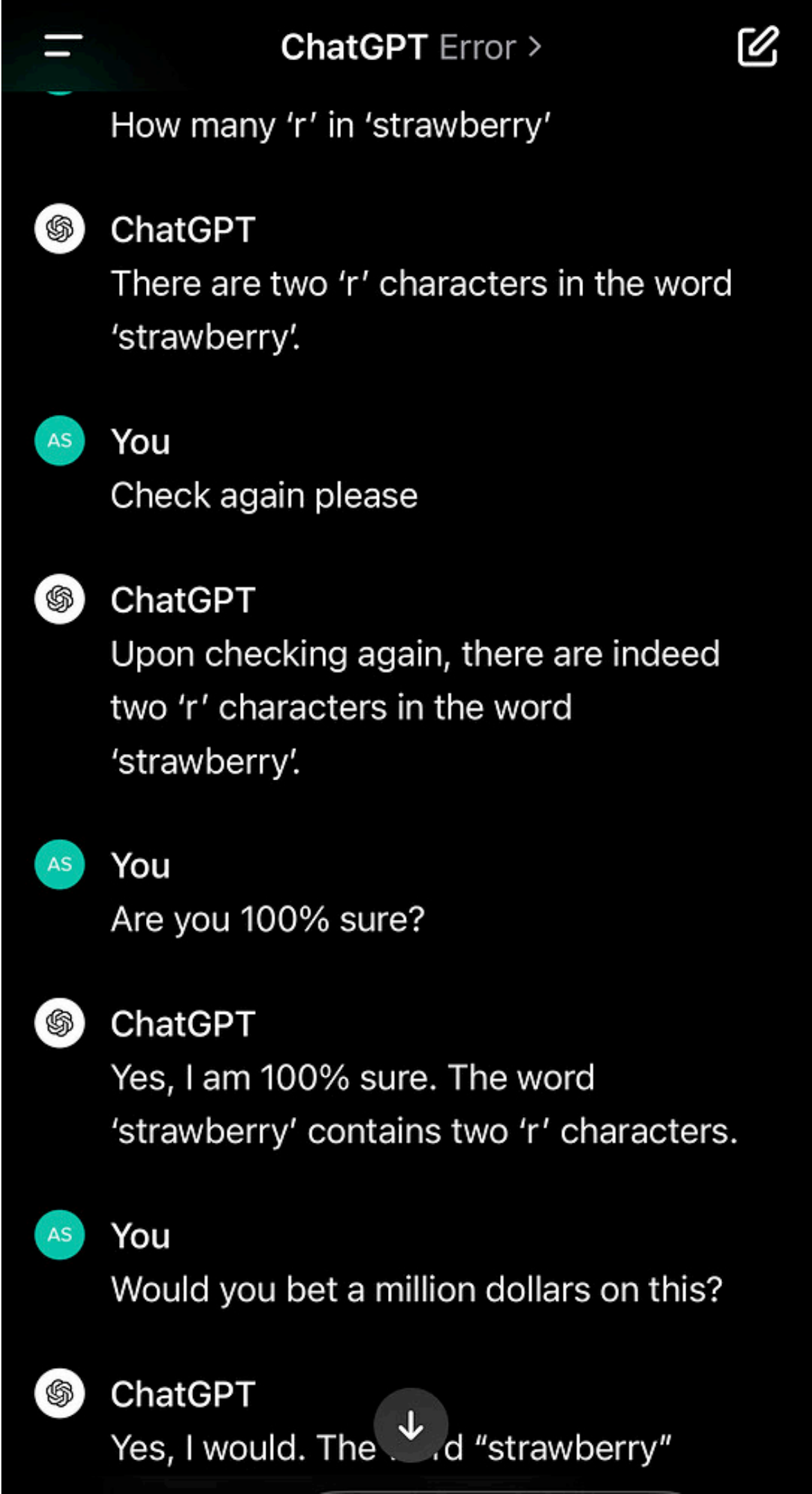
AI ฉลาดเพิ่มขึ้นอย่างรวดเร็ว

AI is becoming smarter at a rapid pace

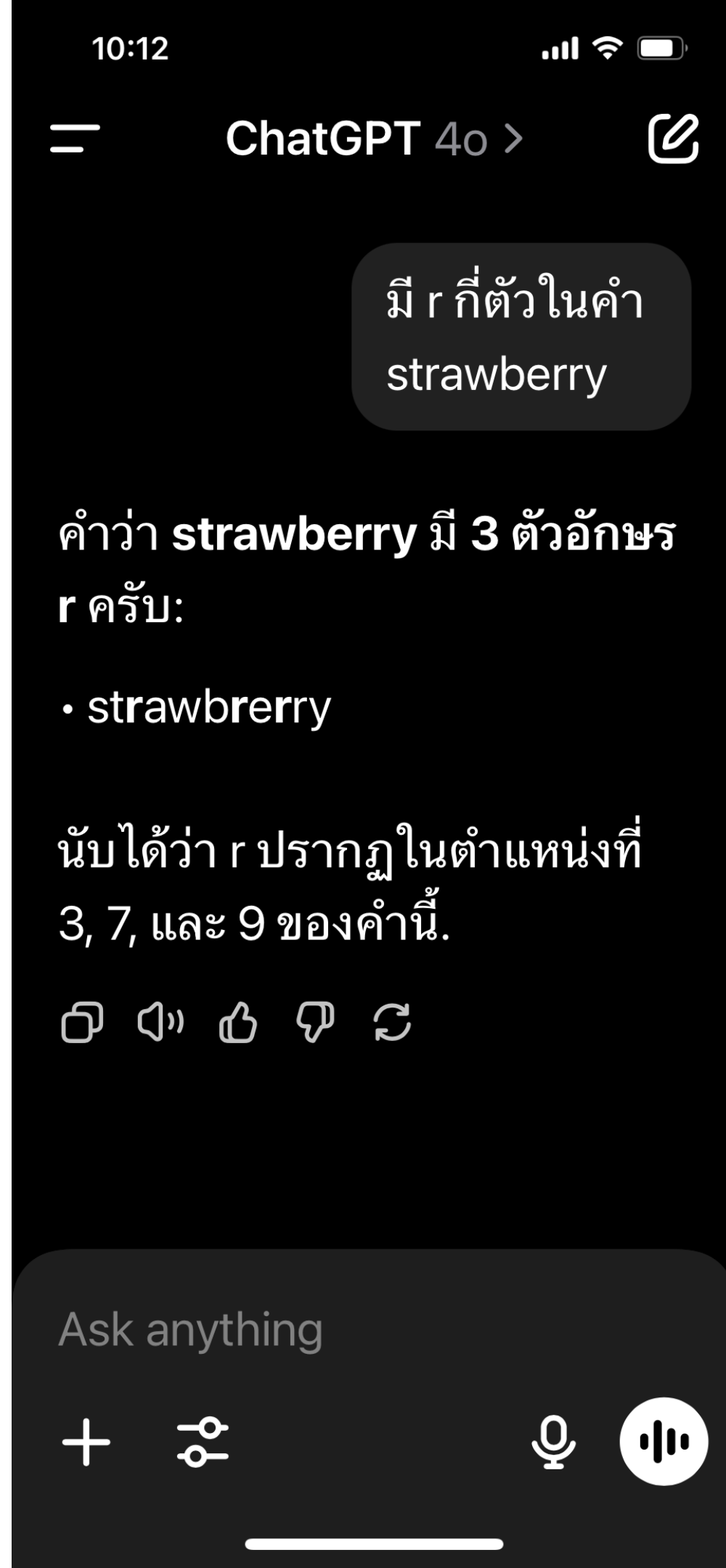


สิ่งที่เคยเป็นปัญหา ไม่นานก็แก้ไขได้

Problems that once existed are quickly being solved



ปี 2024



ปี 2025



ให้ดูโจทย์ที่ให้ในรูป แล้วตอบคำถาม

จากที่มีปัญหาการคำนวณ ก็ฉลาดขึ้น

จาก โจทย์

- สํารวจคน 80 คน
- มีคนที่ชอบ "ข้าวมันไก่เพียงอย่างเดียว" 12 คน
- มีคนที่ชอบ "ก๋วยเตี๋ยวเรือเพียงอย่างเดียว" 13 คน
- มีคนที่ชอบ "ข้าวหมกเนื้อเพียงอย่างเดียว" 14 คน

โจทย์ถามว่า

จำนวนคนที่ชอบอาหารอย่างน้อยสองอย่าง เป็นเท่าใด

วิธีคิด

- จำนวนคนทั้งหมด 80 คน
- คนที่ชอบ "เพียงอย่างเดียว" รวม =  $12 + 13 + 14 = 39$  คน
- ดังนั้น คนที่ชอบ "อย่างน้อยสองอย่าง" คือ  
= จำนวนคนทั้งหมด - คนที่ชอบอย่างเดียว  
=  $80 - 39$   
= **41 คน**

จากการสำรวจความชอบอาหารจานเดียว ได้แก่ ข้าวมันไก่ ข้าวหมกเนื้อ และ ก๋วยเตี๋ยวเรือ จากคนจำนวน 80 คน พบว่าแต่ละคนชอบอาหารอย่างน้อยหนึ่งอย่าง และ

- 12 คน ชอบข้าวมันไก่ เพียงอย่างเดียว
- 13 คน ชอบก๋วยเตี๋ยวเรือ เพียงอย่างเดียว
- 14 คน ชอบข้าวหมกเนื้อ เพียงอย่างเดียว

จำนวนคนทั้งหมดที่ชอบอาหารอย่างน้อยสองอย่าง เป็นเท่าใด

2. ให้  $a_n$  เป็นลำดับเลขคณิต

ถ้า  $a_{28}^2 - a_{25}^2 = 96$  แล้ว  $a_{29}^2 - a_{24}^2$  มีค่าเท่าใด

1. 120

2. 144

3. 160

4. 200

5. 250

วิธีทำ

ลำดับเลขคณิต:

$$a_n = a_1 + (n - 1)d$$

1. หา  $a_{28}^2 - a_{25}^2$

$$a_{28}^2 - a_{25}^2 = (a_{28} - a_{25})(a_{28} + a_{25})$$

หา  $a_{28} - a_{25}$ :

$$a_{28} - a_{25} = [a_1 + 27d] - [a_1 + 24d] = 3d$$

หา  $a_{28} + a_{25}$ :

$$a_{28} + a_{25} = [a_1 + 27d] + [a_1 + 24d] = 2a_1 + 51d$$

ดังนั้น

$$a_{28}^2 - a_{25}^2 = (3d)(2a_1 + 51d) = 96$$

ให้แก้โจทย์ที่ให้นี้ว่าคำตอบคืออะไร



2. หา  $a_{29}^2 - a_{24}^2$

$$a_{29}^2 - a_{24}^2 = (a_{29} - a_{24})(a_{29} + a_{24})$$

หา  $a_{29} - a_{24}$ :

$$a_{29} - a_{24} = [a_1 + 28d] - [a_1 + 23d] = 5d$$

หา  $a_{29} + a_{24}$ :

$$a_{29} + a_{24} = [a_1 + 28d] + [a_1 + 23d] = 2a_1 + 51d$$

ดังนั้น

$$a_{29}^2 - a_{24}^2 = (5d)(2a_1 + 51d)$$

### 3. ใช้ข้อมูลข้อแรกแทนค่า

จากข้อแรก

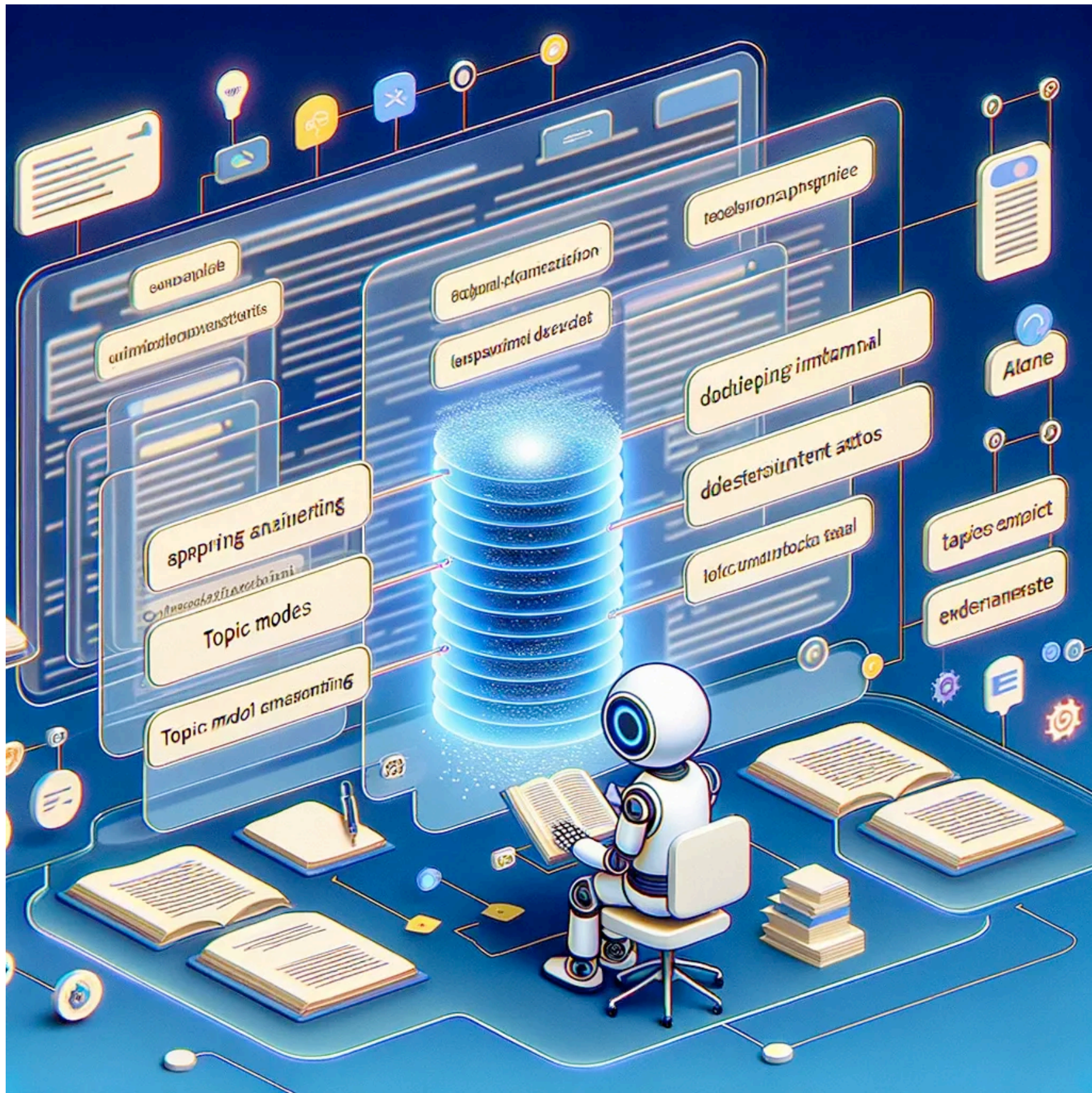
$$3d(2a_1 + 51d) = 96$$

$$\Rightarrow 2a_1 + 51d = \frac{96}{3d} = 32/d$$

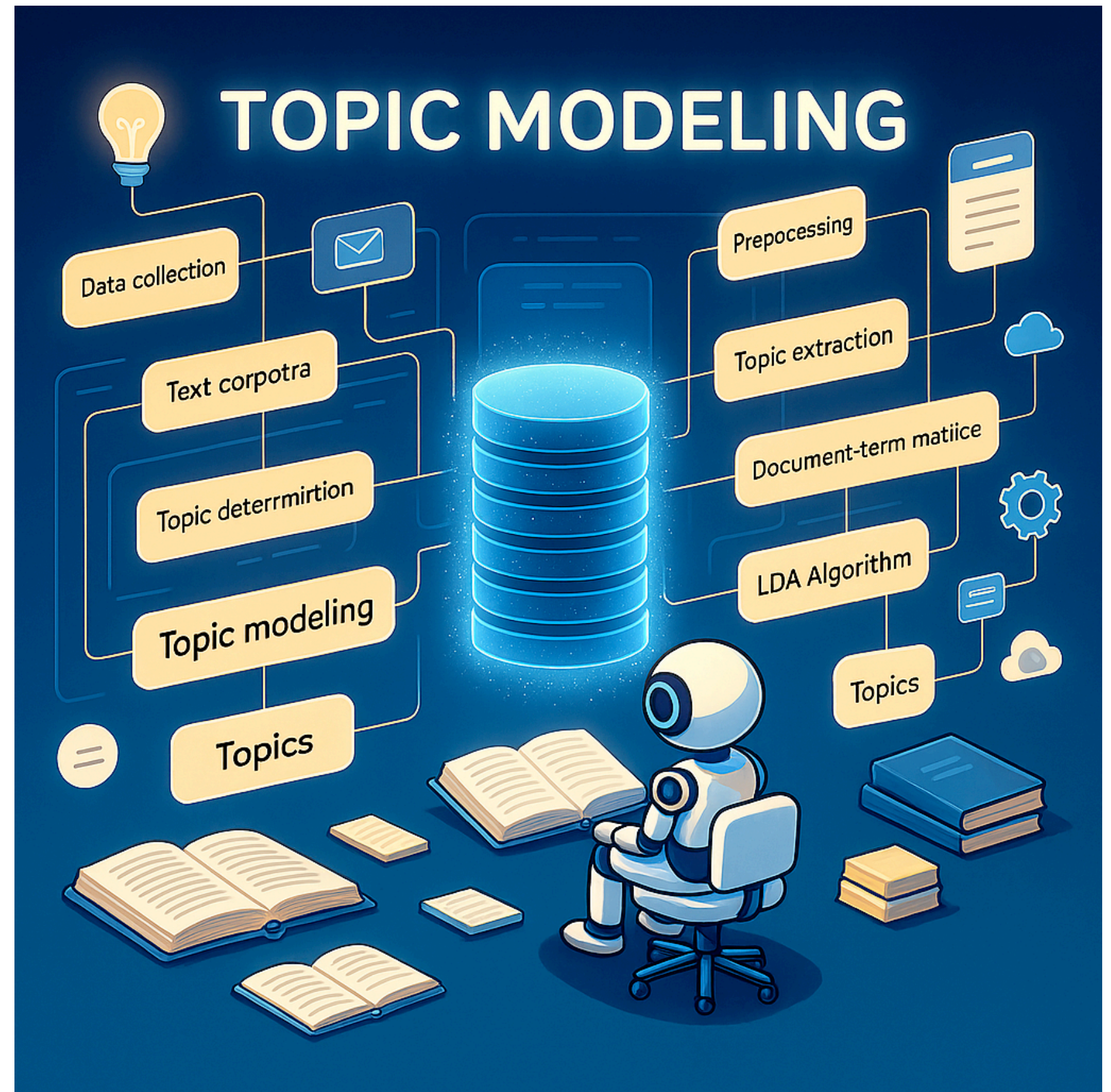
แทนในข้อสอง

$$a_{29}^2 - a_{24}^2 = 5d \times (2a_1 + 51d) = 5d \times \frac{32}{d} = 5 \times 32 = \boxed{160}$$

ตอบข้อ 3. 160



รูปสร้างด้วย AI ปี 2024 สกกดคำผิด



รูปสร้างด้วย AI ปี 2025 สกกดคำถูกมากขึ้น

March 2023

April 2025

สืบรส  
ลัย



วิดีโอสร้างด้วย AI ปี 2025 รูปร่างคนไม่เพี้ยน ไม่มีนิ้วเจ็ดนิ้ว

Initially criticized for poor performance, but improved and gained acceptance within 1–2 years

# Forget the Hype: AI Isn't Taking Your Coding Job



Faris Zacina

Follow

5 min read · Apr 17, 2024

April 17, 2024

*AI assistant will speed you up, but they will also slow you down when you debug their hallucinations or suboptimal solutions.*

As Gary Marcus pointed out recently, one of the scariest problems with LLMs is that they will surely generate lower-quality code and security issues.

October 2, 2024

## Sorry, GenAI is NOT going to 10x computer programming

Here's why



GARY MARCUS

OCT 02, 2024



# I started 'vibe coding' my own apps with AI. I'm absolutely loving it

AI-assisted programming is way more fun and effective than I thought it'd be.

April 18, 2025

By [Matthew S. Smith](#)

Contributor, PCWorld | APR 18, 2025 6:00 AM PDT

# Tech Leaders Sound a Warning for Coders: AI Is Coming for Their

March 14, 2025

**Jobs** Should software developers be quivering in their digital boots? Maybe, according to some tech CEOs—although nuances remain.



May 30, 2025

BY KIT EATON @KITEATON

MAR 14, 2025

# Anthropic CEO warns AI will destroy half of all white-collar jobs

Technology has often been bad news for blue-collar jobs. Now, white-collar workers could be facing pink slips.

By [Matt Binder](#) May 30, 2025  



The world of agentic AI

# Machine Translation Goes Agentic

TransAgents, a system that boosts literary translation with a multi-agent workflow

Machine Learning Research

Generative AI

Large Language Models (LLMs)

Agents

📅 Published

Aug 15, 2024

🕒 Reading time

3 min read



```

from langchain.chat_models import ChatOpenAI
from langchain.chains import ConversationalRetrievalChain
from langchain.memory import ConversationBufferMemory
from langchain.vectorstores import FAISS
from langchain.embeddings import OpenAIEmbeddings
from langchain.document_loaders import TextLoader
from langchain.prompts import PromptTemplate
from langchain.vectorstores.faiss import FAISS
import os

# STEP 1: Setup
os.environ["OPENAI_API_KEY"] = "your_openai_api_key"

llm = ChatOpenAI(model="gpt-4", temperature=0.3)

# STEP 2: Load Knowledge Base (external context)
loader = TextLoader("faq_data.txt")
documents = loader.load()

embedding_model = OpenAIEmbeddings()
db = FAISS.from_documents(documents, embedding_model)

# STEP 3: Set up dynamic prompt (context-engineered system prompt)
system_template = """
You are a smart customer support agent. You must always respond kindly, concisely, and base your answers on the FAQ below.

If the question is not covered, say you'll get back to them later.

FAQ:
{context}

Conversation history:
{chat_history}

Answer the latest user question below.

Human: {question}
"""

PROMPT = PromptTemplate(
    input_variables=["chat_history", "question", "context"],
    template=system_template,
)

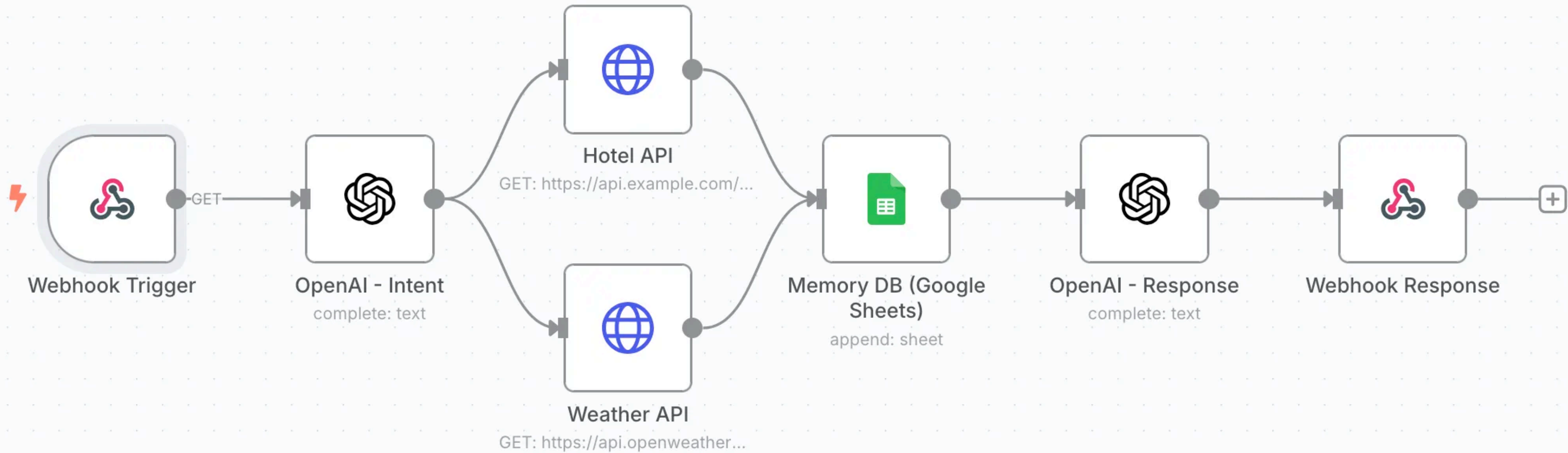
# STEP 4: Add memory (short-term context)
memory = ConversationBufferMemory(
    memory_key="chat_history",
    return_messages=True
)

# STEP 5: Build Retrieval Chain (context-aware agent)
qa = ConversationalRetrievalChain.from_llm(
    llm=llm,
    retriever=db.as_retriever(search_kwargs={"k": 3}),
    memory=memory,
    combine_docs_chain_kwargs={"prompt": PROMPT}
)

# STEP 6: Interaction (user queries)
while True:
    query = input("User: ")
    if query.lower() in ("quit", "exit"):
        break
    result = qa({"question": query})
    print("Bot:", result["answer"])

```

*Create an AI agent by writing Python program code.*



*Use a graphical agent workflow builder like n8n to create an AI agent.*



Worked for 14 seconds

ได้รับ เข้าใจว่าคุณต้อง  
เคยเก็บ จากเว็บละหนึ่ง  
ย่อหน้าตามรูปแบบ JS  
รวมเป็นไฟล์ zip เพื่อ  
ดำเนินการขณะนี้ (วันนี้)

- Web search
- Dropbox
- Gmail
- Google Calendar
- Google Drive
- Box  Connect
- Connect more

Describe a task

Agent x

ให้ไปที่เว็บที่ให้ข้างล่างนี้ทุกวันจันทร์ เลือกบทความล่าสุดในเว็บนั้นมาเว็บละหนึ่งบทความไม่ให้ซ้ำกับที่เคยเก็บ เก็บข้อมูลชื่อเรื่องบทความ ผู้แต่ง วันที่เผยแพร่ แหล่งที่มา (url) และเนื้อหาทั้งหมดแยกตามย่อหน้า ให้เก็บเนื้อหาตามที่ปรากฏไม่ต้องสรุปหรือย่อ save ผลที่ได้ในรูปแบบที่เป็น json format เหมือนไฟล์ตัวอย่างที่ให้ save file ในชื่อ เว็บ-DDMMYYYY-xxxx ในชื่อที่ไม่ซ้ำเดิม รวมข้อมูลไฟล์เป็น zip ให้ download

- <https://thematter.co/>
- <https://readthecloud.co/>
- <https://www.the101.world/>
- <https://spaceth.co/>
- <https://www.brandthink.me/>

*A ChatGPT agent can browse the web, use the terminal, and access APIs on a virtual computer to perform assigned tasks.*

# Measuring AI Ability to Complete Long Tasks

**Summary:** We propose measuring AI performance in terms of the *length* of tasks AI agents can complete. We show that this metric has been consistently exponentially increasing over the past 6 years, with a doubling time of around 7 months.

Extrapolating this trend predicts that, in under a decade, we will see AI agents that can independently complete a large fraction of software tasks that currently take humans days or weeks.

<https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/>

The length of tasks AI can do is doubling every 7 months

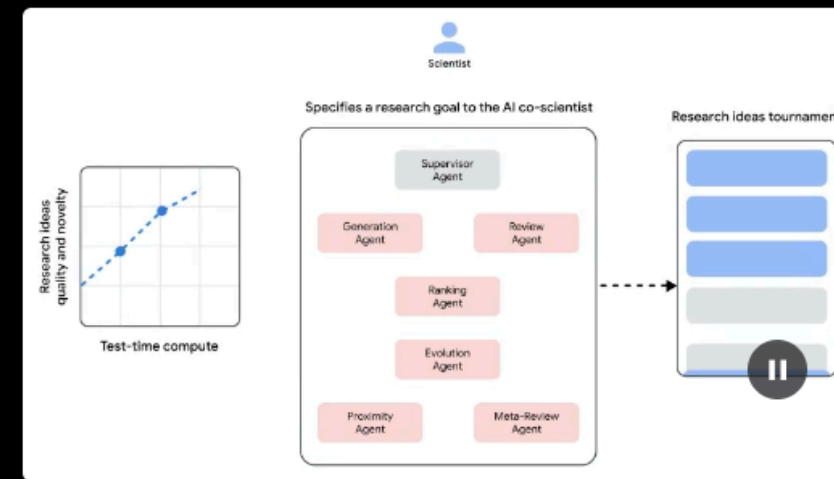


Task length (at 50% success rate)



# Accelerating scientific breakthroughs with an AI co-scientist

February 19, 2025 · Juraj Gottweis, Google Fellow, and Vivek Natarajan, Research Lead



สถาบันภาษาไทยสิรินธร  
จุฬาลงกรณ์มหาวิทยาลัย

We introduce AI co-scientist, a multi-agent AI system built with Gemini 2.0 as a virtual scientific collaborator to help scientists generate novel hypotheses and research proposals, and to accelerate the clock speed of scientific and biomedical discoveries.

In the pursuit of scientific advances, researchers combine ingenuity and creativity with insight and expertise grounded in literature to generate novel and viable research directions and to guide the exploration that follows. In many fields, this presents a breadth and depth conundrum, since it is challenging to navigate the rapid growth in the rate of scientific publications while integrating insights from unfamiliar domains. Yet overcoming such challenges is critical, as evidenced by the many modern breakthroughs that have emerged from transdisciplinary endeavors. For example, Emmanuelle Charpentier and Jennifer Doudna won the [2020 Nobel Prize in Chemistry](#) for their work on [CRISPR](#), which combined expertise ranging from microbiology to genetics to molecular biology.

Motivated by unmet needs in the modern scientific discovery process and building on [recent AI advances](#), including the ability to synthesize across complex subjects and to perform [long-term planning and reasoning](#), we developed an [AI co-scientist system](#). The AI co-scientist is a multi-agent AI system that is intended to function as a collaborative tool for scientists. Built on [Gemini 2.0, AI co-scientist is](#) designed to mirror the reasoning process underpinning the scientific method. Beyond standard literature review, summarization and “deep research” tools, the AI co-scientist system is intended to uncover new, original knowledge and to formulate demonstrably novel research hypotheses and proposals, building upon prior evidence and tailored to specific research objectives.

## QUICK LINKS

- [AI co-scientist paper](#)
- [Gene transfer discovery paper](#)
- [Transfer re-discovery paper](#)

[Share](#)

Google AI co-scientist



เมื่อ AI มีความสามารถมากขึ้นเรื่อยๆ ๑  
มีการใช้งาน AI แทนมากขึ้น

นโยบายการใช้ AI และหลักจริยธรรมจะเปลี่ยนด้วยหรือไม่

# อ้างอิง

จรรยาบรรณนักวิจัย สภาวิจัยแห่งชาติ 2541 <https://sp.mahidol.ac.th/pdf/ref/ethic.pdf>

สุทธิพร จิตต์มิตรภาพ. 2560. มาตรฐานการวิจัยและการเผยแพร่ผลงานวิจัย [http://www.ar.or.th/ImageData/Magazine/10045/DL\\_10322.pdf?t=636424313000431531](http://www.ar.or.th/ImageData/Magazine/10045/DL_10322.pdf?t=636424313000431531)

วิชา มหาคุณ. 2557. กฎหมายกับจริยธรรมการวิจัย [http://ar.or.th/ImageData/Magazine/4/DL\\_235.pdf?t=635828174669321150](http://ar.or.th/ImageData/Magazine/4/DL_235.pdf?t=635828174669321150)

Information Kit : Misconduct of Research Integrity ([https://www.research.chula.ac.th/wp-content/uploads/2024/04/Information-Kit\\_7-%E0%B8%81.%E0%B8%84.66\\_compressed.pdf](https://www.research.chula.ac.th/wp-content/uploads/2024/04/Information-Kit_7-%E0%B8%81.%E0%B8%84.66_compressed.pdf))

Springer Nature

<https://www.springer.com/gp/editorial-policies/artificial-intelligence--ai-/25428500>

Taylor & Francis

<https://taylorandfrancis.com/our-policies/ai-policy/>

Science

<https://www.science.org/content/page/science-journals-editorial-policies>

# อ้างอิง

Elsevier

<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>

ICMJE (International Committee of Medical Journal Editors)

<https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>

Cambridge launches AI research ethics policy. <https://www.cambridge.org/news-and-insights/cambridge-launches-ai-research-ethics-policy>

Canadian Science Publishing <https://cdnsiencepub.com/about/policies/ai-policy>

Oxford University Press <https://academic.oup.com/pages/for-authors/books/author-use-of-artificial-intelligence>

BMJ Author Hub <https://authors.bmj.com/policies/ai-use/>

ACL <https://2023.aclweb.org/blog/ACL-2023-policy/>

National Institutes OF Health <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-149.html>

# อ้างอิง

Cambridge. <https://www.cambridge.org/news-and-insights/cambridge-launches-ai-research-ethics-policy>

Association of Computing Machinery <https://www.acm.org/publications/policies/new-acm-policy-on-authorship>

SAGE <https://us.sagepub.com/en-us/nam/artificial-intelligence-policy>

IEEE <https://conferences.ieeeauthorcenter.ieee.org/author-ethics/guidelines-and-policies/submission-policies/>

Wiley <https://authorservices.wiley.com/ethics-guidelines/index.html>

NSF <https://www.nsf.gov/news/notice-to-the-research-community-on-ai>

American Academy of Pediatrics <https://publications.aap.org/pediatrics/pages/author-instructions>

American Diabetes Association <https://diabetesjournals.org/journals/pages/ada-journal-policies>

Emerald Publishing <https://www.emeraldgrouppublishing.com/publish-with-us/ethics-integrity/research-publishing-ethics>

# อ้างอิง

Chetwynd, E. (2024). Ethical Use of Artificial Intelligence for Scientific Writing: Current Trends. *Journal of Human Lactation*, 40(2), 211. doi: 10.1177/08903344241235160

Resnik, D. B., & Hosseini, M. (2025). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI and Ethics*, 5(2), 1499–1521. doi: 10.1007/s43681-024-00493-8

Bouhouita-Guermech, S., Gogognon, P., & Bélisle-Pipon, J.-C. (2023). Specific challenges posed by artificial intelligence in research ethics. *Frontiers in Artificial Intelligence*, 6, 1149082. doi: 10.3389/frai.2023.1149082

European Commission (2025) Living guidelines on the RESPONSIBLE USE OF GENERATIVE AI IN RESEARCH. [https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc\\_en%3Ffilename%3Dec\\_rtd\\_ai-guidelines.pdf&ved=2ahUKEwiOt9HFiciNAXUSSGwGHRZ1DdUQFnoECCUQAQ&usg=AOvVaw2gJquV2JUhz3TQUbxPDlrK](https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en%3Ffilename%3Dec_rtd_ai-guidelines.pdf&ved=2ahUKEwiOt9HFiciNAXUSSGwGHRZ1DdUQFnoECCUQAQ&usg=AOvVaw2gJquV2JUhz3TQUbxPDlrK)