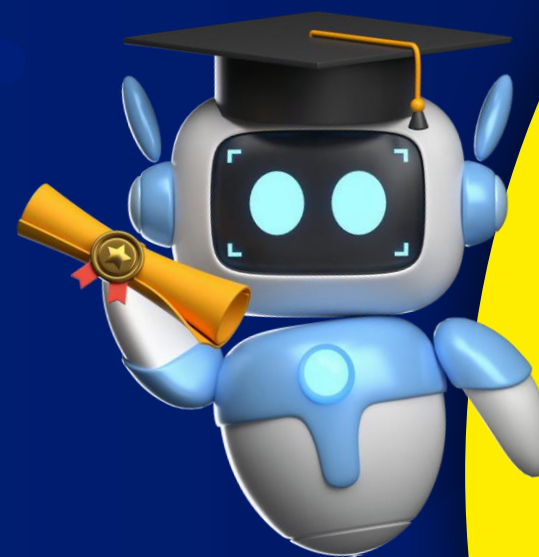


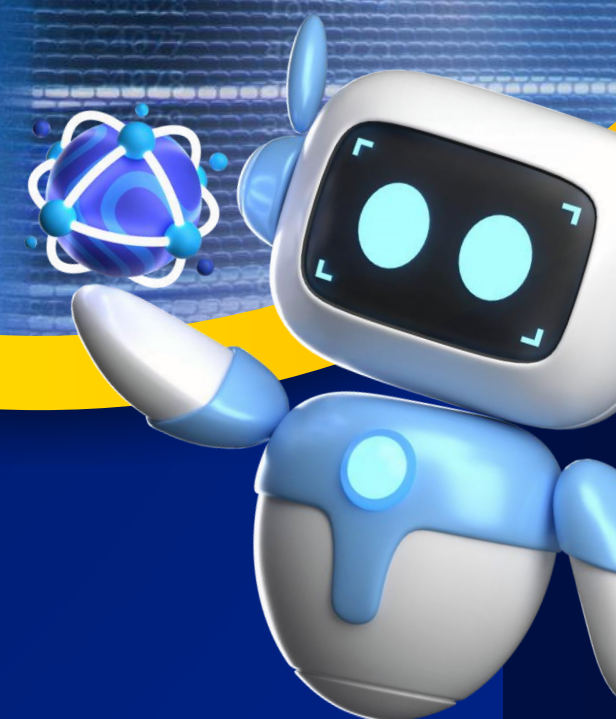


สถาบันภาษาไทยสิรินธรแห่งจุฬาลงกรณ์มหาวิทยาลัย
THE SIRINDHORN THAI LANGUAGE INSTITUTE
Chulalongkorn University



การทดลองใช้ AI ออกข้อสอบจากบทอ่าน

ณัฐพล ฟังน้อย



การใช้คำสั่งให้ AI ออกข้อสอบ

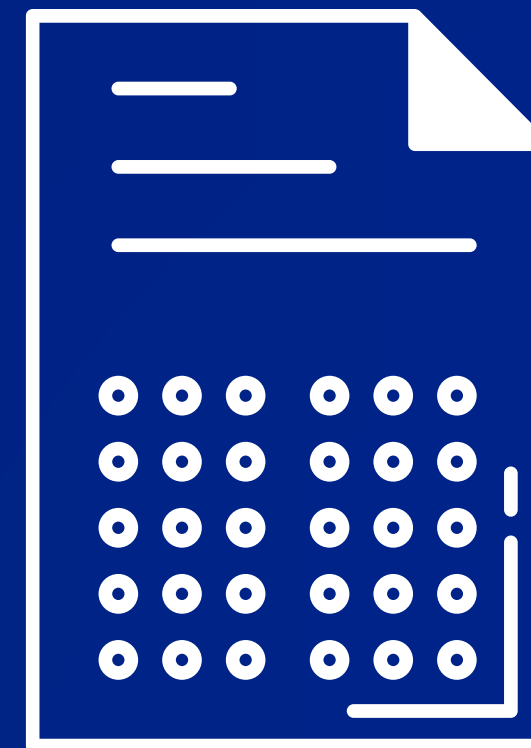


ให้ทำหน้าที่ผู้เชี่ยวชาญ
ออกข้อสอบการอ่าน



ให้ศึกษาคำอธิบาย
ความสามารถด้านการอ่าน
และบทอ่าน

+ ตัวอย่างข้อสอบเดิม

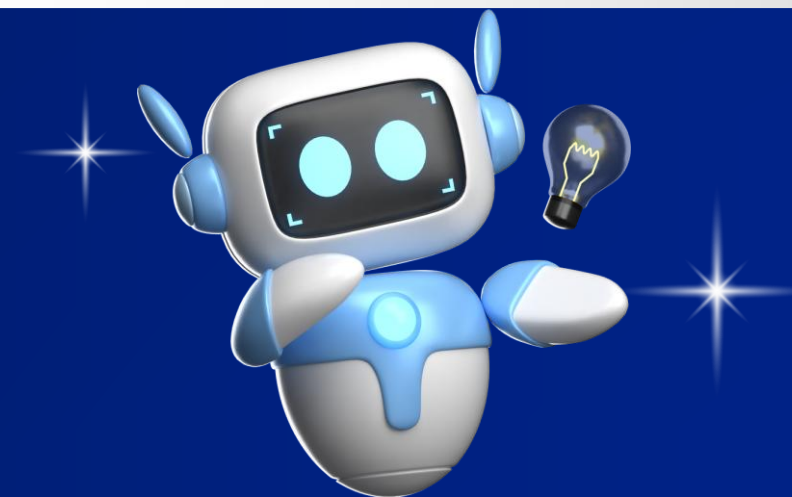


ให้ตั้งคำถามและตัวเลือก
เพื่อวัดความสามารถด้านการอ่าน
ประเภทต่าง ๆ

การใช้คำสั่งให้ AI ออกข้อสอบ

ให้ทำหน้าที่เป็นผู้เชี่ยวชาญออกข้อสอบการอ่าน ให้ใช้คำอธิบายความสามารถในการอ่านแบบต่าง ๆ ที่ให้อยู่ข้างล่างนี้ ทำความเข้าใจว่าการอ่านแต่ละประเภทคืออะไรเพื่อตั้งคำถามที่เหมาะสมกับบทอ่านที่ให้ หลังจากทำความเข้าใจแล้ว จะให้บทอ่าน แล้วให้อ่านบทอ่านนั้น จากนั้นให้ตั้งคำถามวัดการอ่านจับรายละเอียด (.....) ข้อ จับใจความ (.....) ข้อ ดีความ (.....) ข้อ อ่านแบบมีวิจารณญาณ (.....) ข้อ พร้อมตัวเลือก 4 ตัวเลือกในแต่ละข้อ มีตัวเลือกที่เป็นคำตอบที่ถูกต้อง 1 ตัวเลือก และที่เหลือเป็นตัวเลือกที่ผิด

ข้อค้นพบ



- 01 ความแตกต่างระหว่างการใช้ GPT-o1, GPT-4o และการให้ทั้งสองโมเดลเลียนแบบข้อสอบเก่า
- 02 การตั้งคำถามให้ตรงกับความสามารถด้านการอ่านแบบต่าง ๆ ที่กำหนดให้
- 03 การสร้างตัวเลือก
- 04 การสร้าง agent แก่ตัวเลือกที่ถูก 3 ตัวให้ผิด เพื่อให้กลายเป็นตัวลวง

ข้อค้นพบ

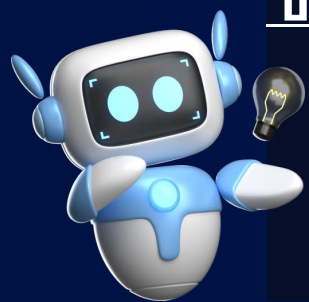
01

ความแตกต่างระหว่างการใช้ GPT-o1, GPT-4o และการให้ทั้งสองโมเดลเลียนแบบข้อสอบเก่า

ข้อสอบที่ได้จากการใช้ AI สร้างทั้งสามรูปแบบไม่ต่างกันมากนัก ในประเด็นต่าง ๆ ต่อไปนี้

- การหาประเด็นมาตั้งคำถาม
- การออกข้อสอบที่ตรงความสามารถด้านการอ่านแบบต่าง ๆ ที่กำหนด
- การตั้งคำถาม
- การสร้างตัวเลือก

ข้อสังเกต



ภาษาของ GPT-4o จะกระชับและอ่านเข้าใจง่ายกว่า GPT-o1

การให้โมเดลเลียนแบบข้อสอบเก่า ก็จะใช้ภาษาที่กระชับและดีขึ้นตามภาษาในข้อสอบเก่า

ตัวอย่างจากเรื่อง : กลไกตลาดจะแก้ปัญหาคาลำสัตว์ใกล้สูญพันธุ์ได้อย่างไร?

GPT-01

ในบทความกล่าวว่า สินค้าเวเบลน (Veblen Goods) หมายถึงสินค้าในลักษณะใด

- ก. สินค้าที่ราคาสูงขึ้นแล้วความต้องการลดลง
- ข. **สินค้าราคาสูงที่คนอยากได้มากขึ้นเพื่อแสดงฐานะ**
- ค. สินค้าประเภทอาหารหลักที่เป็นปัจจัยสี่ของมนุษย์
- ง. สินค้าที่รัฐอุดหนุนเพื่อช่วยลดราคาจำหน่าย

GPT-4o

สินค้าเวเบลน (Veblen Goods) คือสินค้าที่มีลักษณะอย่างไร?

- ก. ยิ่งแพงยิ่งไม่มีใครต้องการ
- ข. **ยิ่งแพงยิ่งมีผู้ต้องการมาก**
- ค. ราคามีผลต่อความต้องการลดลงตามปกติ
- ง. เป็นสินค้าธรรมดาที่พบเห็นได้ทั่วไป

ข้อค้นพบ

02

การตั้งคำถามให้ตรงกับความสามารถด้านการอ่านแบบต่าง ๆ ที่กำหนดให้

ภาพรวม

- สามารถหาประเด็นที่น่าสนใจมาตั้งคำถามได้
- ออกข้อสอบได้ค่อนข้างตรงกับความสามารถด้านการอ่านแบบต่าง ๆ ที่กำหนดให้เป็นส่วนใหญ่

การอ่านตีความ

- จะตรงก็ต่อเมื่อบทอ่านหรือเนื้อหามีส่วนที่เอื้อให้ตั้งคำถามแบบตีความได้

การอ่านอย่างมีวิจารณญาณ

- จะตรงในบางประเภท เช่น การหาข้อเท็จจริงข้อคิดเห็น และการหาข้อมูลที่สนับสนุนข้ออ้างของผู้เขียน

หมายเหตุ : มีข้อสังเกตว่า GPT-o1 ยังใช้ภาษาที่ฟุ่มเฟือยและซับซ้อนในการเขียนคำถาม
ทำให้อ่านแล้วเข้าใจยากกว่า GPT-4o

ตัวอย่างการจับใจความ : เรื่อง ข้าว กับบทบาทต่อการเติบโตและล่มสลายของรัฐอยุธยา

GPT-01

ข้อใดสรุปความจากหลายส่วนของบทอ่านได้ถูกต้องว่า “ข้าว” มีบทบาทอย่างไรต่ออาณาจักรอยุธยา

ก. ข้าวเป็นเพียงสินค้าแลกเปลี่ยนเล็กน้อยของอยุธยา

ข. ข้าวเป็นทั้งพืชเศรษฐกิจหลัก เลี้ยงดูประชากร และเป็นเสบียงสำคัญในยามสงคราม

ค. ข้าวในอยุธยาไม่ค่อยมีบทบาท เพราะประชากรอยู่ได้ด้วยการประมงเป็นหลัก

ง. ข้าวใช้เป็นเครื่องบรรณาการให้รัฐอื่นโดยอยุธยาไม่ได้ใช้บริโภคเอง

หมายเหตุ : จับใจความ สรุปใจความหลักของเรื่อง

คำถามควรแก้เป็น → ข้อใดสรุปความจากหลายส่วนของบทอ่านได้ถูกต้องว่า “ข้าว” มีบทบาทอย่างไรต่ออาณาจักรอยุธยา

ตัวอย่างการอ่านดีความ : เรื่อง วินัย

GPT-01

ให้เลียนแบบข้อสอบเก่า

“ผู้เขียนมีเจตนาหลักใดในการเล่าเรื่อง ‘วินัย’”

- ก. เพื่อกระตุ้นให้ทุกคนตรวจสอบสุขภาพปีละครั้งเป็นอย่างน้อย
- ข. เพื่อบอกว่าการดื่มสุราเป็นเรื่องธรรมดาที่ใคร ๆ ก็ทำได้
- ค. เพื่อเน้นย้ำความสำคัญของการมีวินัยในการดูแลสุขภาพ
- ง. เพื่อวิพากษ์วิจารณ์การรักษาพยาบาลของแพทย์

หมายเหตุ

ดี ความรู้เจตนา/วัตถุประสงค์ของผู้เขียนที่ไม่ได้บอกตรง ๆ (สังเกตได้ว่า GPT-01 ใช้ภาษาดีกว่าเดิมเมื่อมีข้อสอบที่ดีให้เลียนแบบ แต่ประเด็นในการถามไม่แตกต่าง)

ตัวอย่างการอ่านจับใจความ(ที่ไม่ตรง) : เรื่อง วินัย

GPT-4o

ใจความสำคัญของข้อความที่กล่าวถึงการเปรียบเทียบการป่วยเป็นการติดคุกคืออะไร?

ก. คนป่วยต้องมีวินัยในการปฏิบัติตัวเพื่อให้หายจากโรค

ข. คนป่วยต้องหาทางแหกคุกเพื่อพ้นตัวเร็วขึ้น

ค. หมอเป็นผู้คุมที่มีบทบาทสำคัญในการรักษา

ง. การป่วยเป็นสิ่งที่หลีกเลี่ยงไม่ได้และต้องยอมรับ

หมายเหตุ : เป็น*ตีความ*มากกว่า*จับใจความ* เพราะต้องแปลความหมายข้อความที่ผู้เขียนไม่ได้บอกตรง ๆ แล้วต้องอาศัยบริบทในเรื่องช่วยทำความเข้าใจ

ตัวอย่างการอ่านอย่างมีวิจารณญาณ : กลไกตลาดจะแก้ปัญหากล้าสัตว์ใกล้สูญพันธุ์ได้อย่างไร?

GPT-01

ข้อใดที่เป็น “ข้อคิดเห็น” (opinion) มากกว่า “ข้อเท็จจริง” (fact) ตามเนื้อหาบทความ

ก. “กบพิษเลห์มันน์ถูกเรียก ‘เพชรของวงการคนรักกบ’”

ข. “กบพิษเลห์มันน์อาจทำให้คนซื้อสุกไม่ผิดบาปถ้าซื้อมาจากฟาร์ม”

ค. “กบพิษเลห์มันน์พบได้เฉพาะที่หุบเขาทางตะวันตกของโคลอมเบีย”

ง. “ราคากบพิษเลห์มันน์ลายทองคาดดำอาจสูงถึง 50,000 บาทต่อตัว”

หมายเหตุ : วิจารณ์งาน แยกแยะข้อเท็จจริงข้อคิดเห็น

คำถามควรแก้เป็น

ข้อใดเป็น “ข้อคิดเห็น” (opinion) มากกว่า “ข้อเท็จจริง” (fact) ตามเนื้อหาบทความ

ตัวอย่างการอ่านอย่างมีวิจารณญาณ : “ทุเรียน” ราชพลไม้แห่งนุชนันทาราช

GPT-01

ผู้เขียนใช้ข้อมูลหรือหลักฐานใดในการสรุปว่า “อินโดนีเซียไม่ได้ครองตลาดส่งออกทุเรียนของโลก แม้จะผลิตได้จำนวนมาก”

ก. อ้างอิงตัวเลขของ FAO ที่รายงานว่าอินโดนีเซียไม่มีการบริโภคทุเรียนในประเทศ

ข. ระบุข้อมูลเปรียบเทียบมูลค่าการส่งออกทุเรียนของไทย มาเลเซีย เวียดนาม กับอินโดนีเซีย

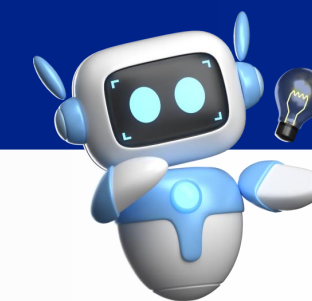
ค. ใช้ตำนานกษัตริย์เสวยทุเรียนเพื่อแสดงอิทธิพลของผู้นำในอดีต

ง. สัมภาษณ์เฉพาะเจ้าหน้าที่จากจีนที่นำเข้าทุเรียนจากไทยเท่านั้น

หมายเหตุ

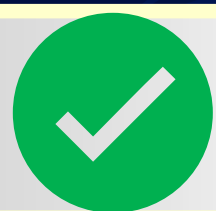
วิจารณญาณ ระบุความสอดคล้องระหว่างข้ออ้างและข้อสรุป

ข้อค้นพบ



03

การสร้างตัวเลือก



ตัวถูก

>> สร้างได้ถูกต้องเป็นส่วนใหญ่



ตัวลวง

>> ส่วนใหญ่ยังไม่เป็นตัวลวงที่มีประสิทธิภาพ
(ผิดชัด, ขัดแย้งตรงข้ามกับเนื้อหา, นอกตัวบท)

>> การอ่านอย่างมีวิจารณญาณ สามารถสร้างตัวเลือกที่เป็นตัวผิดที่ใช้ได้อยู่บ้าง

ตัวอย่างการอ่านจับรายละเอียด : เรือนไทย

GPT-01

ได้ทุนเรือนไทยถูกใช้ประโยชน์ในลักษณะใด ตามเนื้อหาในบทความ

- ก. เป็นพื้นที่จอดรถและประกอบกิจการค้าขาย
- ข. เป็นพื้นที่อเนกประสงค์สำหรับเก็บอุปกรณ์เกษตรหรือเลี้ยงสัตว์
- ค. เป็นห้องนอนหลักสำหรับผู้สูงอายุ
- ง. เป็นสถานที่ประกอบพิธีกรรมทางศาสนา

หมายเหตุ

ตัวผิดทุกตัวใช้คำหรือข้อความที่ไม่มีในบทความ (ตามไฮไลต์สีแดง) ทำให้ผิดชัดเจน

ตัวอย่างการอ่านอย่างมีวิจารณญาณ : เรื่อง กลไกตลาดจะแก้ปัญหาคาสิโนอย่างไร?

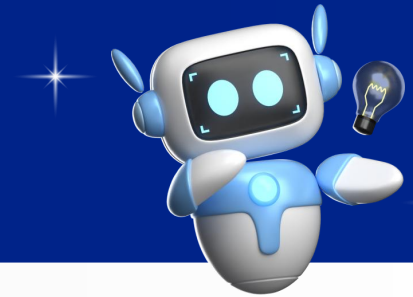
GPT-01

ข้อใดที่เป็น “ข้อคิดเห็น” (opinion) มากกว่า “ข้อเท็จจริง” (fact) ตามเนื้อหาบทความ

- ก. “กบพิษเลห์มันน์ถูกเรียก ‘เพชรของวงการคนรักกบ’”
- ข. “กบพิษเลห์มันน์อาจทำให้คนซื้อรู้สึกไม่ผิดบาปถ้าซื้อมาจากฟาร์ม”
- ค. “กบพิษเลห์มันน์พบได้เฉพาะที่หุบเขาทางตะวันตกของโคลอมเบีย”
- ง. “ราคากบพิษเลห์มันน์ลายทองคาดดำอาจสูงถึง 50,000 บาทต่อตัว”

หมายเหตุ

ตัวถูกเป็นข้อคิดเห็น ส่วนตัวผิดเป็นข้อเท็จจริงจากเรื่องทั้งหมด ทำให้ข้อนี้ถือว่าสร้างตัวเลือกได้ดี



04

การสร้าง agent แก่ตัวเลือกที่ถูก 3 ตัวให้ผิด เพื่อให้กลายเป็นตัวลวง

การอ่านจับใจความและตีความ

- บางครั้งไม่ใช่ตัวเลือกที่ถูกจริงทั้งหมด บางตัวเลือกเป็นตัวเลือกที่เกือบถูกแต่ไม่ถูก จึงกลายเป็นตัวลวงที่ดี ไม่จำเป็นต้องแก้ไขผิดอีก เพราะถ้าแก้ไขผิดจะผิดเกินไป

การอ่านจับรายละเอียด

- เมื่อแก้ไขแล้ว กลายเป็นตัวผิดที่ชัดเจนมากเกินไป ไม่ต่างจากที่ไม่ใช่ agent ในการแก้ไข

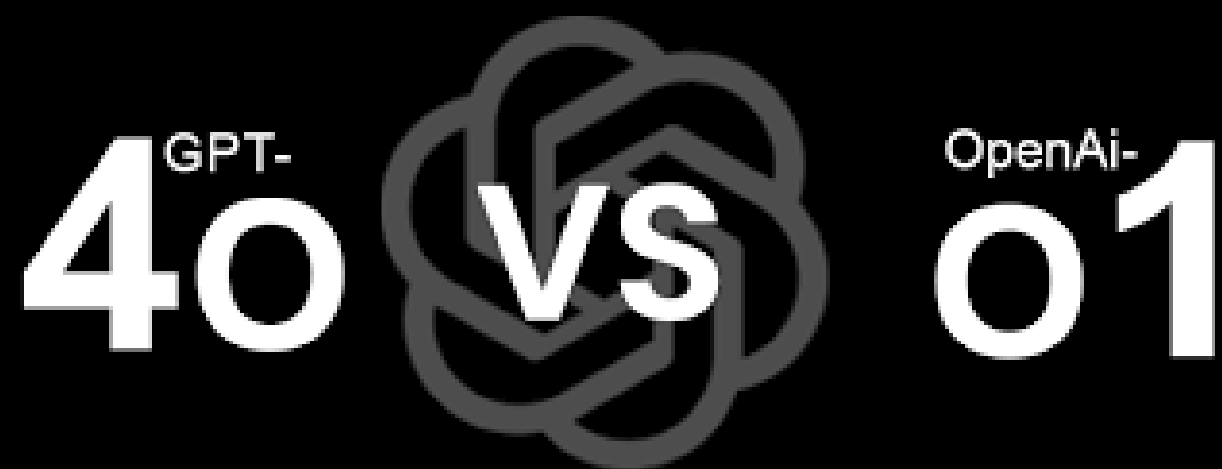
การอ่านอย่างมีวิจารณญาณ

- การให้สร้างตัวถูก 1 ข้อ ตัวผิด 3 ข้อ แบบที่ไม่ต้องใช้ agent มาแก้ไขรอบน่าจะดีกว่า เพราะเมื่อทำให้ถูกทั้ง 4 ข้อ แล้วแก้ไขใหม่ จะทำให้ตัวผิดผิดอย่างชัดเจน



ตัวอย่างจับใจความจากเรื่อง : “ข้าว” กับบทบาทต่อการเติบโตและล่มสลายของรัฐอยุธยา

ตัวถูก 4 ตัว	หลังแก้ไขผิด 3 ตัว	หมายเหตุ
ย่อหน้าที่ 2 ของบทอ่าน (ซึ่งกล่าวถึงอาณาจักรอยุธยากับการตั้งถิ่นฐานบนที่ราบลุ่มแม่น้ำเจ้าพระยา) มีประเด็นหลักเกี่ยวกับเรื่องใด 1. ความอุดมสมบูรณ์ของกลุ่มแม่น้ำเจ้าพระยาที่ส่งเสริมการทำเกษตร 2. บทบาทของข้าวและของป่าในการขับเคลื่อนเศรษฐกิจในอยุธยา 3. การตั้งถิ่นฐานและการขยายตัวของประชากรรอบพื้นที่อุดมสมบูรณ์ 4. พลังทางเศรษฐกิจจากทรัพยากรธรรมชาติที่หล่อเลี้ยงรัฐอยุธยา	ย่อหน้าที่ 2 ของบทอ่าน (ซึ่งกล่าวถึงอาณาจักรอยุธยากับการตั้งถิ่นฐานบนที่ราบลุ่มแม่น้ำเจ้าพระยา) มีประเด็นหลักเกี่ยวกับเรื่องใด 1. (X) อาณาจักรอยุธยามีสภาพภูมิประเทศเป็นภูเขาสูงสลับซับซ้อน 2. (X) ประมงน้ำเค็มและการล่าสัตว์ป่าเป็นกิจกรรมหลักของชุมชน 3. (X) การค้าขายเครื่องเทศกับนานาชาติเป็นรายได้หลักของอยุธยา 4. (✓) ความอุดมสมบูรณ์ทางธรรมชาติเอื้อต่อการขยายตัวของอยุธยา	<u>ก่อนแก้ไข</u> ตัวเลือก 1 และ 4 ถูกต้องจริง แต่ตัวเลือก 2 และ 3 ยังไม่ถูกแต่ใกล้เคียงมาก แต่ถือว่าเป็นตัวผิดที่ดี เพราะดัดแปลงมาจากบทอ่าน <u>หลังแก้ไข</u> ตัวถูกครอบคลุมกว่าตัวถูกก่อนแก้ไข แต่ตัวผิดผิดชัดเจนเกินไป (ตามไฮไลท์สีแดง)



ตัวไหนตั้งคำถามได้ตรงกว่า ?
ตั้งตัวเลือกดีกว่า ?

โดยสรุป

- **GPT-o1**

ตั้งคำถามวิจารณ์ญาณ
และตั้งตัวเลือกของการ
อ่านวิจารณ์ญาณได้ดีกว่า

- **GPT-4o**

ใช้ภาษาในการสร้าง
ข้อสอบได้ดีกว่า

เกณฑ์การประเมิน	GPT-o1	GPT-4o
คำถามจับรายละเอียดตรง	1	1
คำถามจับใจความตรง	1	1
คำถามตีความตรง	1	1
คำถามวิจารณ์ญาณตรง	1	0
เฉลยถูกต้อง	1	1
ตัวลงเหมาะสม	0	0
ภาษากระชับเข้าใจง่าย	0	1
รวม (เต็ม 7)	5	5