SuperAI SS6 Foundation AI Quiz

ถอดข้อสอบ 100 ข้อจาก PDF ภาพ พร้อมเฉลยละเอียดและภาพประกอบโจทย์

Source	Quiz.pdf
Outputs	Markdown + DOCX + question_crops/q001-q100
Reading note	คำตอบที่เลือกในภาพและเฉลยแนะนำถูกแยกไว้ เพื่อไม่ให้สับสนในข้อที่ภาพเลือกผิดหรือสูตรขึ้นกับบริบท

Quick Index

ข้อที่มีภาพประกอบแยก: 20, 74, 94, 95

ข้อที่ควรอ่านหมายเหตุเป็นพิเศษ: 9, 16, 23, 30, 50, 76, 78

ข้อ 1

คำถาม: ทำไม correlation ไม่ใช่ causation

ตัวเลือก

- **อะไรปรากฏคู่ด้วยกัน อาจปรากฏด้วยกันโดยความบังเอิญ [เลือกในภาพ, เฉลย]**
- correlation ดังกล่าวกว่า causation
- เขียนคนละแบบ
- Causation ปรัชญาเกินไป วัดไม่ได้

เลือกในภาพ	อะไรปรากฏคู่ด้วยกัน อาจปรากฏด้วยกันโดยความบังเอิญ
เฉลยแนะนำ	อะไรปรากฏคู่ด้วยกัน อาจปรากฏด้วยกันโดยความบังเอิญ

อธิบาย: Correlation บอกว่าตัวแปรสองตัวเปลี่ยนไปด้วยกัน แต่ยังไม่ได้พิสูจน์ว่าตัวหนึ่งเป็นสาเหตุของอีกตัวหนึ่ง อาจมีตัวแปรแฝงหรือเป็นความบังเอิญได้

ข้อ 2

คำถาม: การเลือก Degree ของ Polynomial จะขึ้นกับข้อมูล แต่หากเลือก Degree ที่สูงเกินไป ซึ่งทำงานได้ดีกับข้อมูล Train แต่ไม่ดีกับข้อมูล Test เรียกเหตุการณ์แบบนี้ว่าอะไร?

ตัวเลือก

- ไม่มีข้อใดถูกต้อง
- **Overfitting [เลือกในภาพ, เฉลย]**
- Just Right (Good fitting)
- Underfitting

เลือกในภาพ	Overfitting
เฉลยแนะนำ	Overfitting

อธิบาย: Overfitting คือโมเดลจำรายละเอียด/noise ของชุด train มากเกินไป ทำให้ generalize ไปยัง test data ได้ไม่ดี

ข้อ 3

คำถาม: ข้อใดต่อไปนี้กล่าวได้ถูกต้องเกี่ยวกับฐานข้อมูล SQL และ NoSQL

ตัวเลือก

- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**
- ฐานข้อมูล SQL เป็นแบบตาราง ในขณะที่ฐานข้อมูล NoSQL เป็น document, key-value, graph หรือ wide-column stores
- ฐานข้อมูล SQL ดีกว่าสำหรับธุรกรรมแบบหลายแถว ในขณะที่ NoSQL ดีกว่าสำหรับข้อมูลที่ไม่มีโครงสร้าง
- ฐานข้อมูล SQL เป็นฐานข้อมูลเชิงสัมพันธ์ ฐานข้อมูล NoSQL ไม่ใช่เชิงสัมพันธ์

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: SQL ใช้โมเดลตารางและ relational schema เหมาะกับธุรกรรมที่ต้องการความสอดคล้องสูง ส่วน NoSQL มีหลายโมเดลและยืดหยุ่นกับข้อมูลกึ่ง/ไม่มีโครงสร้าง

ข้อ 4

คำถาม: หากต้องการพัฒนาระบบเพื่อช่วยคัดกรองอาการผู้ป่วยในโรงพยาบาล คุณคิดว่า named entity ใดเกี่ยวข้องน้อยที่สุด

ตัวเลือก

- ข้อมูลผู้ป่วย
- ชื่อยา
- **ชื่อพืช [เลือกในภาพ, เฉลย]**
- ชื่อโรค

เลือกในภาพ	ชื่อพืช
เฉลยแนะนำ	ชื่อพืช

อธิบาย: บริบทโรงพยาบาลเกี่ยวข้องกับข้อมูลผู้ป่วย ยา และโรคโดยตรง ชื่อพืชอาจเกี่ยวข้องในบางกรณีแต่โดยทั่วไปน้อยที่สุด

ข้อ 5

คำถาม: ข้อใดเป็นเว็บไซต์ของศูนย์กลางข้อมูลเปิดภาครัฐ (Open Government Data) จัดการโดยสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

ตัวเลือก

- data.co.th
- **data.go.th [เลือกในภาพ, เฉลย]**
- data.gov
- data.or.th

เลือกในภาพ	data.go.th
เฉลยแนะนำ	data.go.th

อธิบาย: ศูนย์กลางข้อมูลเปิดภาครัฐของไทยใช้โดเมน data.go.th

ข้อ 6

คำถาม: วิธีการประมวลผลภาพที่กระทำโดยตรงบนพิกเซลของภาพ คือ วิธีการใด

ตัวเลือก

- **Spatial domain [เลือกในภาพ, เฉลย]**
- Transform domain
- Inverse transformation
- ไม่มีข้อใดถูก

เลือกในภาพ	Spatial domain
เฉลยแนะนำ	Spatial domain

อธิบาย: Spatial domain ทำงานโดยตรงกับค่าพิกเซล เช่น filtering/convolution บนภาพ ส่วน transform domain จะทำงานหลังแปลงภาพไปโดเมนอื่น เช่น frequency

ข้อ 7

คำถาม: ข้อใดกล่าวถูกต้อง เกี่ยวกับ Metric ในการวัดประสิทธิภาพของ Regression Model?

ตัวเลือก

- วัดด้วย RMSE (Root Mean Square Error) เพื่อดูว่าค่าที่ทำนายจะเบี่ยงเบนไปเท่าไรโดยเฉลี่ย
- วัดด้วย R^2 (R Square) หากมีค่าใกล้ 1 แสดงว่า Model Fit กับ Data ได้ดี
- ไม่มีข้อใดถูก
- **ถูกทั้งสองข้อ [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ถูกทั้งสองข้อ
เฉลยแนะนำ	ถูกทั้งสองข้อ

อธิบาย: RMSE วัดขนาด error เฉลี่ยในหน่วยเดียวกับ target ส่วน R^2 วัดสัดส่วนความแปรปรวนที่โมเดลอธิบายได้ ค่าใกล้ 1 มักแปลว่า fit ดี

ข้อ 8

คำถาม: ภาษาไทย (Thai) มีระบบการเขียนคล้ายกับภาษาใด?

ตัวเลือก

- ภาษาอาราบิก (Arabic Alphabets)
- ภาษาเกาหลี (Korean - Hangeul)
- **ภาษาพม่า (Burmese) [เลือกในภาพ, เฉลย]**
- ภาษาญี่ปุ่น (Japanese)

เลือกในภาพ	ภาษาพม่า (Burmese)
เฉลยแนะนำ	ภาษาพม่า (Burmese)

อธิบาย: ไทยและพม่าอยู่ในกลุ่มอักษรเชิงพยางค์/abugida ที่มีรากทางประวัติศาสตร์จากตระกูลอักษรพราหมี ต่างจาก Arabic ที่เป็น abjad และ Hangul/Japanese ที่มีกลไกอีกแบบ

ข้อ 9

คำถาม: จากตัวอย่าง System Summary: ฉันไปเข้าค่ายพัฒนานวัตกร และ Reference Summary: ฉันเข้าค่ายนวัตกร จงหาค่า ROUGE-N (F-measure) เมื่อ N=1

ตัวเลือก

- 1.67
- 1.60
- 1.00
- **0.67 [เลือกในภาพ, เฉลย]**

เลือกในภาพ	0.67
เฉลยแนะนำ	0.67

อธิบาย: ตามตัวเลือกของข้อสอบใช้ค่า 0.67 ซึ่งเทียบได้กับ precision ของ unigram overlap เมื่อ system ยาวกว่า reference

หมายเหตุ: หมายเหตุ: ถ้าใช้ ROUGE-1 F1 มาตรฐานและตัดคำเป็น ฉัน|ไป|เข้า|ค่าย|พัฒนา|นวัตกร กับ ฉัน|เข้า|ค่าย|นวัตกร จะได้ P=4/6, R=4/4 และ F1 ประมาณ 0.80 ดังนั้นข้อนี้ควรจำตามบริบทสูตรที่ผู้สอนใช้

ข้อ 10

คำถาม: กฎหมาย PDPA ของประเทศไทยมีผลบังคับใช้อย่างเต็มรูปแบบตั้งแต่เมื่อใด?

ตัวเลือก

- 1 มกราคม 2022
- 1 มกราคม 2020
- 1 มิถุนายน 2023
- **1 มิถุนายน 2022 [เลือกในภาพ, เฉลย]**
- 31 ธันวาคม 2021

เลือกในภาพ	1 มิถุนายน 2022
เฉลยแนะนำ	1 มิถุนายน 2022

อธิบาย: PDPA ของไทยเริ่มบังคับใช้อย่างเต็มรูปแบบวันที่ 1 มิถุนายน 2022 หลังมีการเลื่อนบังคับใช้มาก่อนหน้า

ข้อ 11

คำถาม: ข้อใด คือ เครื่องมือสำเร็จรูปที่ใช้ในการสร้างโมเดล

ตัวเลือก

- Teachable Machine
- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**
- Lobe.ai
- AutoML

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: Teachable Machine, Lobe.ai และ AutoML ล้วนเป็นเครื่องมือช่วยสร้าง/ฝึกโมเดลโดยลดภาระการเขียนโค้ด

ข้อ 12

คำถาม: เมื่อพิจารณาความสามารถในการแข่งขันและการพัฒนาอย่างยั่งยืน (Competitiveness and Sustainability Development) ตามหลักการทางจริยธรรมปัญญาประดิษฐ์ ข้อใดกล่าวไม่ถูกต้อง

ตัวเลือก

- ปัญญาประดิษฐ์ควรสร้างและใช้งานเพื่อสร้างประโยชน์และความผาสุกให้แก่มนุษย์ สังคม เศรษฐกิจ และสิ่งแวดล้อมอย่างยั่งยืน
- **ปัญญาประดิษฐ์ควรนำมาใช้ในการกำหนดชะตาชีวิตของมนุษย์ [เลือกในภาพ, เฉลย]**
- ปัญญาประดิษฐ์ควรใช้งานเพื่อเพิ่มความสามารถในการแข่งขันและสร้างความเจริญให้กับมนุษย์ สังคม ประเทศ ภูมิภาค และโลกอย่างเป็นธรรม
- ปัญญาประดิษฐ์ควรได้รับการวิจัยและพัฒนาอย่างต่อเนื่อง เพื่อให้มนุษย์เกิดการสร้างสรรค์นวัตกรรมและอุตสาหกรรมใหม่

เลือกในภาพ	ปัญญาประดิษฐ์ควรนำมาใช้ในการกำหนดชะตาชีวิตของมนุษย์
เฉลยแนะนำ	ปัญญาประดิษฐ์ควรนำมาใช้ในการกำหนดชะตาชีวิตของมนุษย์

อธิบาย: AI ควรสนับสนุนการตัดสินใจและความยั่งยืน แต่ไม่ควรถูกใช้เป็นผู้กำหนดชะตาชีวิตมนุษย์โดยปราศจากหลักสิทธิ ความรับผิดชอบ และการกำกับดูแล

ข้อ 13

คำถาม: ข้อใดกล่าวถูกต้อง เกี่ยวกับ Linear Regression และ Polynomial Regression?

ตัวเลือก

- Polynomial Regression ใช้สมการ Polynomial Degree ต่าง ๆ ในการ Fit กับ Data
- ในการใช้ Polynomial Regression ค่า Degree ที่เหมาะสมจะขึ้นกับข้อมูลแต่ละชุด ต้องทำการทดลองโดยกระบวนการ Train และ Test
- Linear Regression ใช้สมการเส้นตรงในการ Fit กับ Data
- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: ทั้งสามข้อความถูกต้อง: linear regression fit เส้นตรง ส่วน polynomial regression เพิ่มพจน์กำลัง และ degree ต้องเลือกจากผล train/test

ข้อ 14

คำถาม: ในระบบ RAG ถ้าขั้นตอน Retrieval ค้นหาเอกสารที่ไม่เกี่ยวข้องกับคำถามมาให้ LLM จะเกิดอะไรขึ้น?

ตัวเลือก

- ระบบจะทำ Retrieval ซ้ำอัตโนมัติจนกว่าจะได้ข้อมูลที่ถูกต้อง
- LLM จะตรวจจับได้เองว่าข้อมูลไม่เกี่ยวข้อง และปฏิเสธตอบ
- LLM จะสร้างข้อมูลที่ไม่เกี่ยวข้องและตอบจากความรู้ของตัวเองแทน
- **LLM จะนำข้อมูลที่ไม่เกี่ยวข้องนั้นไปสร้างคำตอบที่ผิดพลาดให้ผู้ใช้ [เลือกในภาพ, เฉลย]**
- ระบบจะหยุดทำงานและแจ้ง Error

เลือกในภาพ	LLM จะนำข้อมูลที่ไม่เกี่ยวข้องนั้นไปสร้างคำตอบที่ผิดพลาดให้ผู้ใช้
เฉลยแนะนำ	LLM จะนำข้อมูลที่ไม่เกี่ยวข้องนั้นไปสร้างคำตอบที่ผิดพลาดให้ผู้ใช้

อธิบาย: RAG พึ่งคุณภาพ retrieval มาก หาก context ผิดหรือไม่เกี่ยวข้อง LLM อาจตอบตามข้อมูลผิดและเกิด hallucination ได้

ข้อ 15

คำถาม: WangchanBERTa โมเดลประมวลผลภาษาไทยที่ใหญ่และก้าวหน้าที่สุดในขณะนั้น พัฒนาโดย VISTEC-depa THAILAND AI RESEARCH INSTITUTE ที่ปล่อยโมเดลออกเมื่อต้นปี 2021 มีพื้นฐานมาจากสถาปัตยกรรมอะไร

ตัวเลือก

- CAMEMBERT
- ALBERT
- **ROBERTA [เลือกในภาพ, เฉลย]**
- BERT

เลือกในภาพ	ROBERTA
เฉลยแนะนำ	ROBERTA

อธิบาย: WangchanBERTa อิงแนวทาง RoBERTa ซึ่งเป็นการปรับปรุงจาก BERT ผ่านวิธี pretraining ที่เหมาะขึ้น

ข้อ 16

คำถาม: Thai elementary discourse unit (TEDU) มุ่งเน้นการจัดโครงสร้างในลักษณะใด?

ตัวเลือก

- **มุ่งเน้นคำกริยาต้องปรากฏในหน่วยนั้นเสมอ [เฉลย]**
- มุ่งเน้นความสำคัญที่ประธานของโครงสร้าง ต้องมีคำนามประกอบเสมอ
- มุ่งเน้นคำกริยาและประธานของหน่วยเสมอ [เลือกในภาพ]
- มุ่งเน้นคำนามที่เป็นกรรมของหน่วย รวมไปถึงวลี

เลือกในภาพ	มุ่งเน้นคำกริยาและประธานของหน่วยเสมอ
เฉลยแนะนำ	มุ่งเน้นคำกริยาต้องปรากฏในหน่วยนั้นเสมอ

อธิบาย: นิยาม TEDU ของไทยเน้นหน่วยที่มี verbal unit/verb phrase เป็นแกน เพราะภาษาไทยละประธานได้บ่อย จึงไม่ควรบังคับว่าต้องมีประธานเสมอ

หมายเหตุ: คำตอบที่เลือกในภาพต่างจากเฉลยตามหลักนิยาม TEDU

ข้อ 17

คำถาม: แนวคิด 'Baseline First' ในการทำ Prototype หมายถึงสิ่งใด?

ตัวเลือก

- การทำแบบสำรวจความพึงพอใจก่อนเริ่มทำโปรแกรม
- การจ้างบริษัทภายนอกมาทำระบบทั้งหมดให้เสร็จในครั้งเดียว
- **การเริ่มจากวิธีที่ง่ายและเร็วที่สุด เช่น Rule-based เพื่อพิสูจน์แนวคิด [เลือกในภาพ, เฉลย]**
- การซื้อซอฟต์แวร์สำเร็จรูปจากต่างประเทศมาใช้งานทันที
- การเริ่มสร้างระบบด้วยโมเดล Deep Learning ที่ซับซ้อนที่สุด

เลือกในภาพ	การเริ่มจากวิธีที่ง่ายและเร็วที่สุด เช่น Rule-based เพื่อพิสูจน์แนวคิด
เฉลยแนะนำ	การเริ่มจากวิธีที่ง่ายและเร็วที่สุด เช่น Rule-based เพื่อพิสูจน์แนวคิด

อธิบาย: Baseline first ช่วยสร้างจุดเทียบและพิสูจน์ feasibility เร็ว ก่อนลงทุนกับโมเดลซับซ้อน

ข้อ 18

คำถาม: ข้อใด ไม่ใช่ โจทย์แบบ Regression

ตัวเลือก

- การทำนายราคาบ้าน
- การทำนายจำนวนเงินที่ลูกค้าจะใช้ในห้างสรรพสินค้า
- **การทำนายว่าลูกค้าจะย้ายค่ายโทรศัพท์มือถือ หรือไม่ [เลือกในภาพ, เฉลย]**
- การทำนายราคาหุ้น

เลือกในภาพ	การทำนายว่าลูกค้าจะย้ายค่ายโทรศัพท์มือถือ หรือไม่
เฉลยแนะนำ	การทำนายว่าลูกค้าจะย้ายค่ายโทรศัพท์มือถือ หรือไม่

อธิบาย: โจทย์ yes/no เป็น classification ส่วนราคาบ้าน จำนวนเงิน และราคาหุ้นเป็นค่าต่อเนื่องจึงเป็น regression

ข้อ 19

คำถาม: ข้อใดกล่าวถูกต้องในความเท่าเทียม หลากหลาย ครอบคลุม และเป็นธรรม (Fairness) ตามหลักการทางจริยธรรมปัญญาประดิษฐ์

ตัวเลือก

- การตัดสินใจที่เกี่ยวข้องกับการวิจัย ออกแบบ พัฒนา ให้บริการ และใช้งานปัญญาประดิษฐ์ควรสามารถพิสูจน์ถึงความเป็นธรรมได้
- การออกแบบและพัฒนาปัญญาประดิษฐ์ควรคำนึงถึงความหลากหลาย หลีกเลี่ยงการผูกขาด ลดการแบ่งแยกและเอนเอียง
- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**
- ควรพัฒนาปัญญาประดิษฐ์ให้ก่อให้เกิดประโยชน์ต่อผู้คนจำนวนมากเท่าที่จะทำได้ โดยเฉพาะกลุ่มคนผู้ด้อยโอกาสในสังคม

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: Fairness ครอบคลุมการลด bias การพิสูจน์ความเป็นธรรม การคำนึงถึงความหลากหลาย และการไม่ทิ้งกลุ่มเปราะบาง

ข้อ 20

คำถาม: จากข้อมูลในตาราง ค่าใช้จ่ายทั้งหมดของรายการเหล่านี้ในปี 2017 คิดเป็นร้อยละของค่าใช้จ่ายทั้งหมดในปี 2021 โดยประมาณกี่เปอร์เซ็นต์

ภาพประกอบข้อ 20

ปี	ค่าใช้จ่ายของบริษัท ABC				
	เงินเดือน	ค่าเดินทาง	ดอกเบี้ยเงินกู้	โบนัส	ภาษี
2017	288	98	23.4	3.00	83
2018	342	112	32.5	2.52	108
2019	324	101	41.6	3.84	74
2020	336	133	36.4	3.68	88
2021	420	142	49.4	3.96	98

ตัวเลือก

- 66%
- **69% [เลือกในภาพ, เฉลย]**
- 62%
- 71%

เลือกในภาพ	69%

เฉลยแนะนำ	69%

อธิบาย: ปี 2017 รวม 288+98+23.4+3+83 = 495.4 ส่วนปี 2021 รวม 420+142+49.4+3.96+98 = 713.36 ดังนั้น 495.4/713.36 ประมาณ 69.4%

ข้อ 21

คำถาม: พิกเซลหนึ่งของภาพ เก็บข้อมูลแบบ float64 โดยมีค่าเท่ากับ 0.27745873 หากนำข้อมูลนี้แปลงเป็น uint8 จะได้คำตอบตรงกับข้อใด

ตัวเลือก

- **71 [เลือกในภาพ, เฉลย]**
- 70
- 17
- 0

เลือกในภาพ	71
เฉลยแนะนำ	71

อธิบาย: เมื่อ float อยู่ช่วง 0 ถึง 1 และแปลงเป็นภาพ 8-bit ให้คูณ 255: 0.27745873 x 255 ประมาณ 70.75 แล้วปัดเป็น 71

ข้อ 22

คำถาม: มีภาษาสองภาษาคือ L1 = {a^m b^n | 0 < m <= n} และ L2 = {a^m b^n | 0 < n <= m} ภาษาใดต่อไปนี้เท่ากับ L1 ∩ L2

ตัวเลือก

- {a^m b^n | 0 < m < n}
- **{a^n b^n | n > 0} [เลือกในภาพ, เฉลย]**
- {a^m b^n | m > 0, n > 0}
- {a^n b^n | n >= 0}

เลือกในภาพ	{a^n b^n | n > 0}
เฉลยแนะนำ	{a^n b^n | n > 0}

อธิบาย: จุดร่วมของ m <= n และ n <= m คือ m = n และทั้งคู่ต้องมากกว่า 0 จึงได้ a^n b^n เมื่อ n > 0

ข้อ 23

คำถาม: ข้อใด เป็น ชื่อเรียกของประเภทการเขียนแบบ Syllabic Alphabets?

ตัวเลือก

- Alphabets
- Abjads
- Syllabaries [เลือกในภาพ]
- **Abugidas [เฉลย]**

เลือกในภาพ	Syllabaries
เฉลยแนะนำ	Abugidas

อธิบาย: Syllabic alphabet หรือ alphasyllabary โดยทั่วไปหมายถึง abugida ซึ่งหน่วยหลักเป็นพยัญชนะพร้อมสระโดยกำเนิดและเปลี่ยนสระด้วยเครื่องหมาย

หมายเหตุ: คำตอบที่เลือกในภาพเป็น Syllabaries แต่ตามนิยาม writing system ควรเป็น Abugidas

ข้อ 24

คำถาม: ข้อใดไม่ถูกต้องเกี่ยวกับ SVD?

ตัวเลือก

- SVD สามารถใช้ในการหาองค์ประกอบหลักของข้อมูลได้
- **SVD สามารถใช้ในการหาเมทริกซ์ผกผันได้เร็วกว่าวิธีการ QR decomposition [เลือกในภาพ, เฉลย]**
- SVD สามารถใช้ในการลดมิติของข้อมูลได้
- SVD สามารถใช้ในการหาเมทริกซ์ผกผันของเมทริกซ์ได้โดยไม่ต้องกังวลเรื่อง Ill-conditioned matrix

เลือกในภาพ	SVD สามารถใช้ในการหาเมทริกซ์ผกผันได้เร็วกว่าวิธีการ QR decomposition
เฉลยแนะนำ	SVD สามารถใช้ในการหาเมทริกซ์ผกผันได้เร็วกว่าวิธีการ QR decomposition

อธิบาย: SVD มีความเสถียรและใช้กับ PCA/dimensionality reduction/pseudoinverse ได้ดี แต่โดยทั่วไปแพงกว่า QR จึงไม่ใช่วิธีที่เร็วกว่าเสมอ

ข้อ 25

คำถาม: RAG ย่อมาจากอะไร?

ตัวเลือก

- **Retrieval Augmented Generation [เลือกในภาพ, เฉลย]**
- Resource Allocation Generator
- Retrieval Augmented Graphics
- Rapid Accelerated Generation
- Real-time Analysis Gateway

เลือกในภาพ	Retrieval Augmented Generation
เฉลยแนะนำ	Retrieval Augmented Generation

อธิบาย: RAG คือการดึงข้อมูลที่เกี่ยวข้องมาเติม context ให้โมเดลก่อน generation

ข้อ 26

คำถาม: ข้อมูลใดเป็นข้อมูลที่เกี่ยวข้องกับตำแหน่งที่ตั้งของข้อมูลต่าง ๆ บนพื้นโลก

ตัวเลือก

- IP Connection Data
- **Geospatial Data [เลือกในภาพ, เฉลย]**
- Metadata
- Internet of Things Data

เลือกในภาพ	Geospatial Data
เฉลยแนะนำ	Geospatial Data

อธิบาย: Geospatial data คือข้อมูลเชิงพื้นที่/ตำแหน่ง เช่น พิกัด แผนที่ เขตพื้นที่ และระยะทางบนพื้นโลก

ข้อ 27

คำถาม: ระบบวิเคราะห์และจัดระดับความยากง่ายของข้อความสำหรับเอกสารภาษาไทย ควรดูจากอะไรได้บ้าง: 1) ประเภทของคำ เช่น คำไทยแท้ คำยืม 2) ชนิดของประโยค เช่น ประโยคความเดียว ความรวม ความซ้อน 3) จำนวนคำในประโยค

ตัวเลือก

- ถูก เฉพาะข้อ 2 กับ 3
- ถูก เฉพาะข้อ 1
- ถูก เฉพาะข้อ 2
- **ถูก เฉพาะข้อ 1, 2, 3 [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ถูก เฉพาะข้อ 1, 2, 3
เฉลยแนะนำ	ถูก เฉพาะข้อ 1, 2, 3

อธิบาย: ความยากของข้อความสัมพันธ์ทั้งระดับคำ โครงสร้างประโยค และความยาว/จำนวนคำ

ข้อ 28

คำถาม: Do calculus มีประโยชน์อะไร

ตัวเลือก

- ช่วยบอกว่า data นี้ซับซ้อน
- เอาไว้หา model
- เอาไว้แก้โจทย์ calculus
- **ช่วยบอกว่า บางครั้งเราไม่ต้องทำการทดลองจริง สามารถพูดถึง causation จาก data ได้เลย [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ช่วยบอกว่า บางครั้งเราไม่ต้องทำการทดลองจริง สามารถพูดถึง causation จาก data ได้เลย
เฉลยแนะนำ	ช่วยบอกว่า บางครั้งเราไม่ต้องทำการทดลองจริง สามารถพูดถึง causation จาก data ได้เลย

อธิบาย: Do-calculus เป็นเครื่องมือใน causal inference
เพื่อแปลง/คำนวณผลเชิงสาเหตุจากข้อมูลเชิงสังเกตภายใต้สมมติฐานเชิงกราฟบางอย่าง

ข้อ 29

คำถาม: ข้อใดกล่าวถูกต้อง เกี่ยวกับ Regression และ Classification?

ตัวเลือก

- Regression มี Target เป็นหมวดหมู่ (Classes) แต่ Classification มี Target เป็นตัวเลข
- **Regression มี Target เป็นตัวเลข แต่ Classification มี Target เป็นหมวดหมู่ (Classes) [เลือกในภาพ, เฉลย]**
- ทั้ง Regression และ Classification มี Target เป็นตัวเลข
- ทั้ง Regression และ Classification มี Target เป็นหมวดหมู่ (Classes)

เลือกในภาพ	Regression มี Target เป็นตัวเลข แต่ Classification มี Target เป็นหมวดหมู่ (Classes)
เฉลยแนะนำ	Regression มี Target เป็นตัวเลข แต่ Classification มี Target เป็นหมวดหมู่ (Classes)

อธิบาย: Regression ทำนายค่าต่อเนื่อง ส่วน classification ทำนาย class/label

ข้อ 30

คำถาม: Google Translation เป็นผลลัพธ์ที่เกิดจากการใช้วิธีการใดเป็นสำคัญ?

ตัวเลือก

- Extraction-based Approach
- **Abstraction-based Approach [เฉลย]**
- Aided Summarization
- Question-Answering Technique [เลือกในภาพ]

เลือกในภาพ	Question-Answering Technique
เฉลยแนะนำ	Abstraction-based Approach

อธิบาย: การแปลภาษาเป็นงาน generation ที่สร้างข้อความใหม่ในภาษาเป้าหมาย ไม่ใช่การตอบคำถามหรือการดึงข้อความเดิมแบบ extractive

หมายเหตุ: คำตอบที่เลือกในภาพไม่ตรงกับหลัก NLP ทั่วไป

ข้อ 31

คำถาม: ข้อมูลชนิดใดที่สามารถนำไปสร้างเป็นภาพได้

ตัวเลือก

- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**
- UINT8
- Float64
- Float32

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: ภาพสามารถเก็บเป็น uint8 หรือ float ได้หลายชนิด เพียงต้องเข้าใจช่วงค่าและการแปลงก่อนแสดงผล

ข้อ 32

คำถาม: สูตรการคำนวณ ROI เบื้องต้นสำหรับโปรเจกต์ AI ตามที่ปรากฏในเนื้อหาคืออะไร?

ตัวเลือก

- ROI = Cost/Benefit
- ROI = Benefit x Cost
- **ROI = (Benefit - Cost)/Cost [เลือกในภาพ, เฉลย]**
- ROI = Benefit - Cost
- ROI = (Reach x Impact)/Effort

เลือกในภาพ	ROI = (Benefit - Cost)/Cost
เฉลยแนะนำ	ROI = (Benefit - Cost)/Cost

อธิบาย: ROI วัดผลตอบแทนเทียบต้นทุน สูตรพื้นฐานคือผลประโยชน์สุทธิต่อต้นทุน

ข้อ 33

คำถาม: วิธีการประมวลผลใด สามารถนำมาใช้ในการหาขนาดของวัตถุในภาพได้

ตัวเลือก

- Histogram equalization
- Opening และ Closing
- Histogram Processing
- **Contour [เลือกในภาพ, เฉลย]**

เลือกในภาพ	Contour
เฉลยแนะนำ	Contour

อธิบาย: Contour ใช้หาเส้นขอบ/พื้นที่ของวัตถุ จากนั้นคำนวณขนาด พื้นที่ หรือ bounding box ได้

ข้อ 34

คำถาม: กระบวนการทำ Data Science ที่ถูกต้องควรเป็นอย่างไร?

ตัวเลือก

- เริ่มต้นด้วยการวิเคราะห์ข้อมูล ก่อนทำความสะอาดชุดข้อมูล และสร้างรายงานผลลัพธ์
- **รวบรวมข้อมูลดิบ ก่อน ทำความสะอาดชุดข้อมูล วิเคราะห์ข้อมูล และสร้างรายงานผลลัพธ์ [เลือกในภาพ, เฉลย]**
- ทำความสะอาดชุดข้อมูล ก่อนรวบรวมข้อมูลดิบ วิเคราะห์ข้อมูล และสร้างรายงานผลลัพธ์
- สร้างรายงานผลลัพธ์ ก่อนทำความสะอาดชุดข้อมูล วิเคราะห์ข้อมูล และรวบรวมข้อมูลดิบ

เลือกในภาพ	รวบรวมข้อมูลดิบ ก่อน ทำความสะอาดชุดข้อมูล วิเคราะห์ข้อมูล และสร้างรายงานผลลัพธ์
เฉลยแนะนำ	รวบรวมข้อมูลดิบ ก่อน ทำความสะอาดชุดข้อมูล วิเคราะห์ข้อมูล และสร้างรายงานผลลัพธ์

อธิบาย: ลำดับพื้นฐานคือ collect raw data -> clean/preprocess -> analyze/model -> visualize/report

ข้อ 35

คำถาม: จากสมการถดถอย yhat = -0.0127 + 0.0180x หาก Bobby ดื่มเบียร์ 4 ขวด และค่าจริงของแอลกอฮอล์ในเลือดคือ 0.085 residual ของ Bobby คือเท่าไร

ตัวเลือก

- +0.005
- -0.0257
- **+0.0257 [เลือกในภาพ, เฉลย]**
- -0.005

เลือกในภาพ	+0.0257
เฉลยแนะนำ	+0.0257

อธิบาย: ค่าทำนาย = -0.0127 + 0.0180(4) = 0.0593 ดังนั้น residual = y - yhat = 0.085 - 0.0593 = 0.0257

ข้อ 36

คำถาม: ถ้าโมเดล regression ทำนายข้อมูลที่ใช้เทรนได้แม่นยำ แต่ทำนายข้อมูลทดสอบไม่ดี ควรทำอย่างไร

ตัวเลือก

- a) เลือก feature ที่จำเป็นเท่านั้น
- b) ใส่ nonlinear feature เพิ่ม
- c) เปลี่ยนไปใช้ L1 Loss
- d) ใช้ L2 regularization
- **a) และ d) [เลือกในภาพ, เฉลย]**
- c) และ d)

เลือกในภาพ	a) และ d)
เฉลยแนะนำ	a) และ d)

อธิบาย: อาการนี้คือ overfitting ควรลดความซับซ้อน/เลือก feature ที่จำเป็น และใช้ regularization เช่น L2

ข้อ 37

คำถาม: บทบาทใดใน Governance Roles ที่ทำหน้าที่เป็นเจ้าของผลลัพธ์ (Outcome) และเป็นผู้รับผิดชอบหลักของ Use Case นั้น ๆ?

ตัวเลือก

- Data Steward
- **Business Owner [เลือกในภาพ, เฉลย]**
- AI Engineer
- Executive Sponsor
- Legal Advisor

เลือกในภาพ	Business Owner
เฉลยแนะนำ	Business Owner

อธิบาย: Business Owner เป็นเจ้าของผลลัพธ์ทางธุรกิจและรับผิดชอบ use case ในภาพรวม

ข้อ 38

คำถาม: Amazon เก็บข้อมูลผู้บริโภคจำนวนมหาศาลเพื่อวิเคราะห์พฤติกรรมการซื้อและค้นหาสินค้า โมเดลข้อมูลใดต่อไปนี้ไม่ใช่โมเดลหลักที่ Amazon ใช้เพื่อวิเคราะห์พฤติกรรมผู้บริโภค?

ตัวเลือก

- Product Data Model
- Customer Data Model
- Search Data Model
- **Delivery Data Model [เลือกในภาพ, เฉลย]**

เลือกในภาพ	Delivery Data Model
เฉลยแนะนำ	Delivery Data Model

อธิบาย: Product, customer และ search data เกี่ยวข้องโดยตรงกับพฤติกรรมซื้อ/ค้นหา ส่วน delivery data เน้นโลจิสติกส์หลังการซื้อ

ข้อ 39

คำถาม: ในขั้นตอนประมวลผลข้อมูลเบื้องต้นสำหรับข้อความภาษาไทย
ขั้นตอนสำคัญที่ตัดแบ่งข้อความที่ได้รับออกเป็นคำหรือส่วนย่อย ๆ เพื่อนำไปประมวลผลต่อ คือข้อใด?

ตัวเลือก

- **Tokenization [เลือกในภาพ, เฉลย]**
- Stemming
- Lemmatization
- Vectorization

เลือกในภาพ	Tokenization
เฉลยแนะนำ	Tokenization

อธิบาย: Tokenization คือการแบ่งข้อความเป็น token เช่น คำหรือหน่วยย่อย ก่อนนำไปทำ embedding/vectorization หรือวิเคราะห์ต่อ

ข้อ 40

คำถาม: หน้าที่หลักของ Champions Network ในช่วง Rollout ระบบ AI คืออะไร?

ตัวเลือก

- การเป็นผู้ถือสิทธิ์ Admin สูงสุดเพื่อลบข้อมูลที่ผิดพลาด
- การตัดสินใจอนุมัติงบประมาณในเฟสถัดไป
- การทำหน้าที่เป็นพรีเซนเตอร์ในโฆษณาของบริษัท
- **การเป็นผู้ใช้งานแกนนำที่ช่วยสอนเพื่อนร่วมงานและสะท้อนปัญหาจริง [เลือกในภาพ, เฉลย]**
- การเขียนโปรแกรมเพื่อแก้ไขบั๊กของระบบ

เลือกในภาพ	การเป็นผู้ใช้งานแกนนำที่ช่วยสอนเพื่อนร่วมงานและสะท้อนปัญหาจริง
เฉลยแนะนำ	การเป็นผู้ใช้งานแกนนำที่ช่วยสอนเพื่อนร่วมงานและสะท้อนปัญหาจริง

อธิบาย: Champions ช่วยให้การนำระบบไปใช้จริงเกิด adoption โดยเป็นผู้สนับสนุน/ให้ feedback จากหน้างาน

ข้อ 41

คำถาม: ฟังก์ชันที่ใช้ในการแสดงรูปภาพ คือฟังก์ชันใด

ตัวเลือก

- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**
- ฟังก์ชัน cv2.imshow()
- ฟังก์ชัน plt.imshow()
- ฟังก์ชัน Image.show()

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: OpenCV, Matplotlib และ PIL ต่างมีฟังก์ชันแสดงภาพ แต่พฤติกรรมและสภาพแวดล้อมที่รองรับต่างกัน

ข้อ 42

คำถาม: ทำไมการทดสอบ AI จึงต้องวัดทั้งในส่วนของ AI Output และ User Experience?

ตัวเลือก

- เพื่อให้นักพัฒนาและนักออกแบบไม่ต้องคุยกัน
- เพราะกฎหมายบังคับให้ต้องส่งรายงานสองฉบับ
- เพราะเป็นวิธีเดียวที่จะทำให้ AI เรียนรู้ได้ด้วยตัวเอง
- เพื่อลดจำนวนพนักงานในองค์กรลงครึ่งหนึ่ง
- **เพื่อให้มั่นใจว่าคำตอบถูกต้องและผู้ใช้สามารถนำไปใช้งานต่อได้จริง [เลือกในภาพ, เฉลย]**

เลือกในภาพ	เพื่อให้มั่นใจว่าคำตอบถูกต้องและผู้ใช้สามารถนำไปใช้งานต่อได้จริง
เฉลยแนะนำ	เพื่อให้มั่นใจว่าคำตอบถูกต้องและผู้ใช้สามารถนำไปใช้งานต่อได้จริง

อธิบาย: AI ที่ output ดีแต่ใช้ยากอาจล้มเหลวในการใช้งานจริง เช่นเดียวกับ UX ดีแต่ output ผิดก็เสี่ยงต่อผู้ใช้

ข้อ 43

คำถาม: ข้อใดถูกต้องเกี่ยวกับ Validation set?

ตัวเลือก

- Validation set เป็นชุดข้อมูลที่ไม่เคยถูกใช้ในกระบวนการฝึกและทดสอบโมเดลเลย
- Validation set เป็นชุดข้อมูลที่ใช้สำหรับฝึกโมเดลเพื่อปรับ Learning parameter
- Validation set เป็นชุดข้อมูลที่ใช้เพื่อทดสอบโมเดลที่แม่นยำกว่า Test set
- **Validation set ใช้ในการปรับค่า Hyperparameters ของโมเดล [เลือกในภาพ, เฉลย]**

เลือกในภาพ	Validation set ใช้ในการปรับค่า Hyperparameters ของโมเดล
เฉลยแนะนำ	Validation set ใช้ในการปรับค่า Hyperparameters ของโมเดล

อธิบาย: Training set ใช้เรียนพารามิเตอร์ ส่วน validation ใช้เลือก hyperparameters/model setting ก่อนวัดผลสุดท้ายบน test set

ข้อ 44

คำถาม: อะไรคือ Data science ตามหลักวิชาการ

ตัวเลือก

- Buzzword
- **การค้นหาองค์ความรู้ จากข้อมูล [เลือกในภาพ, เฉลย]**
- Big data
- Machine learning

เลือกในภาพ	การค้นหาองค์ความรู้ จากข้อมูล
เฉลยแนะนำ	การค้นหาองค์ความรู้ จากข้อมูล

อธิบาย: Data science เป็นกระบวนการใช้สถิติ คอมพิวเตอร์ และ domain knowledge เพื่อสกัด insight/knowledge จากข้อมูล

ข้อ 45

คำถาม: ข้อใด เป็น Machine learning approach

ตัวเลือก

- A* algorithm
- **Linear Regression [เลือกในภาพ, เฉลย]**
- ANOVA
- Bootstrapping

เลือกในภาพ	Linear Regression
เฉลยแนะนำ	Linear Regression

อธิบาย: Linear regression เป็น supervised learning model ส่วน A* เป็น search algorithm, ANOVA/bootstrapping เป็นสถิติ

ข้อ 46

คำถาม: ข้อใดไม่ถูกต้องเกี่ยวกับ K-mean

ตัวเลือก

- a) ต้องกำหนดค่า K ก่อนการใช้งาน
- **b) ถ้าใช้ข้อมูลชุดเดียวกัน จะได้คำตอบเดิมเสมอ [เลือกในภาพ, เฉลย]**
- c) เป็น model ประเภท unsupervised จึงไม่ต้องมี label ก็ได้
- d) ไม่เหมาะกับข้อมูลที่มี feature จำนวนมาก ๆ
- e) ถ้าข้อมูลมีการกระจายในแต่ละ feature ไม่เท่ากัน ควรปรับให้การกระจายเท่า ๆ กันก่อนใช้งาน
- ข้อ c) และ d) ไม่ถูกต้อง

เลือกในภาพ	b) ถ้าใช้ข้อมูลชุดเดียวกัน จะได้คำตอบเดิมเสมอ
เฉลยแนะนำ	b) ถ้าใช้ข้อมูลชุดเดียวกัน จะได้คำตอบเดิมเสมอ

อธิบาย: K-means มักมี random initialization ดังนั้นใช้ dataset เดิมแต่ seed/initial centroid ต่างกันอาจได้ผลต่างกัน

ข้อ 47

คำถาม: การเปรียบเทียบในข้อใดต่อไปนี้ให้ผลลัพธ์เป็น True?

ตัวเลือก

- 'a' != 'A'
- 'a' == 'A'
- **'a' > 'A' [เลือกในภาพ, เฉลย]**
- 'a' < 'A'

เลือกในภาพ	'a' > 'A'
เฉลยแนะนำ	'a' > 'A'

อธิบาย: ถ้าเปรียบเทียบตามรหัสอักขระ ASCII/Unicode ตัวพิมพ์เล็ก 'a' มีค่า 97 และ 'A' มีค่า 65 จึงมากกว่า

ข้อ 48

คำถาม: Model คืออะไร

ตัวเลือก

- **สิ่งที่ใช้อธิบายปรากฏการณ์ ธรรมชาติ ระบบ ฯลฯ [เลือกในภาพ, เฉลย]**
- Gunpla เท่านั้น
- กฎธรรมชาติเท่านั้น
- สมการคณิตศาสตร์เท่านั้น

เลือกในภาพ	สิ่งที่ใช้อธิบายปรากฏการณ์ ธรรมชาติ ระบบ ฯลฯ
เฉลยแนะนำ	สิ่งที่ใช้อธิบายปรากฏการณ์ ธรรมชาติ ระบบ ฯลฯ

อธิบาย: Model คือ representation หรือแบบจำลองของสิ่งจริง อาจเป็นสมการ กฎ รูปภาพ โปรแกรม หรือ conceptual model ก็ได้

ข้อ 49

คำถาม: ข้อดีของ ReLU ที่นำมาใช้ใน Deep learning คือข้อใด?

ตัวเลือก

- ทำให้ Deep learning เป็น Universal approximator เนื่องจากมีผลลัพธ์เป็นรูป S curve
- สามารถปรับค่า Learning rate ได้อัตโนมัติ
- **ReLU ช่วยลดปัญหา Gradient Vanishing ในการฝึกโมเดล [เลือกในภาพ, เฉลย]**
- ความสามารถในการแปลงค่าบวกเป็น 0

เลือกในภาพ	ReLU ช่วยลดปัญหา Gradient Vanishing ในการฝึกโมเดล
เฉลยแนะนำ	ReLU ช่วยลดปัญหา Gradient Vanishing ในการฝึกโมเดล

อธิบาย: ReLU มี gradient คงที่ 1 เมื่อ input เป็นบวก จึงช่วยให้ gradient ไหลผ่านเครือข่ายลึกได้ดีกว่า sigmoid/tanh ในหลายกรณี

ข้อ 50

คำถาม: ข้อใดถูกต้องเกี่ยวกับ Bias Variance ใน classical ML

ตัวเลือก

- เราสามารถลด Bias และ Variance ของโมเดลได้พร้อม ๆ กัน
- การใช้ deep learning ทำให้ไม่ต้องสนใจ bias และ variance ของโมเดล
- ถ้า Bias สูง ควรไปเก็บข้อมูลมาทดสอบเพิ่ม [เลือกในภาพ]
- ถ้า Bias มาก แสดงว่าโมเดล Overfit
- **โมเดลยิ่งซับซ้อน ยิ่งมี Variance สูง [เฉลย]**
- ถ้า Bias สูง ควรไปเก็บข้อมูลเทรนนิ่งมาเพิ่ม

เลือกในภาพ	ถ้า Bias สูง ควรไปเก็บข้อมูลมาทดสอบเพิ่ม
เฉลยแนะนำ	โมเดลยิ่งซับซ้อน ยิ่งมี Variance สูง

อธิบาย: โมเดลซับซ้อนขึ้นมักลด bias แต่เพิ่ม variance; high bias คือ underfitting และควรเพิ่มความสามารถของโมเดล/feature ไม่ใช่เพิ่ม test data

หมายเหตุ: คำตอบที่เลือกในภาพไม่ตรงกับหลัก bias-variance

ข้อ 51

คำถาม: จงพิจารณาภาษาดังต่อไปนี้ L1={a^m b^n | m>0,n>0}, L2={a^n b^n c^n d^n | n>0}, L3={a^m b^n c^m | m>0,n>0}, L4={a^n b^n c^n | n>0}

ตัวเลือก

- L1 < L2 < L3 < L4
- **L1 < L3 < L4 < L2 [เลือกในภาพ, เฉลย]**
- L1 < L4 < L3 < L2
- L1 < L3 < L2 < L4

เลือกในภาพ	L1 < L3 < L4 < L2
เฉลยแนะนำ	L1 < L3 < L4 < L2

อธิบาย: L1 เป็น regular, L3 เป็น context-free ที่ต้องจับคู่ a กับ c, L4 ต้องจับสามจำนวนเท่ากัน, และ L2 เพิ่มเงื่อนไขสี่กลุ่มจึงซับซ้อนกว่าในตัวเลือก

ข้อ 52

คำถาม: ในงาน Triage ที่มีความเสี่ยงสูง การตั้งค่า Error Budget ควรให้ความสำคัญกับเรื่องใดเป็นพิเศษ?

ตัวเลือก

- ลด False Positive ให้เหลือน้อยที่สุด
- ยกเลิก Error Budget เพื่อให้ AI ทำงานถูกต้อง 100%
- **ลด False Negative ให้เหลือน้อยที่สุด [เลือกในภาพ, เฉลย]**
- ทำให้ค่าเฉลี่ยความผิดพลาดทั้งสองแบบเท่ากันเป๊ะ
- เน้นความเร็วในการประมวลผลโดยไม่สนใจความแม่นยำ

เลือกในภาพ	ลด False Negative ให้เหลือน้อยที่สุด
เฉลยแนะนำ	ลด False Negative ให้เหลือน้อยที่สุด

อธิบาย: ใน triage ความเสี่ยงสูง การพลาดเคสสำคัญมักเสียหายมากกว่า false alarm จึงต้องคุม false negative ให้ต่ำ

ข้อ 53

คำถาม: เราใช้การกรองข้อมูลภาพ โดยใช้ค่ามัธยฐาน (Median filter) เพื่อทำอะไร

ตัวเลือก

- **ทำให้ภาพไม่มีสัญญาณรบกวน [เลือกในภาพ, เฉลย]**
- ทำให้ภาพมีคอนทราสต์ต่ำ
- ทำให้ภาพคมชัด
- ทำให้ภาพมีคอนทราสต์สูง

เลือกในภาพ	ทำให้ภาพไม่มีสัญญาณรบกวน
เฉลยแนะนำ	ทำให้ภาพไม่มีสัญญาณรบกวน

อธิบาย: Median filter เหมาะกับการลด impulse/salt-and-pepper noise
โดยแทนค่าพิกเซลด้วยค่ามัธยฐานของบริเวณรอบข้าง

ข้อ 54

คำถาม: เพราะเหตุใดภาษาไทยจึงจัดเป็นภาษาคำโดด (analytic language)

ตัวเลือก

- จะใช้คำขยายเพื่ออธิบายรายละเอียดทางไวยากรณ์ เช่น พจน์หรือกาล ไม่ใช้การผันท้ายคำ
- มีลำดับคำแน่นอนตายตัว เช่น ประธาน-กริยา-กรรม หากเปลี่ยนลำดับคำ ความหมายจะเปลี่ยนทันที
- หน่วยความหมายของคำเขียนขาดเรือ จึงต้องนำคำมาประกอบสร้างเพื่ออธิบายความหมายที่ซับซ้อน
- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: ภาษาไทยพึ่งคำช่วย/คำขยายและลำดับคำมากกว่าการผันรูปคำ จึงจัดเป็น analytic language

ข้อ 55

คำถาม: Programming Language ใดต่อไปนี้ที่ได้รับความนิยมมากที่สุดสำหรับการใช้งานใน Data Science ในปี 2024?

ตัวเลือก

- C++
- Java
- **Python [เลือกในภาพ, เฉลย]**
- R

เลือกในภาพ	Python
เฉลยแนะนำ	Python

อธิบาย: Python เป็นภาษาเด่นใน data science เพราะมี ecosystem เช่น NumPy, pandas, scikit-learn, PyTorch และ TensorFlow

ข้อ 56

คำถาม: เมื่อเลือกใช้ Embedding Model สิ่งสำคัญที่ต้องพิจารณาคืออะไร?

ตัวเลือก

- ความเร็วในการดาวน์โหลดโมเดล
- ชนิดของฐานข้อมูลที่ใช้เก็บข้อมูล
- **ภาษาที่โมเดลรองรับ ว่าตรงกับภาษาของเอกสารที่ใช้หรือไม่ [เลือกในภาพ, เฉลย]**
- ภาษาโปรแกรมที่ใช้ในการพัฒนา
- จำนวน GPU ที่ใช้ในโมเดล

เลือกในภาพ	ภาษาที่โมเดลรองรับ ว่าตรงกับภาษาของเอกสารที่ใช้หรือไม่
เฉลยแนะนำ	ภาษาที่โมเดลรองรับ ว่าตรงกับภาษาของเอกสารที่ใช้หรือไม่

อธิบาย: Embedding ที่ไม่รองรับภาษา/โดเมนของเอกสารจะทำให้ similarity และ retrieval แย่ลง

ข้อ 57

คำถาม: กรณีต้องการนับจำนวนของวัตถุในภาพ ต้องใช้โมเดลการเรียนรู้เชิงลึกแบบใด

ตัวเลือก

- a) Image Classification
- b) Image segmentation
- c) Object detection
- **ถูกทั้งข้อ b และ c [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ถูกทั้งข้อ b และ c
เฉลยแนะนำ	ถูกทั้งข้อ b และ c

อธิบาย: การนับวัตถุต้องระบุตำแหน่งหรือแยก instance; object detection ใช้ bounding boxes ส่วน segmentation ใช้ mask แล้วนับได้

ข้อ 58

คำถาม: ข้อใด เป็นคำอธิบายของ ระบบการเขียน (Writing System)?

ตัวเลือก

- use sets of symbols to represent the consonants, punctuation and numerals.
- use sets of symbols to represent the syllables, punctuation and numerals.
- **use sets of symbols to represent the sounds of speech, punctuation and numerals. [เลือกในภาพ, เฉลย]**
- use sets of symbols to represent the vowels, punctuation and numerals.

เลือกในภาพ	use sets of symbols to represent the sounds of speech, punctuation and numerals.
เฉลยแนะนำ	use sets of symbols to represent the sounds of speech, punctuation and numerals.

อธิบาย: Writing system คือระบบสัญลักษณ์แทนเสียง/หน่วยภาษา รวมถึงเครื่องหมายวรรคตอนและตัวเลข ไม่จำกัดเฉพาะพยัญชนะหรือสระ

ข้อ 59

คำถาม: จากคำสั่ง Python: img_shape=(1280,720,3); img=np.random.rand(*img_shape); plt.imshow(img) จงเลือกคำอธิบายที่ถูกต้องเกี่ยวกับขนาดภาพและประเภทข้อมูลของ img

ตัวเลือก

- สูง 720 กว้าง 1280 มี 3 ระนาบ เป็น float32
- สูง 1280 กว้าง 720 มี 3 ระนาบ เป็น float32
- **สูง 1280 กว้าง 720 มี 3 ระนาบ เป็น float64 [เลือกในภาพ, เฉลย]**
- สูง 720 กว้าง 1280 มี 3 ระนาบ เป็น float64

เลือกในภาพ	สูง 1280 กว้าง 720 มี 3 ระนาบ เป็น float64
เฉลยแนะนำ	สูง 1280 กว้าง 720 มี 3 ระนาบ เป็น float64

อธิบาย: NumPy image shape สำหรับ imshow คือ (height, width, channel) และ np.random.rand สร้าง float64 เป็นค่าเริ่มต้น

ข้อ 60

คำถาม: ข้อใดเป็นคุณสมบัติที่ถูกต้องของ Eigenvalues และ Eigenvectors?

ตัวเลือก

- **Eigenvalues ของเมทริกซ์เดิมจะไม่เปลี่ยนแปลงเมื่อทำการ Transpose เมทริกซ์นั้น [เลือกในภาพ, เฉลย]**
- Eigenvector ของเมทริกซ์หนึ่ง ๆ สามารถเป็นเวกเตอร์ศูนย์ได้
- Eigenvalues และ Eigenvectors ใช้ในการแปลงเมทริกซ์หนึ่ง ๆ ให้เป็นเมทริกซ์ศูนย์
- Eigenvector ของเมทริกซ์หนึ่ง ๆ จะเปลี่ยนทิศทางเมื่อถูกคูณด้วย Eigenvalue

เลือกในภาพ	Eigenvalues ของเมทริกซ์เดิมจะไม่เปลี่ยนแปลงเมื่อทำการ Transpose เมทริกซ์นั้น
เฉลยแนะนำ	Eigenvalues ของเมทริกซ์เดิมจะไม่เปลี่ยนแปลงเมื่อทำการ Transpose เมทริกซ์นั้น

อธิบาย: A และ A^T มี characteristic polynomial เดียวกันจึงมี eigenvalues ชุดเดียวกัน ส่วน eigenvector ต้องไม่ใช่เวกเตอร์ศูนย์

ข้อ 61

คำถาม: ข้อใดคือตัวอย่างของการกำหนดข้อมูลประเภท Dictionary ใน Python ที่ถูกต้อง?

ตัวเลือก

- **dict = {"name": "John", "age": 30} [เลือกในภาพ, เฉลย]**
- dict = ["name": "John", "age": 30]
- dict = < "name": "John", "age": 30 >
- dict = ("name": "John", "age": 30)

เลือกในภาพ	dict = {"name": "John", "age": 30}
เฉลยแนะนำ	dict = {"name": "John", "age": 30}

อธิบาย: Dictionary ใน Python ใช้ปีกกาและ key-value pairs คั่นด้วย colon

ข้อ 62

คำถาม: ในการสร้าง Sentiment lexicon สำหรับ Lexicon-based sentiment analysis หากไม่ต้องการให้คำคำหนึ่งมีค่า Polarity score สูงเกินควร สัดส่วนการปรากฏคำในข้อใดจำเป็นต้องนำมาใช้ในการคำนวณคะแนนความรู้สึกของคำนั้นด้วย?

ตัวเลือก

- ไม่มีข้อใดถูก
- สัดส่วนการปรากฏคำในเอกสารเชิงลบ (Negative)
- **สัดส่วนการปรากฏคำในเอกสารที่เป็นกลาง (Neutral) [เลือกในภาพ, เฉลย]**
- สัดส่วนการปรากฏคำในเอกสารเชิงบวก (Positive)

เลือกในภาพ	สัดส่วนการปรากฏคำในเอกสารที่เป็นกลาง (Neutral)
เฉลยแนะนำ	สัดส่วนการปรากฏคำในเอกสารที่เป็นกลาง (Neutral)

อธิบาย: ถ้าคำปรากฏบ่อยในเอกสาร neutral ควรลดน้ำหนัก polarity เพราะคำอาจเป็นคำทั่วไป ไม่ใช่สัญญาณอารมณ์แรง

ข้อ 63

คำถาม: ในระบบ RAG ขั้นตอนใดเกิดขึ้นก่อนการส่งข้อมูลให้ LLM สร้างคำตอบ?

ตัวเลือก

- แปลคำถามเป็นภาษาอังกฤษก่อนประมวลผล
- สร้าง Summary ของคำถามก่อนส่งให้ LLM
- Fine-tuning โมเดล LLM ด้วยข้อมูลใหม่
- **ค้นหาเอกสารที่เกี่ยวข้องจาก Knowledge Base แล้วส่งเป็น Context [เลือกในภาพ, เฉลย]**
- ตรวจสอบ Grammar ของคำถามผู้ใช้

เลือกในภาพ	ค้นหาเอกสารที่เกี่ยวข้องจาก Knowledge Base แล้วส่งเป็น Context
เฉลยแนะนำ	ค้นหาเอกสารที่เกี่ยวข้องจาก Knowledge Base แล้วส่งเป็น Context

อธิบาย: RAG เริ่มจากรับ query แล้ว retrieve เอกสาร/ชิ้นข้อมูลที่เกี่ยวข้องเพื่อแนบเป็น context ให้ LLM

ข้อ 64

คำถาม: Pattern แบบ 'AI suggests, human decides' จัดเป็นรูปแบบการทำงานประเภทใด?

ตัวเลือก

- Static Rule-based Engine
- **Human-in-the-loop [เลือกในภาพ, เฉลย]**
- Fully Autonomous System
- Manual Data Entry
- Database Management System

เลือกในภาพ	Human-in-the-loop
เฉลยแนะนำ	Human-in-the-loop

อธิบาย: Human-in-the-loop คือ AI ช่วยเสนอแนะหรือประมวลผล แต่คนยังตรวจหรืออนุมัติการตัดสินใจสำคัญ

ข้อ 65

คำถาม: ข้อใดไม่ใช่ลักษณะของ Volume ของ Big Data

ตัวเลือก

- ข้อมูลแบบมีโครงสร้างและไม่มีโครงสร้าง
- จำนวนของข้อมูลที่ถูกสร้างขึ้น
- **การจัดเก็บข้อมูลในตาราง หรือไฟล์ [เลือกในภาพ, เฉลย]**
- ธุรกรรมออนไลน์และออฟไลน์

เลือกในภาพ	การจัดเก็บข้อมูลในตาราง หรือไฟล์
เฉลยแนะนำ	การจัดเก็บข้อมูลในตาราง หรือไฟล์

อธิบาย: Volume เน้นปริมาณ/ขนาดของข้อมูล ไม่ใช่วิธีจัดเก็บเป็นตารางหรือไฟล์

ข้อ 66

คำถาม: ฟังก์ชันหลัก 4 ประการของ NIST AI RMF (Risk Management Framework) ประกอบด้วยอะไรบ้าง?

ตัวเลือก

- Reach, Impact, Confidence, Effort
- Input, Processing, Output, Outcome
- Identify, Protect, Detect, Recover
- Plan, Do, Check, Act
- **Govern, Map, Measure, Manage [เลือกในภาพ, เฉลย]**

เลือกในภาพ	Govern, Map, Measure, Manage
เฉลยแนะนำ	Govern, Map, Measure, Manage

อธิบาย: AI RMF Core ของ NIST แบ่งเป็น 4 functions คือ Govern, Map, Measure และ Manage

ข้อ 67

คำถาม: หากต้องการทำ Clustering ด้วยวิธี K-Means อย่างมีประสิทธิภาพ ควรคำนึงถึงเรื่องใด?

ตัวเลือก

- การกำหนดค่า K ที่เหมาะสม
- ต้องมีการทำ Scaler และกำจัด Outliers ในชุดข้อมูล
- การกำหนดจุดเริ่มต้นของตัวแทนแต่ละ Cluster ที่ดี
- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: K-means ไวต่อจำนวน cluster, scale ของ feature, outlier และ initial centroid

ข้อ 68

คำถาม: ข้อใด คือ Hypothesis ของ Polynomial Regression Degree = 3

ตัวเลือก

- $h_\theta(x)=\theta_0+\theta_1x_1+\theta_2x_1^2$
- $h_\theta(x)=\theta_0+\theta_1x_1+\theta_2x_1^2+\ldots+\theta_nx_1^n$
- **$h_\theta(x)=\theta_0+\theta_1x_1+\theta_2x_1^2+\theta_3x_1^3$ [เลือกในภาพ, เฉลย]**
- $h_\theta(x)=\theta_0+\theta_1x_1$

เลือกในภาพ	$h_\theta(x)=\theta_0+\theta_1x_1+\theta_2x_1^2+\theta_3x_1^3$
เฉลยแนะนำ	$h_\theta(x)=\theta_0+\theta_1x_1+\theta_2x_1^2+\theta_3x_1^3$

อธิบาย: Polynomial degree 3 ต้องมีพจน์ถึง x^3 พร้อมสัมประสิทธิ์ θ_3

ข้อ 69

คำถาม: System Summary: the cat was found under the bed; Reference Summary: the cat was under the bed จงหาค่า ROUGE-N (recall) เมื่อ N=2

ตัวเลือก

- 0.86
- 0.73
- 0.67
- **0.80 [เลือกในภาพ, เฉลย]**

เลือกในภาพ	0.80
เฉลยแนะนำ	0.80

อธิบาย: Reference bigrams มี 5 ตัว: the cat, cat was, was under, under the, the bed; System ตรงกัน 4 ตัว จึง recall = 4/5 = 0.80

ข้อ 70

คำถาม: ขั้นตอนที่สำคัญของการทำ Extraction-based Approach?

ตัวเลือก

- **(1) ค้นหาคำ วลี ข้อความ ที่สำคัญและคำนวณ (2) เลือก คำ วลี ข้อความ ที่สำคัญ (3) สร้างข้อความ [เลือกในภาพ, เฉลย]**
- (1) ค้นหาคำ วลี ข้อความ ที่สำคัญ (2) เลือก คำ วลี ข้อความ ที่สำคัญ (3) สร้างข้อความ
- (1) ตัดคำ (2) แบ่งข้อความ (3) สร้างข้อความ
- (1) ค้นหาคำ วลี ข้อความ ที่สำคัญ (2) เลือกคำ วลี ข้อความ มาเรียบเรียงใหม่

เลือกในภาพ	(1) ค้นหาคำ วลี ข้อความ ที่สำคัญและคำนวณ (2) เลือก คำ วลี ข้อความ ที่สำคัญ (3) สร้างข้อความ
เฉลยแนะนำ	(1) ค้นหาคำ วลี ข้อความ ที่สำคัญและคำนวณ (2) เลือก คำ วลี ข้อความ ที่สำคัญ (3) สร้างข้อความ

อธิบาย: Extractive summarization มักให้คะแนน/คำนวณความสำคัญของหน่วยข้อความ เลือกหน่วยที่สำคัญ แล้วรวมเป็นสรุปโดยไม่สร้างข้อความใหม่มากนัก

ข้อ 71

คำถาม: ข้อใดไม่ถูกต้องเกี่ยวกับ Open Data

ตัวเลือก

- ไม่มีข้อใดถูก
- Open data คือ ข้อมูลเปิดที่ทุกคนสามารถนำไปใช้ได้โดยอิสระ เช่น การนำไปใช้ การนำกลับมาใช้ใหม่ หรือนำไปเผยแพร่ได้โดยใครก็ตาม
- Open Data จัดเป็น Publicly available data ทั้งหมด แต่ Publicly available data ทั้งหมดไม่จัดเป็น Open Data
- **Government data คือข้อมูลจากรัฐบาลและเป็น Open data ทั้งหมด [เลือกในภาพ, เฉลย]**

เลือกในภาพ	Government data คือข้อมูลจากรัฐบาลและเป็น Open data ทั้งหมด
เฉลยแนะนำ	Government data คือข้อมูลจากรัฐบาลและเป็น Open data ทั้งหมด

อธิบาย: ข้อมูลภาครัฐบางส่วนอาจเป็นความลับ มีข้อมูลส่วนบุคคล หรือมีข้อจำกัดทางกฎหมาย จึงไม่ใช่ open data ทั้งหมด

ข้อ 72

คำถาม: ข้อใดต่อไปนี้แสดงให้เห็นถึงข้อดีของการใช้ Version control ของ GitHub ใน Google Colab ได้ดีที่สุด?

ตัวเลือก

- สมาชิกในทีมหลายคนสามารถแก้ไขส่วนต่าง ๆ ของ Notebook เดียวกันพร้อม ๆ กันได้
- Notebook จะ Save โดยอัตโนมัติทุก ๆ 5 นาที
- **ทีมงานสามารถ Maintain Notebook เวอร์ชันต่าง ๆ สำหรับการทดลองต่าง ๆ ได้ [เลือกในภาพ, เฉลย]**
- โปรแกรมเมอร์สามารถ Undo การเปลี่ยนแปลงที่ทำกับ Notebook เมื่อสัปดาห์ที่แล้วได้

เลือกในภาพ	ทีมงานสามารถ Maintain Notebook เวอร์ชันต่าง ๆ สำหรับการทดลองต่าง ๆ ได้
เฉลยแนะนำ	ทีมงานสามารถ Maintain Notebook เวอร์ชันต่าง ๆ สำหรับการทดลองต่าง ๆ ได้

อธิบาย: Version control ช่วยติดตามประวัติ แยก branch/เวอร์ชัน และย้อนกลับ experiment ต่าง ๆ ได้เป็นระบบ

ข้อ 73

คำถาม: ในมุมมองของ Responsible AI และ Google PAIR การระบุในหน้าจอว่า 'Draft generated from FAQ v3' จัดเป็นการสร้างคุณสมบัติใดให้แก่ระบบ?

ตัวเลือก

- Fairness
- Accountability
- Inclusiveness
- Reliability
- **Transparency & Explanation [เลือกในภาพ, เฉลย]**

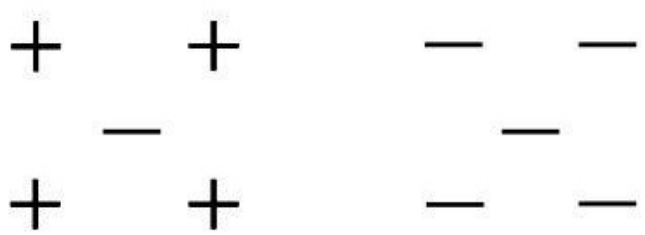
เลือกในภาพ	Transparency & Explanation
เฉลยแนะนำ	Transparency & Explanation

อธิบาย: การบอกแหล่งที่มาหรือสถานะของ output ทำให้ผู้ใช้เข้าใจว่าข้อความมาจากไหนและควรเชื่อมากแค่ไหน

ข้อ 74

คำถาม: จากภาพ ค่า k ใดต่อไปนี้ที่จะให้ความแม่นยำในการตรวจสอบแบบ leave-one-out cross validation ต่ำที่สุด

ภาพประกอบข้อ 74



ตัวเลือก

- 1
- 3
- 2
- **5 [เลือกในภาพ, เฉลย]**

เลือกในภาพ	5
เฉลยแนะนำ	5

อธิบาย: เมื่อ k ใหญ่เกินไปในข้อมูลที่คลาสแยกเป็นกลุ่มเล็ก ๆ เพื่อนบ้านส่วนใหญ่จะครอบคลุมข้ามกลุ่ม ทำให้ทำนายผิดมากขึ้นเมื่อทำ leave-one-out

ข้อ 75

คำถาม: เหตุใดจึงต้องใช้เทคนิค RAG แทนที่จะให้ LLM ตอบคำถามโดยตรง?

ตัวเลือก

- LLM ไม่สามารถสร้างข้อความยาวได้
- LLM ไม่สามารถเข้าใจภาษาธรรมชาติได้ดีพอ
- LLM ใช้พลังงานในการประมวลผลมากเกินไป
- **LLM อาจสร้างข้อมูลที่ไม่ถูกต้องเมื่อไม่มีแหล่งข้อมูลอ้างอิง [เลือกในภาพ, เฉลย]**
- LLM ต้องการ Fine-tuning ทุกครั้งก่อนใช้งาน

เลือกในภาพ	LLM อาจสร้างข้อมูลที่ไม่ถูกต้องเมื่อไม่มีแหล่งข้อมูลอ้างอิง
เฉลยแนะนำ	LLM อาจสร้างข้อมูลที่ไม่ถูกต้องเมื่อไม่มีแหล่งข้อมูลอ้างอิง

อธิบาย: RAG ลด hallucination และช่วยให้คำตอบอิงแหล่งข้อมูลปัจจุบัน/เฉพาะองค์กรได้มากขึ้น

ข้อ 76

คำถาม: ทำไมการนิยาม Input ให้ชัดเจนจึงสำคัญมากในการ Define โจทย์ AI?

ตัวเลือก

- เพื่อป้องกันไม่ให้ผู้ใช้ป้อนข้อมูลผิดพลาด
- **เพื่อประเมินว่าระบบทำงานบนพื้นฐานของความจริงหรือความฝัน [เฉลย]**
- เพื่อให้ทีมกราฟิกออกแบบไอคอนได้ถูกต้อง
- เพื่อลดค่าใช้จ่ายในการซื้อ Cloud Computing [เลือกในภาพ]

เลือกในภาพ	เพื่อลดค่าใช้จ่ายในการซื้อ Cloud Computing
เฉลยแนะนำ	เพื่อประเมินว่าระบบทำงานบนพื้นฐานของความจริงหรือความฝัน

อธิบาย: การนิยาม input ทำให้รู้ว่ามีข้อมูลจริงพอหรือไม่ คุณภาพเป็นอย่างไร และโจทย์ AI ทำได้จริงหรือเป็นเพียงความคาดหวัง

หมายเหตุ: คำตอบที่เลือกในภาพไม่ตรงกับหลักการ define AI problem

ข้อ 77

คำถาม: ข้อใดถูกต้องเกี่ยวกับปัญหา Class Imbalance ใน Machine Learning?

ตัวเลือก

- Class Imbalance สามารถแก้ไขได้โดยการลบตัวอย่างจากคลาสที่มีจำนวนตัวอย่างน้อยกว่า
- Class Imbalance เกิดขึ้นเมื่อทุกคลาสมีจำนวนตัวอย่างเท่ากัน
- **การใช้เทคนิคการ Oversampling หรือ Undersampling เป็นวิธีที่ใช้แก้ไขปัญหา Class Imbalance [เลือกในภาพ, เฉลย]**
- Class Imbalance สามารถแก้ไขได้โดยการเพิ่มตัวอย่างจากคลาสที่มีจำนวนตัวอย่างมากกว่า

เลือกในภาพ	การใช้เทคนิคการ Oversampling หรือ Undersampling เป็นวิธีที่ใช้แก้ไขปัญหา Class Imbalance
เฉลยแนะนำ	การใช้เทคนิคการ Oversampling หรือ Undersampling เป็นวิธีที่ใช้แก้ไขปัญหา Class Imbalance

อธิบาย: Oversampling เพิ่มตัวอย่างคลาสน้อย และ undersampling ลดคลาสมาก เพื่อให้โมเดลไม่เอนเอียงตามคลาสใหญ่

ข้อ 78

คำถาม: ข้อใดต่อไปนี้เป็นตัวอย่างข้อความที่เกี่ยวข้องกับการสรุปความหรือย่อความ: 1) Digests (TV Guide) 2) Biography (Resume) 3) Minutes (of a meeting) 4) Previews (of movies)

ตัวเลือก

- 2), 4)
- 2), 3), 4)
- **1), 2), 3), 4) [เฉลย]**
- 1), 2), 3) [เลือกในภาพ]

เลือกในภาพ	1), 2), 3)
เฉลยแนะนำ	1), 2), 3), 4)

อธิบาย: ตามหลักทั่วไป ทั้ง digest, resume/biography, meeting minutes และ movie preview ล้วนเป็นรูปแบบของข้อความสรุปหรือย่อความในระดับต่าง ๆ

หมายเหตุ: ภาพเลือก 1),2),3) แต่ previews of movies ก็มักถือเป็นข้อความสรุปได้ด้วย จึงควรจำแนวคิดมากกว่าท่องตัวเลือก

ข้อ 79

คำถาม: อะไรคือนิยามการ learning ด้วย data สำหรับ machine ที่ไม่เจาะจงกับ task

ตัวเลือก

- machine ถูกสอนหนังสือ
- Machine จดจำภาพได้
- **machine ทำงานใน task ที่กำหนดเก่งขึ้น เมื่อเราให้ข้อมูลแก่ machine เทียบกับตอนไม่มีข้อมูล [เลือกในภาพ, เฉลย]**
- machine ช่วยสอนหนังสือ

เลือกในภาพ	machine ทำงานใน task ที่กำหนดเก่งขึ้น เมื่อเราให้ข้อมูลแก่ machine เทียบกับตอนไม่มีข้อมูล
เฉลยแนะนำ	machine ทำงานใน task ที่กำหนดเก่งขึ้น เมื่อเราให้ข้อมูลแก่ machine เทียบกับตอนไม่มีข้อมูล

อธิบาย: Machine learning คือการปรับปรุง performance จากข้อมูล/ประสบการณ์ในงานที่กำหนด

ข้อ 80

คำถาม: ข้อใดคือผลลัพธ์จากการใช้ regular expression ของข้อความต่อไปนี้ 'Th[AI]land'

ตัวเลือก

- **ThIland [เลือกในภาพ, เฉลย]**
- Thland
- Thalland
- ThAiland

เลือกในภาพ	ThIland
เฉลยแนะนำ	ThIland

อธิบาย: [AI] คือ character class ที่เลือกตัวอักษร A หรือ I เพียงหนึ่งตัว จึง match ThAland หรือ ThIland; จากตัวเลือกมี ThIland

ข้อ 81

คำถาม: การใช้ Deepfake ในการสร้างข่าวปลอมอาจส่งผลกระทบต่อสังคมอย่างไร

ตัวเลือก

- **ลดความน่าเชื่อถือของข้อมูลในสังคม [เลือกในภาพ, เฉลย]**
- เพิ่มความไว้วางใจในสื่อมวลชน
- ทำให้ผู้คนมีความสุขมากขึ้น
- สร้างโอกาสในการพัฒนาทางเทคโนโลยี

เลือกในภาพ	ลดความน่าเชื่อถือของข้อมูลในสังคม
เฉลยแนะนำ	ลดความน่าเชื่อถือของข้อมูลในสังคม

อธิบาย: Deepfake ทำให้ผู้คนแยกจริง/ปลอมยากขึ้น จึงลดความน่าเชื่อถือของข่าวและหลักฐานดิจิทัล

ข้อ 82

คำถาม: ในระบบ AI Triage รูปแบบ Human-in-the-Loop (HITL) แบบใดที่เหมาะสมที่สุดตามเนื้อหา?

ตัวเลือก

- **AI ให้คำแนะนำเบื้องต้น และให้มนุษย์เป็นผู้ยืนยันหรือตัดสินใจในขั้นตอนสุดท้าย [เลือกในภาพ, เฉลย]**
- มนุษย์ทำงานทั้งหมดและใช้ AI สำหรับเก็บประวัติการทำงานเท่านั้น
- AI จะทำงานเฉพาะเมื่อมนุษย์ไม่ว่างเท่านั้น
- ให้ AI ตัดสินใจในเคสที่ยาก และมนุษย์ทำเคสที่ง่าย
- AI ตัดสินใจและดำเนินการโดยไม่ต้องแจ้งให้มนุษย์ทราบ

เลือกในภาพ	AI ให้คำแนะนำเบื้องต้น และให้มนุษย์เป็นผู้ยืนยันหรือตัดสินใจในขั้นตอนสุดท้าย
เฉลยแนะนำ	AI ให้คำแนะนำเบื้องต้น และให้มนุษย์เป็นผู้ยืนยันหรือตัดสินใจในขั้นตอนสุดท้าย

อธิบาย: ในงานเสี่ยงสูง ควรให้ AI ช่วยคัดกรอง/แนะนำ แต่คนเป็นผู้รับผิดชอบการตัดสินใจขั้นสุดท้าย

ข้อ 83

คำถาม: ข้อใดต่อไปนี้ไม่เป็นความจริง

ตัวเลือก

- ภาพดิจิตอลสามารถมีได้มากกว่า 3 ช่องสี
- ภาพถ่ายดาวเทียมสามารถเก็บได้มากกว่า 3 ช่องสีและมากกว่า 8 bit ต่อช่องสี
- **ภาพ RGB สามารถเก็บได้เพียง 24 bit ต่อจุดภาพเท่านั้น [เลือกในภาพ, เฉลย]**
- ภาพดิจิตอลสามารถเก็บข้อมูลได้มากกว่า 8 bit ต่อช่องสี

เลือกในภาพ	ภาพ RGB สามารถเก็บได้เพียง 24 bit ต่อจุดภาพเท่านั้น
เฉลยแนะนำ	ภาพ RGB สามารถเก็บได้เพียง 24 bit ต่อจุดภาพเท่านั้น

อธิบาย: RGB มักพบแบบ 8 bit x 3 = 24 bit แต่ไม่จำเป็นต้องเป็นได้เพียงเท่านั้น อาจใช้ 16 bit/float ต่อ channel ได้

ข้อ 84

คำถาม: ข้อใดต่อไปนี้ แสดงจุดประสงค์การทำ Data Discretization ได้ถูกต้องที่สุด

ตัวเลือก

- เพื่อขยายขนาดข้อมูล
- เพื่อการนำข้อมูลจากหลายแหล่งเข้ามาจัดรูปแบบให้คล้ายกัน
- เพื่อแปลงแอตทริบิวต์ที่ไม่ต่อเนื่องเป็นแอตทริบิวต์แบบต่อเนื่อง
- **เพื่อแปลงแอตทริบิวต์ต่อเนื่องเป็นแอตทริบิวต์ที่ไม่ต่อเนื่อง [เลือกในภาพ, เฉลย]**

เลือกในภาพ	เพื่อแปลงแอตทริบิวต์ต่อเนื่องเป็นแอตทริบิวต์ที่ไม่ต่อเนื่อง
เฉลยแนะนำ	เพื่อแปลงแอตทริบิวต์ต่อเนื่องเป็นแอตทริบิวต์ที่ไม่ต่อเนื่อง

อธิบาย: Discretization แบ่งค่าต่อเนื่องเป็นช่วง/กลุ่ม เช่น อายุเป็นเด็ก/วัยทำงาน/สูงอายุ

ข้อ 85

คำถาม: ข้อใด 'ไม่ใช่' ข้อดีของ Estimation statistics ที่เหนือกว่า Null hypothesis test

ตัวเลือก

- **Estimation statistics ทำงานเร็วกว่า hypothesis test [เลือกในภาพ, เฉลย]**
- เรารู้ magnitude ความแตกต่างของประชากร ไม่ใช่แค่ yes/no answer แบบ hypothesis test
- Estimation statistics เปิดโอกาสให้ผู้วิเคราะห์เห็น insight ของข้อมูลมากกว่า hypothesis test
- Estimation statistics ให้ข้อมูลละเอียดกว่า null hypothesis test

เลือกในภาพ	Estimation statistics ทำงานเร็วกว่า hypothesis test
เฉลยแนะนำ	Estimation statistics ทำงานเร็วกว่า hypothesis test

อธิบาย: ข้อดีของ estimation statistics คือการรายงาน effect size/uncertainty ไม่ใช่ความเร็วในการคำนวณ

ข้อ 86

คำถาม: หากผลการทดสอบ PoC ของระบบ Auto-triage พบว่าความแม่นยำอยู่ที่ 70-84% และต้นทุนดีแต่คุณภาพยังไม่นิ่ง ตามเกณฑ์ Success Criteria ควรตัดสินใจอย่างไร?

ตัวเลือก

- Scale up
- Stop
- Anonymize
- **Pivot [เลือกในภาพ, เฉลย]**
- Go

เลือกในภาพ	Pivot
เฉลยแนะนำ	Pivot

อธิบาย: ผลยังไม่ถึงระดับพร้อม production แต่มีสัญญาณว่าคุ้ม จึงควร pivot/refine แนวทางก่อน ไม่ใช่ scale หรือ go ทันที

ข้อ 87

คำถาม: ข้อใดคือตัวอย่างของการกำหนดข้อมูลประเภท Tuple ใน Python ที่ถูกต้อง?

ตัวเลือก

- tuple = <1, 2, 3, 4>
- tuple = {1, 2, 3, 4}
- **tuple = (1, 2, 3, 4) [เลือกในภาพ, เฉลย]**
- tuple = [1, 2, 3, 4]

เลือกในภาพ	tuple = (1, 2, 3, 4)
เฉลยแนะนำ	tuple = (1, 2, 3, 4)

อธิบาย: Tuple literal ใช้วงเล็บและคั่นสมาชิกด้วย comma; ปีกกาเป็น set/dict และวงเล็บเหลี่ยมเป็น list

ข้อ 88

คำถาม: ข้อความใดต่อไปนี้ พูดถึง Google Colab ได้ถูกต้องที่สุด?

ตัวเลือก

- Google Colab รองรับภาษาโปรแกรมได้ทุกภาษา
- **Google Colab ให้สิทธิ์ในการเข้าถึงทรัพยากรการประมวลผลประสิทธิภาพสูงฟรี [เลือกในภาพ, เฉลย]**
- Google Colab จะมีการ Optimize code ให้อัตโนมัติ
- Google Colab มีเครื่องมือสำหรับการวิเคราะห์ข้อมูลขั้นสูงในตัว

เลือกในภาพ	Google Colab ให้สิทธิ์ในการเข้าถึงทรัพยากรการประมวลผลประสิทธิภาพสูงฟรี
เฉลยแนะนำ	Google Colab ให้สิทธิ์ในการเข้าถึงทรัพยากรการประมวลผลประสิทธิภาพสูงฟรี

อธิบาย: Colab ให้ notebook บน cloud พร้อมทรัพยากรเช่น GPU/TPU แบบมีข้อจำกัด จึงเหมาะกับการทดลอง ML

ข้อ 89

คำถาม: ข้อใดต่อไปนี้ถูกต้องเกี่ยวกับ bias-variance decomposition ระหว่าง ridge regression และ ordinary least squares regression

ตัวเลือก

- ridge regression มี Bias และ variance สูงกว่า
- ridge regression มี Bias และ variance ต่ำกว่า
- ridge regression มี Bias ต่ำกว่า แต่มี variance สูงกว่า
- **ridge regression มี Bias สูงกว่า และมี variance ต่ำกว่า [เลือกในภาพ, เฉลย]**

เลือกในภาพ	ridge regression มี Bias สูงกว่า และมี variance ต่ำกว่า
เฉลยแนะนำ	ridge regression มี Bias สูงกว่า และมี variance ต่ำกว่า

อธิบาย: Ridge เพิ่ม regularization ทำให้ค่าสัมประสิทธิ์หดลง โมเดล bias เพิ่มเล็กน้อยแต่ variance ลดลงเมื่อเทียบกับ OLS

ข้อ 90

คำถาม:
เทคนิคใดต่อไปนี้ใช้สำหรับวิเคราะห์ข้อมูลที่อาจถูกกำหนดให้เป็นข้อมูลที่ไม่สอดคล้องกับพฤติกรรมทั่วไปหรือรูปแบบของข้อมูลที่มีอยู่

ตัวเลือก

- Prediction
- Classification
- Evolution Analysis
- **Outlier Analysis [เลือกในภาพ, เฉลย]**

เลือกในภาพ	Outlier Analysis
เฉลยแนะนำ	Outlier Analysis

อธิบาย: Outlier analysis ใช้ตรวจหาจุดข้อมูลที่เบี่ยงเบนจาก pattern ปกติ

ข้อ 91

คำถาม: โมเดลในชุดใดต่อไปนี้ถูกสร้างมาเพื่องาน Generative (ให้ AI มาเขียนต่อจากประโยคเริ่มต้นได้) ทั้งหมด

ตัวเลือก

- DPR, RAG, Longformer
- BERT, GPT, Google T5
- **GPT, CTRL, Transformer-XL [เลือกในภาพ, เฉลย]**
- BERT, Roberta, XLM

เลือกในภาพ	GPT, CTRL, Transformer-XL
เฉลยแนะนำ	GPT, CTRL, Transformer-XL

อธิบาย: GPT, CTRL และ Transformer-XL เป็น autoregressive/generative language models ส่วน BERT/RoBERTa เป็น masked language models

ข้อ 92

คำถาม: องค์ประกอบของโมเดลธุรกิจข้อมูลของ Amazon ประกอบไปด้วยอะไรบ้าง

ตัวเลือก

- **Information-based differentiation, Information-based delivery network, Information-based brokering [เลือกในภาพ, เฉลย]**
- ไม่มีข้อใดถูก
- Information-based differentiation
- Information-based differentiation, Information-based delivery network

เลือกในภาพ	Information-based differentiation, Information-based delivery network, Information-based brokering
เฉลยแนะนำ	Information-based differentiation, Information-based delivery network, Information-based brokering

อธิบาย: โมเดลข้อมูลของ Amazon ใช้ข้อมูลเพื่อสร้างความแตกต่าง จัดการเครือข่ายส่งมอบ และเป็นตัวกลาง/marketplace ผ่านข้อมูล

ข้อ 93

คำถาม: ภาษาอังกฤษ ภาษาเกรก และภาษาเกาหลี (Korean - Hangeul) มีระบบการเขียนแบบใด?

ตัวเลือก

- **Alphabets [เลือกในภาพ, เฉลย]**
- Abjads
- Abugidas
- Syllabaries

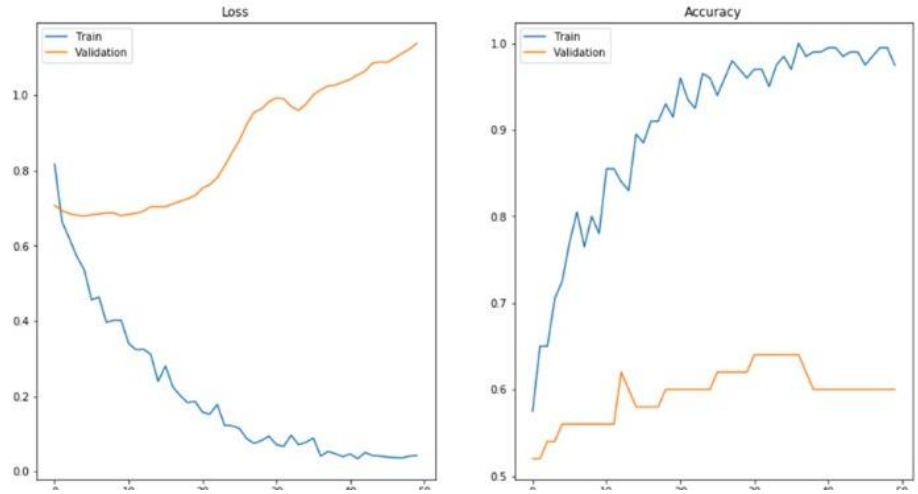
เลือกในภาพ	Alphabets
เฉลยแนะนำ	Alphabets

อธิบาย: ทั้งสามระบบมีตัวอักษรแทนหน่วยเสียง consonant/vowel ในระดับ alphabet แม้ Hangul จะจัดเป็นบล็อกพยางค์ในเชิงรูป

ข้อ 94

คำถาม: กราฟแสดงผล Loss และ Accuracy ในการ Train Model Transformer หากไม่ต้อง Train ใหม่ ควรใช้ Model ที่ได้รับการ Train ณ Epoch ใดเพื่อผลดีที่สุดเหมาะสมกับข้อมูลนำเข้า

ภาพประกอบข้อ 94



ตัวเลือก

- 10
- 0

- **13 [เลือกในภาพ, เฉลย]**
- 35

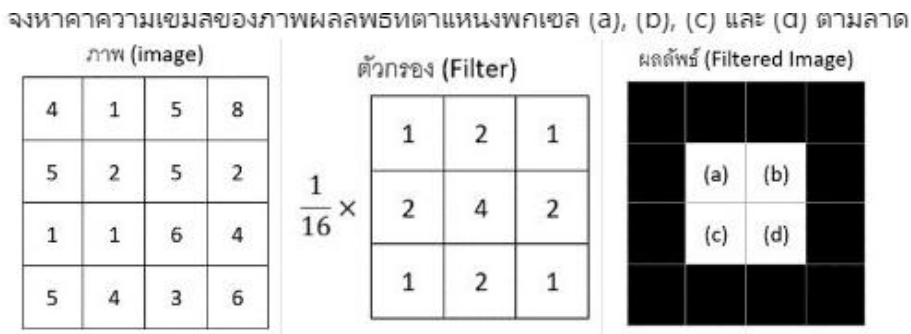
เลือกในภาพ	13
เฉลยแนะนำ	13

อธิบาย: Validation accuracy สูงสุดช่วงประมาณ epoch 13 ขณะที่หลังจากนั้น validation loss เริ่มสูงและเกิด overfitting มากขึ้น จึงควรเลือก checkpoint แถว epoch นี้

ข้อ 95

คำถาม: ภาพ 4x4 ถูกผ่านตัวกรอง 3x3 คูณ 1/16 จงหาค่าผลลัพธ์ตำแหน่ง (a), (b), (c), (d)

ภาพประกอบข้อ 95



ตัวเลือก

- **3,4,3,4 [เลือกในภาพ, เฉลย]**
- 3,3,4,4
- 4,4,3,3
- 4,3,4,3

เลือกในภาพ	3,4,3,4
เฉลยแนะนำ	3,4,3,4

อธิบาย: ทำ convolution แบบ valid ด้วย kernel [[1,2,1],[2,4,2],[1,2,1]]/16 จะได้ a=48/16=3, b=64/16=4, c=48/16=3, d=64/16=4

ข้อ 96

คำถาม: ข้อใดไม่ใช่ส่วนประกอบของ Information Flow ในการประมวลผลฐานข้อมูล

ตัวเลือก

- Interactive
- **Clustering [เลือกในภาพ, เฉลย]**
- Straming
- Batch

เลือกในภาพ	Clustering
เฉลยแนะนำ	Clustering

อธิบาย: Interactive, streaming และ batch เป็นรูปแบบการไหล/ประมวลผลข้อมูล ส่วน clustering เป็นเทคนิควิเคราะห์/ML

ข้อ 97

คำถาม: Application ใดเหมาะกับการนำ model ประเภท regression ไปประยุกต์ใช้งานที่สุด

ตัวเลือก

- **ทำนายปริมาณรังสีที่เพิ่มขึ้นในวันรุ่งขึ้น [เลือกในภาพ, เฉลย]**
- แยกภาพรถเป็นยี่ห้อต่าง ๆ
- หาหัวข้อที่กำลังเป็นที่พูดถึงจากข้อมูล twitter ในวันที่ผ่านมา
- เปลี่ยนเสียงพูดเป็นตัวอักษร
- สรุปข้อมูลลูกค้าบริษัทให้อยู่ในรูปแบบของกราฟ
- แยกประเภทของจดหมายเพื่อส่งไปยังแผนกที่เกี่ยวข้อง

เลือกในภาพ	ทำนายปริมาณรังสีที่เพิ่มขึ้นในวันรุ่งขึ้น
เฉลยแนะนำ	ทำนายปริมาณรังสีที่เพิ่มขึ้นในวันรุ่งขึ้น

อธิบาย: Regression เหมาะกับการทำนายค่าต่อเนื่อง เช่น ปริมาณรังสี ส่วนตัวเลือกอื่นเป็น classification, topic modeling, speech recognition หรือ reporting

ข้อ 98

คำถาม: ปรากฏการณ์ Hallucination ใน Generative AI หมายถึงอะไร?

ตัวเลือก

- การที่ AI ไม่สามารถตอบคำถามที่อยู่นอกเหนือจากคู่มือการใช้งานได้
- การที่ AI แสดงอคติที่มาจากชุดข้อมูลที่ใช้ในการฝึกฝน
- การที่ AI ทำงานช้าลงเมื่อได้รับคำสั่งที่ซับซ้อนเกินไป
- การที่ AI ถูกแฮ็กเพื่อนำข้อมูลส่วนบุคคลออกไปภายนอก
- **การที่ AI สร้างข้อมูลที่ดูน่าเชื่อถือแต่ไม่เป็นความจริง [เลือกในภาพ, เฉลย]**

เลือกในภาพ	การที่ AI สร้างข้อมูลที่ดูน่าเชื่อถือแต่ไม่เป็นความจริง
เฉลยแนะนำ	การที่ AI สร้างข้อมูลที่ดูน่าเชื่อถือแต่ไม่เป็นความจริง

อธิบาย: Hallucination คือคำตอบที่ฟังดูมั่นใจและเป็นไปได้ แต่ไม่มีหลักฐานหรือขัดกับข้อเท็จจริง

ข้อ 99

คำถาม: เพราะเหตุใดการทำ Empathize สำหรับงาน AI จึงมีความซับซ้อนกว่าการออกแบบแอปพลิเคชันทั่วไป?

ตัวเลือก

- เพราะการทำ Empathize เน้นเพียงแค่การเลือกโมเดลที่ทันสมัยที่สุด
- เพราะ AI ไม่จำเป็นต้องมีส่วนประสานงานผู้ใช้ (UI)
- เพราะระบบ AI ไม่ส่งผลกระทบต่อบุคคลอื่นนอกเหนือจากผู้ใช้โดยตรง
- **เพราะผู้ใช้ต้องมีความเชื่อใจในระบบที่เหมาะสม (Appropriate Trust) และต้องจัดการกับความไม่แน่นอน [เลือกในภาพ, เฉลย]**
- เพราะ AI ทำงานได้แม่นยำ 100% เสมอจึงไม่ต้องทดสอบกับคน

เลือกในภาพ	เพราะผู้ใช้ต้องมีความเชื่อใจในระบบที่เหมาะสม (Appropriate Trust) และต้องจัดการกับความไม่แน่นอน
เฉลยแนะนำ	เพราะผู้ใช้ต้องมีความเชื่อใจในระบบที่เหมาะสม (Appropriate Trust) และต้องจัดการกับความไม่แน่นอน

อธิบาย: AI มี uncertainty/error และผลกระทบต่อการตัดสินใจ ผู้ใช้จึงต้องเข้าใจขอบเขตความน่าเชื่อถือของระบบ ไม่ใช่เชื่อหรือไม่เชื่อสุดโต่ง

ข้อ 100

คำถาม: Data science ใช้ศาสตร์ไหนเป็นตัวช่วยบ้าง

ตัวเลือก

- Computer science
- Statistics
- **ถูกทุกข้อ [เลือกในภาพ, เฉลย]**
- Math

เลือกในภาพ	ถูกทุกข้อ
เฉลยแนะนำ	ถูกทุกข้อ

อธิบาย: Data science ผสม computer science, statistics, mathematics และ domain knowledge เพื่อสร้าง insight จากข้อมูล