น้องเอ็ม: ระบบคัดกรองสุขภาพจิตเบื้องต้นด้วยปัญญาประดิษฐ์และฐานข้อมูลสำหรับคนไทย

NongM: An Artificial Intelligence Based Preliminary Mental Health Screening System with a Mental Health Dataset for Thais

[1]วรวรรธน์ ศิริสวิสดิ์ [1]พิชชยุทธ รักษาโคตร
ที่ปรึกษา: [1]เอกชัย วัฒนไชย
[1]โรงเรียนวิทยาศาสตร์จุฬาภรณราชวิทยาลัย บุรีรัมย์ ตำบลสตึก อำเภอสตึก จังหวัดบุรีรัมย์ 31150
*อีเมลล์: worawat.sirisawat0@gmail.com

บทคัดย่อ

ปัจจุบันประเทศไทยเผชิญกับวิกฤตสุขภาพจิตอย่างต่อเนื่อง โดยมีผู้ป่วยโรคซึมเศร้าประมาณ 2.9 ล้านคน ขณะที่อัตราส่วนจิตแพทย์ต่อประชากรอยู่ที่เพียง 1 ต่อ 50,000 คน ส่งผลให้ประชาชนส่วนใหญ่ไม่สามารถเข้าถึง บริการสุขภาพจิตได้อย่างทันท่วงที โครงงานนี้นำเสนอ "น้องเอ็ม" ระบบปัญญาประดิษฐ์สำหรับคัดกรองสุขภาพจิต เบื้องต้นที่ออกแบบเฉพาะสำหรับภาษาไทย โดยผสานเทคโนโลยี Retrieval-Augmented Generation เข้ากับ Large Language Model เชี่ยวชาญภาษาไทย และแบบประเมินซึมเศร้ามาตรฐานของกรมสุขภาพจิต กระทรวงสาธารณสุข ระบบอนุญาตให้ผู้ใช้สนทนาด้วยภาษาธรรมชาติผ่านเว็บแอปพลิเคชัน โดยไม่ต้องเปิดเผยตัวตนและ เข้าถึงได้ตลอด 24 ชั่วโมง นอกจากนี้ระบบยังได้รับการออกแบบตามหลัก Privacy by Design และสอดคล้องกับ พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 เพื่อรองรับการเก็บชุดข้อมูลสุขภาพจิต ภาษาไทยแบบไม่ระบุตัวตน สำหรับการพัฒนา AI ด้านสุขภาพจิตในอนาคต

คำสำคัญ: *สุขภาพจิต, ปัญญาประดิษฐ์, การคัดกรองซึมเศร้า, RAG,* แบบประเมินซึมเศร้า, *LLM, Dataset.*

บทนำ

สุขภาพจิตเป็นองค์ประกอบสำคัญของสุขภาพโดยรวม อย่างไรก็ตาม ระบบบริการสุขภาพจิตในประเทศไทยยังคงประสบปัญหาด้านการเข้าถึงอย่างต่อเนื่อง ข้อมูลจากกรมสุขภาพจิต กระทรวงสาธารณสุข พบว่าในปี พ.ศ. 2566 มีผู้ป่วยโรคซึมเศร้าประมาณ 2.9 ล้านคน และอัตราส่วนจิตแพทย์ต่อประชากรอยู่ที่ 1 ต่อ 50,000 คน ซึ่งต่ำกว่าเกณฑ์ที่องค์การอนามัยโลกแนะนำอย่างมีนัยสำคัญ (WHO Thailand, 2025) สถานการณ์ดังกล่าวก่อให้เกิดช่องว่างระหว่างความต้องการและการได้รับบริการ ทำให้ผู้ป่วยส่วนใหญ่ไม่ได้รับการตรวจพบและรักษาอย่างทันท่วงทีในทศวรรษที่ผ่านมา เทคโนโลยีปัญญาประดิษฐ์โดยเฉพาะ Natural Language Processing ได้แสดงให้เห็นถึงศักยภาพอย่างมากในการสนับสนุนการดูแลสุขภาพจิต (Fitzpatrick et al., 2017; Shatte et al., 2019; Zhang et al., 2022) โมเดลภาษาขนาดใหญ่สามารถวิเคราะห์และตอบสนองต่อข้อความภาษาธรรมชาติได้อย่างมีประสิทธิภาพ แต่ส่วนใหญ่พัฒนาขึ้นสำหรับภาษาอังกฤษ ทำให้การประยุกต์ใช้กับภาษาไทยมีข้อจำกัดด้านความเข้าใจบริบทและวัฒนธรรม แบบประเมินซึมเศร้าที่กรมสุขภาพจิตปรับใช้เป็นแบบประเมินถือเป็นมาตรฐานระดับชาติสำหรับการคัดกรองโรคซึมเศร้าในบริบทสุขภาพจิตของประเทศไทย (Kroenke et al., 2001; Lotrakul et al., 2008; Kongsuk et al., 2018) อย่างไรก็ตาม การกรอกแบบประเมินด้วยกระดาษยังมีข้อจำกัดด้านการเข้าถึง และอาจก่อให้เกิดความลำเอียงจากผลของสังคมเมื่อผู้ป่วยทราบว่าบุคลากรทางการแพทย์กำลังสังเกตการณ์อยู่ โครงงานนี้จึงมีวัตถุประสงค์เพื่อ (1) พัฒนาระบบ AI สำหรับการคัด

กรองสุขภาพจิตเบื้องต้นในภาษาไทย โดยผสาน LLM เข้ากับ RAG และแบบประเมินสุขภาพจิตตามมาตรฐาน (2) ลดอุปสรรคในการเข้าถึงบริการสุขภาพจิตเบื้องต้นผ่านเว็บแอปพลิเคชันที่ไม่ระบุตัวตน และ (3) สร้างฐานข้อมูลการสนทนาด้านสุขภาพจิตภาษาไทยเพื่อสนับสนุนการวิจัย AI ในอนาคต

วิธีดำเนินการวิจัย

1. การออกแบบและสถาปัตยกรรมระบบ

ระบบน้องเอ็มออกแบบตามแนวทาง Microservices แบ่งเป็นส่วน RAG Pipeline สำหรับประมวลผลภาษาธรรมชาติ ส่วนประเมินสุขภาพจิตและจัดการข้อมูล และส่วนติดต่อผู้ใช้แบบเว็บแอปพลิเคชัน

2. กระบวนการดึงข้อมูลและสร้างคำตอบ

เทคโนโลยี Retrieval Augmented Generation (RAG) ที่ Lewis et al. (2020) นำเสนอ ช่วยลดปัญหา Hallucination ของ LLM ด้วยการดึงข้อมูลจากฐานความรู้จริงก่อนสร้างคำตอบ RAG Pipeline ของน้องเอ็มประกอบด้วยกระบวนการสำคัญ 3 ขั้นตอน ได้แก่ (1) การสกัดและเตรียมข้อความจากเอกสารการแพทย์ภาษาไทย ทั้งในรูปแบบข้อความและรูปภาพ โดยใช้ PyThaiNLP ในการ normalize ข้อความภาษาไทยก่อนประมวลผล (Phatthiyaphaibun et al., 2023) (2) การแปลงเอกสารเป็นเวกเตอร์เพื่อจัดเก็บในฐานข้อมูลเวกเตอร์ โดยใช้โมเดลภาษาแบบหลายภาษาตามแนวทางของ Reimers & Gurevych (2019) และ (3) การค้นหาด้วย Semantic Similarity ร่วมกับการ Reranking (Nogueira & Cho, 2019; Xiao et al., 2023) เพื่อคัดเลือกเอกสารที่เกี่ยวข้องมากที่สุด ซึ่งช่วยเพิ่มความแม่นยำในการค้นหาได้ 15–30% เมื่อเทียบกับการใช้ Embedding เพียงอย่างเดียว

3. การพัฒนาระบบสนทนาเพื่อประเมินสุขภาพจิตแบบครอบคลุมทุกช่วงวัย

ระบบน้องเอ็มได้รับการพัฒนาให้รองรับการคัดกรองภาวะซึมเศร้าสำหรับผู้ใช้ทุกช่วงวัย โดยใช้แบบประเมินมาตรฐานที่แตกต่างกันตามกลุ่มอายุ ซึ่งระบบจะเลือกแบบประเมินที่เหมาะสมโดยอัตโนมัติจากข้อมูลอายุที่ผู้ใช้ระบุในขั้นตอนก่อนเริ่มการประเมิน สำหรับกลุ่มเด็กอายุต่ำกว่า 11 ปี ระบบใช้แบบประเมิน CDI (Children's Depression Inventory) ซึ่งนำมาจากต้นฉบับภาษาไทยที่ใช้ในสถานพยาบาล แบบประเมินนี้ออกแบบมาเพื่อประเมินพฤติกรรมและการแสดงออกทางอารมณ์ของเด็ก เนื่องจากเด็กในวัยนี้อาจไม่สามารถรายงานอาการทางจิตใจของตนเองได้อย่างตรงไปตรงมาเหมือนผู้ใหญ่ โดยเน้นประเมินพฤติกรรม เช่น ความหงุดหงิด การเก็บตัว และความเบื่อหน่ายต่อกิจกรรมประจำวัน (Kovacs, 1992) สำหรับกลุ่มวัยรุ่นอายุ 11–20 ปี ระบบใช้แบบประเมิน PHQ-A (Patient Health Questionnaire for Adolescents) ซึ่งเป็นฉบับภาษาไทยที่นำมาจากกรมสุขภาพจิต กระทรวงสาธารณสุข แบบประเมินนี้มีโครงสร้าง 9 ข้อ ประเมินอาการย้อนหลัง 2 สัปดาห์ โดยปรับภาษาและบริบทให้เหมาะสมกับวัยรุ่น และให้ความสำคัญกับมิติความสัมพันธ์ทางสังคมกับเพื่อนและครอบครัว ผู้ใช้ตอบด้วยตัวเลือก 4 ระดับ ได้แก่ "ไม่มีเลย" (0) "เป็นบางวัน" (1) "เป็นบ่อย" (2) และ "เป็นทุกวัน" (3) โดยคะแนนรวมสูงสุด 24 คะแนน แปลผลเป็น 5 ระดับ ได้แก่ ไม่มีภาวะซึมเศร้า (0–4) เล็กน้อย (5–9) ปานกลาง (10–14) มาก (15–19) และรุนแรง (20–24) สำหรับกลุ่มผู้ใหญ่อายุ 20 ปีขึ้นไป ระบบใช้แบบประเมิน PHQ-9 ที่กรมสุขภาพจิตนำมาปรับใช้เป็นแบบประเมิน 9Q ซึ่งเป็นมาตรฐานระดับชาติสำหรับการคัดกรองโรคซึมเศร้าในบริบทสุขภาพจิตของประเทศไทย (Kroenke et al., 2001; Lotrakul et al., 2008; Kongsuk et al., 2018)

ผู้ใช้ตอบด้วยตัวเลือก 4 ระดับเช่นเดียวกัน คะแนนรวมสูงสุด 27 คะแนน แปลผลเป็น 5 ระดับ ได้แก่ ไม่มีภาวะซึมเศร้า (0–4) เล็กน้อย (5–9) ปานกลาง (10–14) ค่อนข้างรุนแรง (15–19) และรุนแรง (20–27) ในทุกแบบประเมิน ระบบสนทนาถูกออกแบบให้ถามคำถามทีละข้อผ่านภาษาธรรมชาติโดยใช้ LLM ภาษาไทย Typhoon (Pipatanakul et al., 2023, 2024) เป็นเครื่องมือในการดำเนินบทสนทนา โดยระบบจะตรวจสอบและคำนวณคะแนนจากคำตอบของผู้ใช้ฝั่ง backend เสมอ ไม่พึ่งพาการแปลผลจาก LLM โดยตรง เพื่อให้มั่นใจว่าคะแนนที่ได้มีความถูกต้องตามเกณฑ์มาตรฐานทางคลินิก ทั้งนี้ระบบมีกลไกพิเศษสำหรับคำถามที่เกี่ยวข้องกับความคิดทำร้ายตนเอง โดยแสดงข้อมูลสายด่วนสุขภาพจิต 1323 ทันทีหากผู้ใช้ตอบในระดับที่มีความเสี่ยง อย่างไรก็ตาม การประเมินผ่านบทสนทนาภาษาธรรมชาติมีข้อจำกัดด้าน validity ที่ควรระบุไว้ในการตีความผล กล่าวคือ คะแนนที่ได้อาจได้รับอิทธิพลจาก question-order effect และ conversational priming ทำให้ construct validity ของคะแนนอาจแตกต่างจากการทำแบบทดสอบมาตรฐาน (Batra et al., 2025; Dosovitsky et al., 2021) รวมถึงความเสี่ยงด้าน bias ที่พบในระบบ AI สำหรับสุขภาพจิตโดยทั่วไป (Timmons et al., 2023) ระบบจึงนำเสนอผลการประเมินทั้งหมดในฐานะ "การคัดกรองเบื้องต้น" เท่านั้น และแนะนำให้ผู้ใช้ที่มีคะแนนในระดับเสี่ยงพบผู้เชี่ยวชาญด้านสุขภาพจิตเพื่อการประเมินที่ละเอียดยิ่งขึ้น

4. จริยธรรมการวิจัยและการคุ้มครองข้อมูลส่วนบุคคล

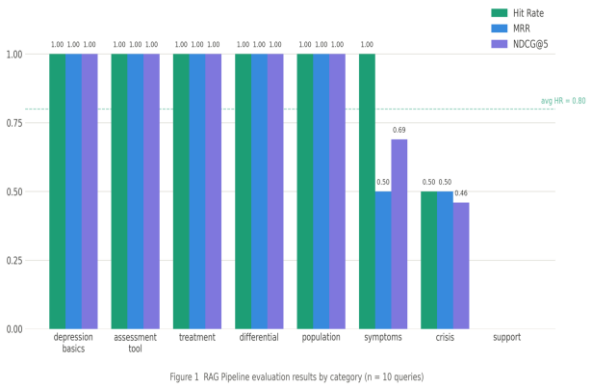
ระบบน้องเอ็มออกแบบตามหลักการ Privacy by Design ของ Cavoukian (2011) โดยไม่จัดเก็บข้อมูลที่สามารถระบุตัวตนของผู้ใช้ เช่น ชื่อ อีเมล หรือหมายเลขโทรศัพท์ และสอดคล้องกับ พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 ระบบมี หน้าต่างยินยอม (Consent Modal) 2 ระดับ ได้แก่ การยินยอมวิเคราะห์ ด้วย AI (บังคับ) และการ ยินยอมให้ใช้ข้อมูลเพื่อการวิจัย (ไม่บังคับ) รวมถึงรองรับสิทธิ์การลบข้อมูลผ่าน API Endpoint ข้อมูล ที่เก็บรวบรวมได้แก่ Session ID ที่ไม่ระบุตัวตน คะแนน แบบประเมิน ระดับความเสี่ยง และบทสนทนาในกรณีที่ผู้ใช้ยินยอมเพื่อการวิจัย ข้อมูลทั้งหมด จัดเก็บในรูปแบบ JSON เพื่อป้องกันการเสียหายของข้อมูล

5. การประเมินประสิทธิภาพ RAG Pipeline

ผู้พัฒนาทดสอบด้วยชุดคำถาม 10 ข้อ ครอบคลุม 8 หมวดหมู่ วัดผลด้วย Hit Rate, MRR และ NDCG ตามแนวทางของ Nogueira & Cho (2019)

สรุปผล

งานวิจัยนี้นำเสนอระบบน้องเอ็ม ซึ่งเป็น AI Chatbot สำหรับคัดกรองสุขภาพจิตเบื้องต้นที่ผสาน RAG Pipeline เข้ากับ LLM ภาษาไทยและแบบประเมินมาตรฐานครอบคลุมทุกช่วงวัย ได้แก่ CDI สำหรับเด็ก PHQ-A สำหรับวัยรุ่น และ PHQ-9 สำหรับผู้ใหญ่ ระบบได้รับการออกแบบให้สามารถดำเนินการสนทนาและคัดกรองสุขภาพจิตผ่านภาษาธรรมชาติ และออกแบบตามหลัก Privacy by Design ที่สอดคล้องกับ PDPA พ.ศ. 2562 เพื่อรองรับการเก็บชุดข้อมูลสุขภาพจิตภาษาไทยแบบไม่ระบุตัวตนเพื่อตอบโจทย์การนำข้อมูลไปใช้ต่อ



ภาพที่ 1

ผลการประเมินประสิทธิภาพ RAG Pipeline ด้วยชุดทดสอบ 10 queries ใน 8 หมวดหมู่ (ภาพที่ 1) พบว่าระบบมี Hit Rate 0.80 และ MRR 0.75 โดยมีค่า Latency เฉลี่ย 908.2 มิลลิวินาที ซึ่งอยู่ในระดับที่เหมาะสมสำหรับการใช้งานจริงผ่านเว็บแอปพลิเคชัน เมื่อพิจารณาแยกตามหมวดหมู่ พบว่า 5 หมวดหมู่หลัก ได้แก่ depression_basics, assessment_tool, treatment, differential และ population มีประสิทธิภาพสูงสุดโดยได้ NDCG@5 = 1.00 และ Hit Rate = 1.00 ทุกหมวด นอกจากนี้กลไก CrossEncoder Reranking ช่วยเลื่อนตำแหน่งเอกสารที่เกี่ยวข้องเฉลี่ย 2.23 ตำแหน่ง สอดคล้องกับงานวิจัยที่ระบุว่า CrossEncoder Reranking เพิ่มความแม่นยำได้ 15–30% เมื่อเทียบกับการใช้ Embedding เพียงอย่างเดียว (Nogueira & Cho, 2019; Xiao et al., 2023)

อย่างไรก็ตาม หมวด symptoms มี Hit Rate = 1.00 แต่ MRR ลดลงเหลือ 0.50 และ NDCG@5 = 0.69 ซึ่งบ่งชี้ว่าระบบสามารถดึงเอกสารที่เกี่ยวข้องได้ แต่เอกสารที่ตรงที่สุดไม่ได้ถูกจัดลำดับไว้อันดับแรกเสมอไป ส่วนหมวด crisis มี Hit Rate = 0.50 และ NDCG@5 = 0.46 โดยพบว่า query mh_004 ไม่สามารถดึงเอกสารที่เกี่ยวข้องได้เลย ซึ่งถือเป็นข้อจำกัดที่สำคัญที่สุดของระบบ เนื่องจากหมวด crisis เป็นหมวดที่มีความอ่อนไหวสูงและส่งผลโดยตรงต่อความปลอดภัยของผู้ใช้ และหมวด support มี Hit Rate = 0.00 ในทุก metric สะท้อนให้เห็นว่าฐานความรู้ปัจจุบันยังขาดเอกสารที่ครอบคลุมด้านการสนับสนุนทางสังคมและจิตใจอย่างมีนัยสำคัญ ผลการประเมินดังกล่าวเป็นเพียงการวัดประสิทธิภาพของ RAG retrieval เท่านั้น เนื่องจากการประเมินแบบ end-to-end ร่วมกับ LLM จะดำเนินการในระยะต่อไปเมื่อระบบมีความเสถียรมากขึ้น โดยในระยะถัดไปจะดำเนินการเพิ่มเอกสารในหมวด crisis และ support เป็นลำดับแรก เพื่อยกระดับประสิทธิภาพในหมวดที่ยังอ่อนแอและเพิ่มความปลอดภัยของระบบโดยรวม

ข้อเสนอแนะ

ปัจจุบันระบบอยู่ในระยะการทดสอบและเก็บข้อมูลเพื่อการพัฒนา โดยมีแผนการดำเนินงานใน 4 ด้านหลัก ได้แก่ (1) การพัฒนาระบบความปลอดภัยและ Safe Messaging ให้สอดคล้องกับแนวทางของกรมสุขภาพจิตและ WHO โดยเพิ่มการแทรกแซงทันทีเมื่อตรวจพบความเสี่ยงระหว่างบทสนทนา (2) การพัฒนาโมเดลภาษาไทยเฉพาะทางด้านสุขภาพจิต โดยใช้เทคนิค LoRA fine-tuning บนโมเดล Qwen2.5 ร่วมกับการศึกษาวิธีการ Quantum-Inspired LoRA Extension ซึ่งเป็นแนวทางใหม่ในการยกระดับ ประสิทธิภาพของ Small-Medium Language Model ให้สามารถประมวลผลบริบทด้านสุขภาพจิตภาษาไทย ได้ใกล้เคียงกับโมเดลขนาดใหญ่ เพื่อลดการพึ่งพา API ภายนอกและเผยแพร่เป็น open-source (3) การพัฒนาโมดูลตรวจจับอารมณ์และเจตนา (Emotion/Intent Detection) โดยใช้ BiLSTM ร่วมกับ WangchanBERTa ตามแนวทาง aRAG ของ Ravenda et al. (2025) เพื่อเสริมการประเมินเชิงคลินิกโดยไม่ override คะแนนแบบทดสอบ และ (4) การพัฒนา Thai Mental Health NLP Open Source Dataset เพื่อสนับสนุนการวิจัย AI สุขภาพจิตภาษาไทยในวงกว้าง ระบบนี้มีศักยภาพในการช่วยลดช่องว่างระหว่างความต้องการบริการสุขภาพจิตและความสามารถในการให้บริการ โดยเฉพาะในพื้นที่ห่างไกลที่ขาดแคลนจิตแพทย์ (Shin et al., 2024)

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณครูเอกชัย วัฒนไชย ครูที่ปรึกษาโครงงานที่ให้การสนับสนุนและคำแนะนำอันเป็นประโยชน์ ตลอดการดำเนินโครงงานนี้

เอกสารอ้างอิง

Cavoukian, A. (2011). Privacy by design: The 7 foundational principles. Toronto: Information and Privacy Commissioner of Ontario.

Kovacs, M. (1992). Children's Depression Inventory (CDI) manual. New York: Multi-Health Systems.

Batra, S., et al. (2025). Differences in PHQ-9 scores by administration mode in adolescents: A population-based study. Journal of Primary Care & Community Health. 16: 1-8.

Dosovitsky, G., Pineda, B. S., Jacobson, N. C., Chang, C., Aguilera, A., & Bunge, E. L. (2021). Artificial intelligence chatbot for depression: Descriptive study of usage and acceptability. JMIR Formative Research. 5(9): e28914.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health. 4(2): e19.

Kongsuk, T., Arunpongpaisal, S., Janthong, S., Prukkanone, B., Sukhawaha, S., & Leejongpermpoon, J. (2018). Criterion-related validity of the 9 questions depression rating scale revised for Thai central dialect. Journal of the Psychiatric Association of Thailand. 63(4): 321-334.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine. 16(9): 606-613.

Lotrakul, M., Sumrithe, S., & Saipanish, R. (2008). Reliability and validity of the Thai version of the PHQ-9. BMC Psychiatry. 8: 46.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. Psychological Medicine. 49(9): 1426-1448.

Shin, D., Kim, H., Lee, S., Cho, Y., & Jung, W. (2024). Using large language models to detect depression from user-generated diary text data as a novel approach in digital mental health screening: Instrument validation study. Journal of Medical Internet Research. 26: e54617.

Timmons, A. C., Duong, J. B., Simo Fiallo, N., Lee, T., Vo, H. P. Q., Ahle, M. W., Comer, J. S., Brewer, L. C., Frazier, S. L., & Chaspari, T. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. Perspectives on Psychological Science. 18(5): 1062-1096.

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A systematic review. Journal of Medical Internet Research. 24(4): e37363.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. ใน Advances in Neural Information Processing Systems 33 (NeurIPS 2020). หน้า 9459–9474. Vancouver: Curran Associates.

Nogueira, R., & Cho, K. (2019). Passage re-ranking with BERT. ใน arXiv preprint

arXiv:1901.04085. สืบค้นเมื่อ 3 เมษายน 2568, จาก https://doi.org/10.48550/arXiv.1901.04085

Phatthiyaphaibun, W., Chaovavanich, K., Polpanumas, C., Suriyawongkul, A., Lowphansirikul, L., Chormai, P., Limkonchotiwat, P., Suntorntip, T., & Udomcharoenchaikit, C. (2023). PyThaiNLP: Thai natural language processing in Python. ใน Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023). หน้า 25-36. Singapore: Association for Computational Linguistics.

Pipatanakul, K., Jirabovonvisut, P., Manakul, P., Sripaisarnmongkol, S., Patomwong, R., Chokchainant, P., & Tharnpipitchai, K. (2023). Typhoon: Thai large language models. ใน arXiv preprint arXiv:2312.13951. สืบค้นเมื่อ 8 เมษายน 2568, จาก https://doi.org/10.48550/arXiv.2312.13951

Pipatanakul, K., Manakul, P., Nitarach, N., Sirichotedumrong, W., Nonesung, S., Jaknamon, T., Pengpun, P., Taveekitworachai, P., Na-Thalang, A., Sripaisarnmongkol, S., Jirayoot, K., & Tharnpipitchai, K. (2024). Typhoon 2: A family of open text and multimodal Thai large language models. ใน arXiv preprint arXiv:2412.13702. สืบค้นเมื่อ 9 เมษายน 2568, จาก https://doi.org/10.48550/arXiv.2412.13702

Ravenda, M., et al. (2025). Are LLMs effective psychological assessors? Leveraging adaptive RAG for interpretable mental health screening through psychometric practice. ใน Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Long Papers). หน้า 8975–8991. Bangkok: Association for Computational Linguistics.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. ใน Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). หน้า 3982-3992. Hong Kong: Association for Computational Linguistics.

Xiao, S., Liu, Z., Zhang, P., & Muennighoff, N. (2023). C-Pack: Packaged resources to advance general Chinese embedding. ใน arXiv preprint arXiv:2309.07597. สืบค้นเมื่อ 4 เมษายน 2568, จาก https://doi.org/10.48550/arXiv.2309.07597

Guo, Y., et al. (2025). Development and evaluation of HopeBot: An LLM-based chatbot for structured and interactive PHQ-9 depression screening. สืบค้นเมื่อ 1 พฤษภาคม 2568, จาก https://doi.org/10.48550/arXiv.2507.05984

WHO Thailand. (2025). Mental health and social connection in Thailand. Geneva: World Health Organization. สืบค้นเมื่อ 17 เมษายน 2568, จาก https://www.who.int/thailand/news/feature-stories/detail/mental-health-and-social-connection-in-thailand