วารสารแม่โจ้เทคโนโลยีสารสนเทศและนวัตกรรม ปีที่ 12 ฉบับที่ 3 กันยายน – ธันวาคม 2569

Received: 27 ม.ค. 2569 **Revised:** 20 มี.ค. 2569 **Accepted:** 10 เม.ย. 2569

ระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะด้วย RAG และ AI Agent

AI Agent-based Internet Promotion Recommendation System using RAG

อธิศ ฟองไข่มุกข์[1], ชัยยุทธ สุนทะโรจน์[1], สมนึก สินธุปวน[1], พาสน์ ปราโมกข์ชน[1] และ ก่องกาญจน์ ดุลยไชย[1]*

[1]สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้

Athit Fongkhaimuk[1], Chaiyut Suntharote[1], Somnuek Sinthupuan[1], Part Pramokchon[1] and Kongkarn Dullayachai[1*]

[1]Computer Science Department, Faculty of Science, Maejo University, Chaingmai, Thailand, 50290

*Corresponding author: kongkarn@gmaejo.mju.ac.th

Abstract

This study presents an intelligent internet promotion recommendation system designed to address the inaccuracy of chatbots that rely solely on large language models. The system integrates Retrieval-Augmented Generation (RAG) with an AI Agent to retrieve domain-specific knowledge and generate accurate responses. It is implemented using FastAPI and PostgreSQL to support web-based conversational interactions.

The system was evaluated using 50 conversation cases with both AI Judge and rule-based evaluation. The results indicate good to very good performance, achieving scores of 4.60 for politeness, 4.28 for relevance, and 4.98 for readability. These findings demonstrate that the system can provide accurate, clear, and user-friendly recommendations suitable for customer service applications.

Keywords: *Intelligent Internet Promotion Recommendation System; AI Agent; Retrieval-Augmented Generation (RAG); Open-WebUI; Gemini LLM; FastAPI; PostgreSQL*

บทคัดย่อ

ระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะด้วย RAG และ AI Agent นี้พัฒนาขึ้นเพื่อแก้ปัญหาความไม่แม่นยำของแชตบอทที่ใช้โมเดลภาษาขนาดใหญ่ โดยประยุกต์ใช้เทคนิค Retrieval-Augmented Generation (RAG) ร่วมกับ AI Agent เพื่อดึงข้อมูลจากคลังความรู้เฉพาะและสร้างคำตอบที่ถูกต้อง ระบบถูกพัฒนาด้วย FastAPI และเชื่อมต่อฐานข้อมูล PostgreSQL เพื่อรองรับการใช้งานผ่านเว็บแชต

ผลการประเมินจากบทสนทนา 50 ชุด โดยใช้ AI Judge และการประเมินเชิงกฎ พบว่าระบบมีประสิทธิภาพอยู่ในระดับดีถึงดีมาก โดยมีคะแนนความสุภาพ 4.60 ความตรงประเด็น 4.28 และความง่ายในการอ่าน 4.98 แสดงให้เห็นว่าระบบสามารถให้คำแนะนำได้อย่างถูกต้อง ชัดเจน และเหมาะสมต่อการใช้งานในงานบริการลูกค้า

คำสำคัญ: *AI Agent; Retrieval-Augmented Generation (RAG); Open-WebUI; Gemini LLM; ระบบแนะนำโปรโมชั่นอินเทอร์เน็ต*

1. บทนำ

ในยุคปัจจุบัน ปัญญาประดิษฐ์ (Artificial Intelligence: AI) และโมเดลภาษาขนาดใหญ่ (Large Language Models: LLMs) ถูกนำมาใช้ในงานบริการลูกค้าและการตัดสินใจทางธุรกิจอย่างแพร่หลาย เนื่องจากสามารถประมวลผลข้อมูลขนาดใหญ่และโต้ตอบกับผู้ใช้งานด้วยภาษาธรรมชาติได้อย่างมีประสิทธิภาพ (Google AI, 2026; Open-WebUI Team, 2026) โดยเฉพาะในธุรกิจโทรคมนาคมที่มีข้อมูลแพ็กเกจและโปรโมชั่นจำนวนมาก การใช้แชตบอทอัจฉริยะจึงถูกมองว่าเป็นแนวทางสำคัญในการช่วยให้ลูกค้าเข้าถึงข้อมูลได้รวดเร็วและแม่นยำยิ่งขึ้น

อย่างไรก็ตาม งานวิจัยที่ผ่านมาได้ชี้ให้เห็นว่าแชตบอทที่ใช้ LLM เพียงอย่างเดียวมักประสบปัญหาการสร้างคำตอบที่ไม่ตรงกับข้อมูลจริง รวมถึงขาดความสามารถในการอ้างอิงข้อมูลจากแหล่งความรู้เฉพาะขององค์กร (Madaan et al., 2024) ส่งผลให้การนำ LLM ไปใช้ในงานที่ต้องการความถูกต้องสูง เช่น การแนะนำโปรโมชั่นอินเทอร์เน็ต อาจทำให้เกิดความคลาดเคลื่อนและลดความเชื่อมั่นของลูกค้า นอกจากนี้ แม้ว่าจะมีงานวิจัยที่ผสานระบบแนะนำกับแชตบอทเพื่อปรับคำแนะนำให้เหมาะสมกับผู้ใช้ (Ganesh et al., 2024) และงานที่ศึกษาบทบาทของรูปแบบการสื่อสารของแชตบอทต่อความพึงพอใจของลูกค้า (Thomas et al., 2022) แต่ยังขาดงานที่บูรณาการ RAG (Lewis et al., 2020), LLM และ AI Agent เข้าด้วยกันในบริบทของ การแนะนำโปรโมชั่นอินเทอร์เน็ตในองค์กรจริง ด้วยเหตุนี้ จึงเกิด

ช่องว่างทางการวิจัยในการพัฒนาระบบที่สามารถให้คำแนะนำโปรโมชั่นได้อย่างถูกต้อง อ้างอิงจากข้อมูลจริง และสื่อสารกับลูกค้าได้อย่างเป็นธรรมชาติและเหมาะสมกับบริบททางธุรกิจ

บริษัทโทรคมนาคมแห่งชาติ จำกัด (มหาชน) จึงได้จัดตั้งกลุ่ม “Nexus” ซึ่งประกอบด้วยผู้เชี่ยวชาญด้าน AI และ LLM จากหลายแผนก เพื่อพัฒนาระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะด้วย RAG และ AI Agent โดยประยุกต์ใช้เทคนิค Retrieval-Augmented Generation (RAG) ร่วมกับแพลตฟอร์ม Open-WebUI และโมเดล LLM ผ่าน API เพื่อแก้ไขข้อจำกัดของแชตบอทแบบเดิม และยกระดับการให้บริการลูกค้าในด้านความแม่นยำ ความน่าเชื่อถือ และประสบการณ์ผู้ใช้

2. วัตถุประสงค์

วัตถุประสงค์หลัก 4 ประการ:

1. เพื่อศึกษาและประยุกต์ใช้การจัดการโมเดลภาษาขนาดใหญ่แบบแยกตามประเภทโปรโมชั่น (อินเทอร์เน็ตบ้านและอินเทอร์เน็ตมือถือ) ผ่านระบบ AI Agent เพื่อเพิ่มความเหมาะสมของคำตอบต่อบริบทของผู้ใช้งาน
2. เพื่อพัฒนาแชตบอทอัจฉริยะสำหรับแนะนำโปรโมชั่นอินเทอร์เน็ต โดยใช้โมเดลภาษาขนาดใหญ่ร่วมกับเทคนิค Retrieval-Augmented Generation (RAG) เพื่อให้สามารถตอบคำถามจากข้อมูลในคลังความรู้ได้อย่างถูกต้องและสอดคล้องกับเงื่อนไขโปรโมชั่นจริง
3. เพื่อประเมินความแม่นยำและคุณภาพของคำตอบของระบบ โดยวัดระดับความสอดคล้องของคำตอบกับข้อมูลในคลังความรู้ รวมถึงตัวชี้วัดด้านความตรงประเด็น (Relevance) ความครบถ้วนของข้อมูล (Completeness) และความเหมาะสมของการนำเสนอ
4. เพื่อประเมินประสิทธิภาพเชิงการใช้งานของระบบแชตบอท จากบทสนทนาจริงของผู้ใช้ ผ่านกรอบการประเมินแบบผสมผสานระหว่าง AI-based evaluation และ rule-based evaluation

3. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องสามารถจัดเป็น 3 ประเด็นหลัก ได้แก่ (1) Retrieval-Augmented Generation (RAG) (2) ระบบแนะนำที่ผสานแชตบอท และ (3) คุณภาพการสื่อสารของ Conversational AI

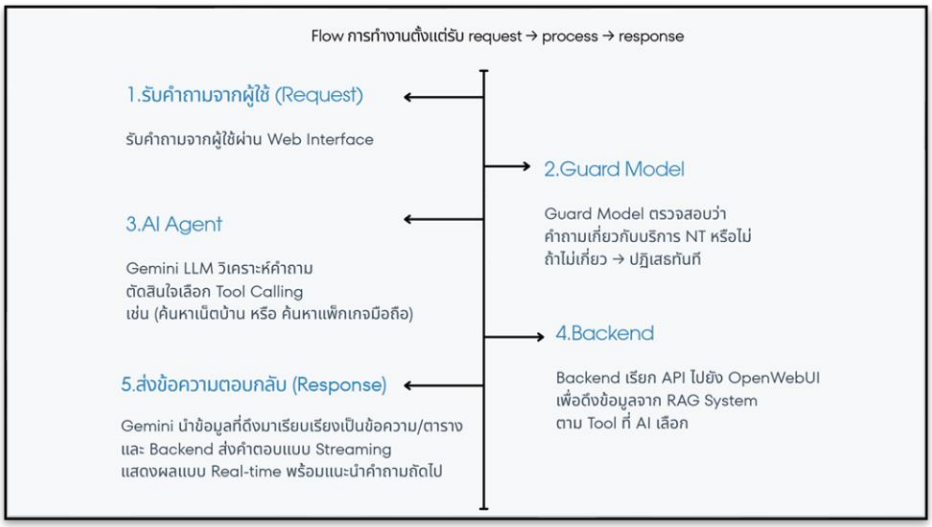
1) Retrieval-Augmented Generation (RAG) แนวคิด RAG ถูกพัฒนาขึ้นเพื่อแก้ปัญหา hallucination ของ LLM โดยผสานกลไกดึงข้อมูลกับการสร้างคำตอบ (Lewis et al., 2020) ทำให้โมเดลสามารถอ้างอิงข้อมูลจากคลังความรู้ได้โดยตรง งานของ Madaan et al. (2024) แสดงให้เห็นว่า RAG ให้คุณภาพคำตอบสูงกว่าวิธีพื้นฐานในเมตริก เช่น ROUGE และ BERTScore จึงเหมาะกับระบบที่ต้องการความถูกต้องของข้อมูลเชิงข้อเท็จจริง เช่น ระบบแนะนำโปรโมชั่น

2) ระบบแนะนำร่วมกับแชตบอท Ganesh et al. (2024) เสนอการผสาน LLM เข้ากับ collaborative และ content-based filtering เพื่อสร้างระบบแนะนำที่สามารถอธิบายเหตุผลและปรับตามผู้ใช้แบบเรียลไทม์ แนวคิดนี้สนับสนุนการพัฒนา Conversational Recommendation System ที่ให้คำแนะนำเฉพาะบุคคล

3) คุณภาพการสื่อสารของ Conversational AI Thomas et al. (2022) พบว่าสไตล์การตอบแบบ social-oriented ช่วยเพิ่มความพึงพอใจมากกว่าสไตล์เชิงภารกิจเพียงอย่างเดียว ชี้ให้เห็นว่าระบบแนะนำควรคำนึงถึงโทนภาษา ความสุภาพ และประสบการณ์ผู้ใช้ควบคู่กับความถูกต้องของข้อมูล สรุปงานวิจัยนี้จึงผสานแนวคิด RAG เพื่อความแม่นยำของข้อมูล แนวคิดระบบแนะนำเพื่อความเฉพาะบุคคล และหลักการสื่อสารของ Conversational AI เพื่อยกระดับประสบการณ์ผู้ใช้ในบริบทงานบริการลูกค้า

4. วิธีการดำเนินการ

การพัฒนาระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะด้วย RAG และ AI Agent นี้ อาศัยแนวคิดการพัฒนาระบบเชิงสถาปัตยกรรมแบบแยกส่วน (Modular Architecture) โดยผสานเทคโนโลยีปัญญาประดิษฐ์ร่วมกับ Retrieval-Augmented Generation (RAG) จาก Open-WebUI เพื่อให้ระบบสามารถให้คำตอบที่ถูกต้องและอ้างอิงข้อมูลจริงจากคลังความรู้ได้ กระบวนการทำงานโดยรวมของระบบสามารถสรุปได้ดังนี้



ภาพที่ 1 Workflow การทำงานของระบบ

รายละเอียดการทำงานของระบบ

1. รับคำถามจากผู้ใช้ (Frontend Service)

ผู้ใช้พิมพ์คำถามผ่านหน้าเว็บแชทของระบบ เช่น สอบถามเกี่ยวกับโปรโมชั่นอินเทอร์เน็ต ราคา ความเร็ว หรือแพ็กเกจที่เหมาะสม จากนั้นคำถามจะถูกส่งมายัง Backend ของระบบที่พัฒนาด้วย FastAPI เพื่อเริ่มกระบวนการประมวลผล

2. Guard Model ตรวจสอบคำถาม (Validation Service)

ระบบจะส่งคำถามไปยัง Guard Model เพื่อตรวจสอบว่าคำถามเกี่ยวข้องกับบริการอินเทอร์เน็ตของ NT หรือไม่ หากคำถามไม่อยู่ในขอบเขตของระบบ ระบบจะปฏิเสธคำถามทันที เพื่อลดการประมวลผลที่ไม่จำเป็นและควบคุมขอบเขตของ AI

3. AI Agent วิเคราะห์คำถามและเลือกเครื่องมือ (AI Orchestrator Service)

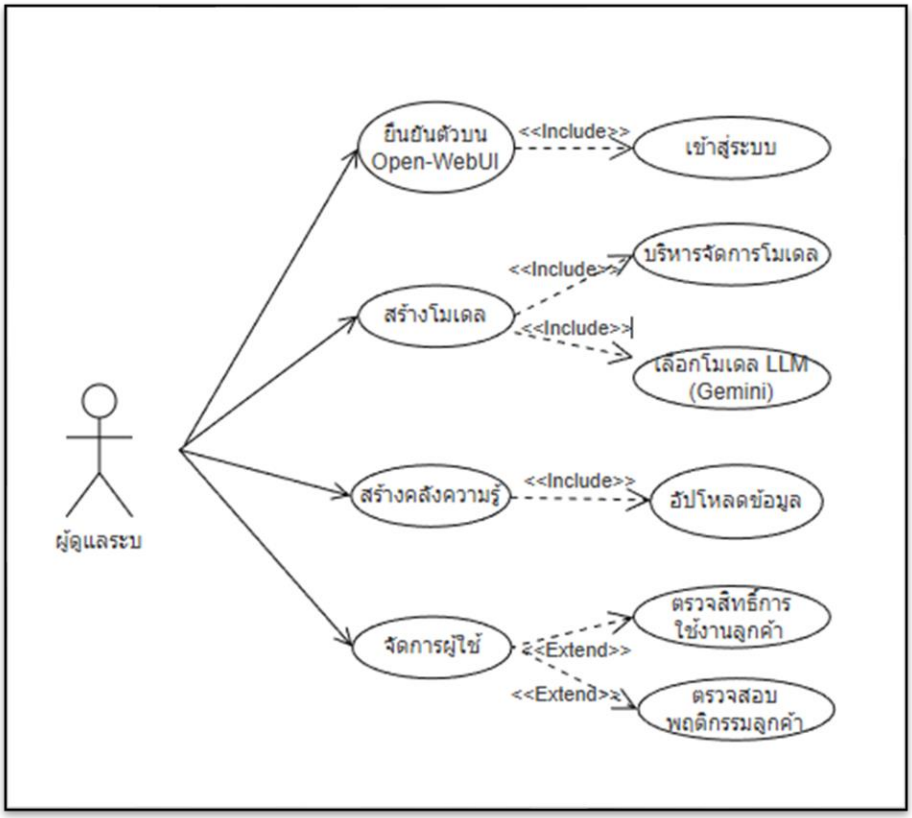
ในขั้นตอนนี้ AI Agent ซึ่งใช้โมเดลภาษา Gemini จะทำการวิเคราะห์เจตนาของผู้ใช้ (User Intent) เพื่อทำความเข้าใจว่าคำถามเกี่ยวข้องกับบริการประเภทใด เช่น อินเทอร์เน็ตบ้านหรืออินเทอร์เน็ตมือถือ จากนั้น AI Agent จะใช้แนวคิด Tool Calling เพื่อเลือกเครื่องมือหรือโมเดลที่เหมาะสมสำหรับการค้นหาข้อมูลที่เกี่ยวข้อง

4. Backend และ OpenWebUI ดึงข้อมูลจาก Knowledge Base ด้วย RAG Pipeline (Knowledge Service)

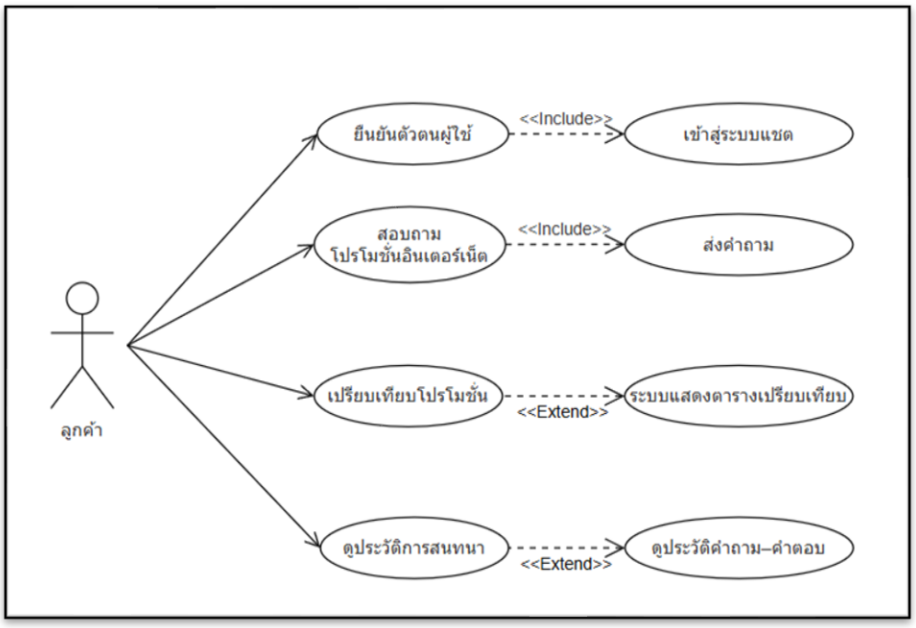
Backend จะเรียกใช้ API ไปยังระบบ Open-WebUI ตาม Tool ที่ AI Agent เลือก จากนั้น Open-WebUI จะค้นหาข้อมูลที่เกี่ยวข้องจาก Knowledge Base ของระบบ เช่น ข้อมูลแพ็กเกจอินเทอร์เน็ตหรือรายละเอียดโปรโมชั่น โดยใช้กระบวนการ RAG Pipeline (Retrieval-Augmented Generation) เพื่อดึงข้อมูลที่เกี่ยวข้องมาใช้เป็นบริบทในการสร้างคำตอบ

5. สร้างและส่งคำตอบกลับไปยังผู้ใช้ (Response)

โมเดล Gemini จะนำข้อมูลที่ได้จาก RAG Pipeline มาประมวลผลร่วมกับคำถามของผู้ใช้เพื่อสร้างคำตอบที่ถูกต้องและสอดคล้องกับข้อมูลจริง จากนั้น Backend จะส่งคำตอบกลับไปยังหน้าเว็บแชทแบบ Streaming Response ทำให้ผู้ใช้สามารถเห็นคำตอบแบบ Real-time และสามารถสอบถามข้อมูลเพิ่มเติมต่อได้ทันที



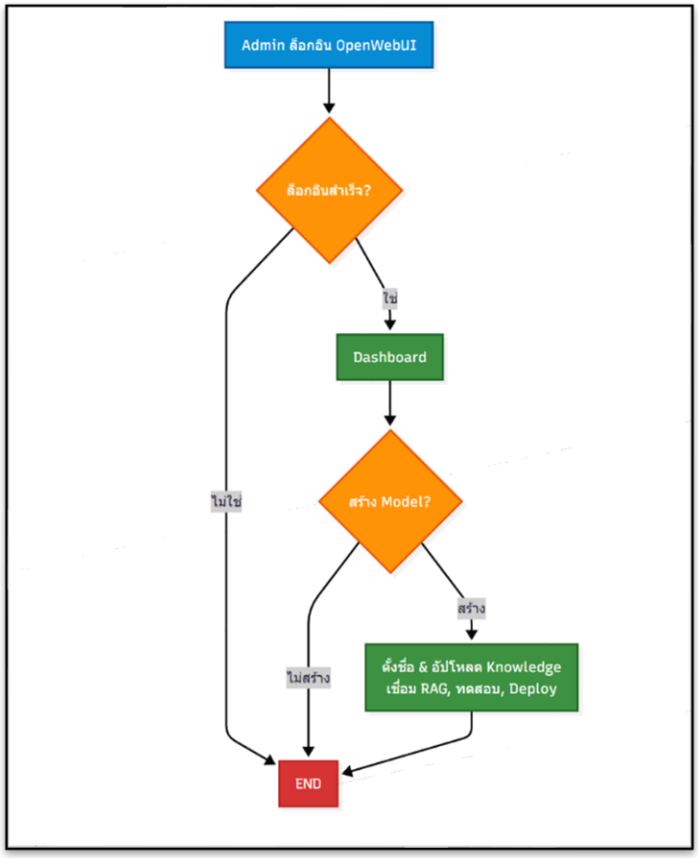
ภาพที่ 2 แผนภาพยูสเคสของผู้ดูแลระบบ



ภาพที่ 3 แผนภาพยูสเคสของลูกค้า

4.6 การทำงานของระบบ (System Flowchart)

4.6.1 รายละเอียดและแผนภาพการทำงานของผู้ดูแลระบบ

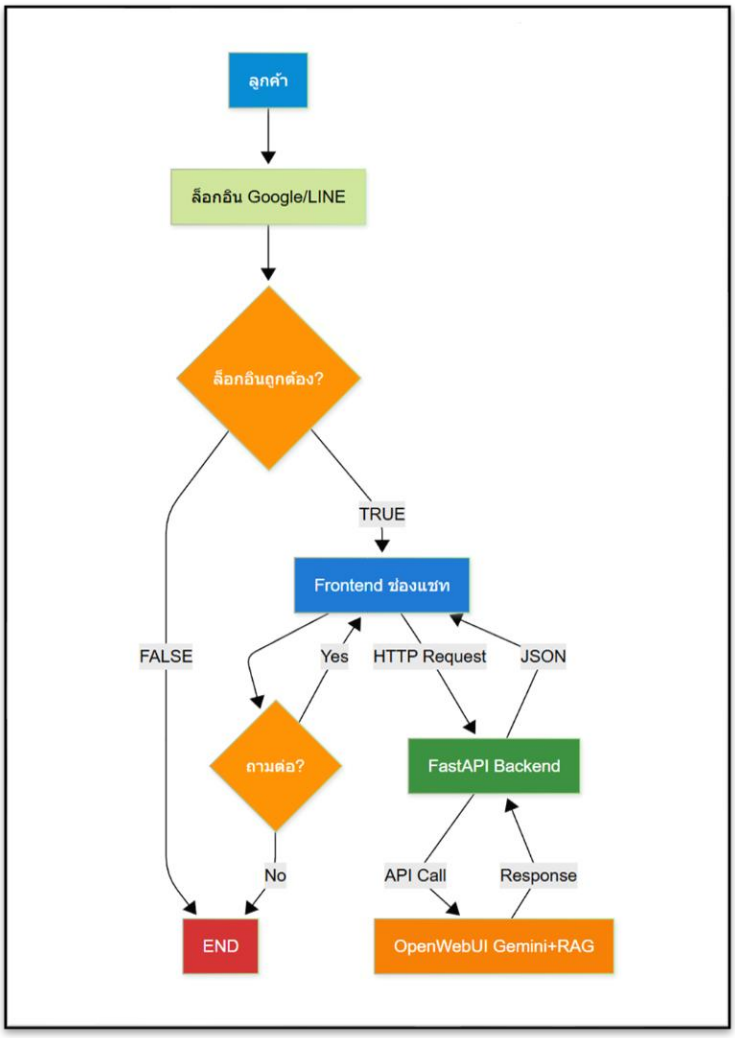


ภาพที่ 4 การทำงานของผู้ดูแลระบบ

รายละเอียดการทำงานของผู้ดูแลระบบ

ผู้ดูแลระบบเริ่มต้นจากการเข้าสู่ระบบ OpenWebUI หากการยืนยันตัวตนไม่ถูกต้อง ระบบจะสิ้นสุดการทำงานทันทีเมื่อเข้าสู่ระบบสำเร็จ ผู้ดูแลระบบจะเข้าสู่หน้า Dashboard เพื่อจัดการโมเดลปัญญาประดิษฐ์ โดยสามารถเลือกสร้างโมเดลใหม่ได้ ในกรณีที่มีการสร้างโมเดล ผู้ดูแลระบบจะดำเนินการตั้งค่าโมเดล อัปโหลดองค์ความรู้ เชื่อมต่อระบบ Retrieval-Augmented Generation (RAG) ทดสอบการทำงานของโมเดล และนำโมเดลไปใช้งานจริง (Deploy) เมื่อกระบวนการเสร็จสิ้น ระบบจะยุติการทำงานตามลำดับขั้นตอนที่กำหนดไว้

4.6.2 รายละเอียดและแผนภาพการทำงานของลูกค้า



ภาพที่ 5 ระบบการทำงานของลูกค้า

รายละเอียดการทำงานของลูกค้า

ลูกค้าเริ่มต้นจากการเข้าสู่ระบบผ่าน Google หรือ LINE หากการยืนยันตัวตนไม่ถูกต้อง ระบบจะยุติการทำงานทันที เมื่อเข้าสู่ระบบสำเร็จ ลูกค้าสามารถเข้าถึงหน้าถาม-ตอบเพื่อสอบถามข้อมูลโปรโมชั่นอินเทอร์เน็ต โดยระบบจะส่งคำถามในรูปแบบ HTTP/JSON ไปยัง FastAPI Backend ซึ่งทำหน้าที่ตรวจสอบ Token และจัดรูปแบบ Prompt ก่อนส่งต่อไปยังระบบ Open-WebUI ที่ผสานการทำงานร่วมกับ Gemini และเทคนิค Retrieval-Augmented Generation (RAG) จากนั้นระบบจะประมวลผลและส่งผลลัพธ์เป็นข้อมูล JSON กลับมายังหน้าถาม-ตอบ เพื่อแสดงผลแก่ลูกค้า ทั้งนี้ลูกค้าสามารถเลือกสอบถามข้อมูลเพิ่มเติมได้จนกว่าจะสิ้นสุดการใช้งานระบบ

5. ผลการทดสอบระบบ

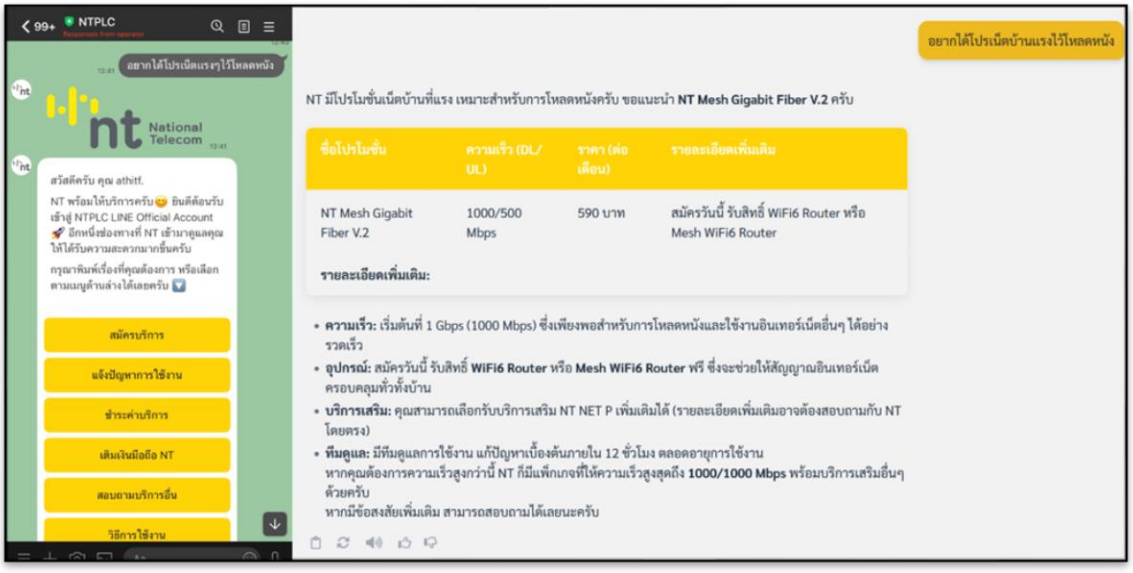
ชุดข้อมูลที่ใช้ทดสอบ (Dataset) ชุดข้อมูลที่ใช้ในงานวิจัยนี้แบ่งออกเป็น 2 ส่วน ได้แก่

(1) ข้อมูลคลังความรู้สำหรับระบบ RAG

(2) ชุดข้อมูลสำหรับการประเมินผล

(1) ข้อมูลคลังความรู้ (Knowledge Base) ข้อมูลที่ใช้ในการฝัง (embedding) ลงในคลังความรู้เป็นข้อมูลรายละเอียดโปรโมชั่นอินเทอร์เน็ตจริงขององค์กรนามสกุลไฟล์ CSV ซึ่งประกอบด้วยข้อมูลด้านราคา แพ็กเกจ และเงื่อนไขการใช้งาน โดยแบ่งเป็น โปรโมชั่นอินเทอร์เน็ตบ้าน จำนวน 4 ไฟล์ โปรโมชั่นอินเทอร์เน็ตมือถือ จำนวน 11 ไฟล์ ทั้งนี้ ข้อมูลดังกล่าวเป็นข้อมูลจริงที่เคยใช้งานในองค์กร แต่เป็นข้อมูลที่ไม่ได้ใช้งานแล้วในปัจจุบัน เพื่อหลีกเลี่ยงผลกระทบเชิงธุรกิจและรักษาความเป็นส่วนตัวของข้อมูล

(2) ชุดข้อมูลสำหรับการประเมินผล (Evaluation Dataset) ชุดข้อมูลที่ใช้ทดสอบประสิทธิภาพของระบบประกอบด้วยบทสนทนาจำนวน 50 ชุด ในรูปแบบคู่คำถาม–คำตอบ (Question–Answer Pairs) โดยคำถามถูกออกแบบให้ครอบคลุมลักษณะการใช้งานจริงของผู้ใช้ เช่น การสอบถามโปรโมชั่นที่เหมาะสม, การเปรียบเทียบแพ็กเกจ, การสอบถามเงื่อนไขและรายละเอียดของโปรโมชั่นคำตอบในแต่ละชุดถูกสร้างโดยระบบ AI Agent ที่ผสานเทคนิค Retrieval-Augmented Generation (RAG) จากคลังข้อมูลดังกล่าว โดยไม่มีการปรับแก้โดยมนุษย์ เพื่อสะท้อนสมรรถนะของระบบในสภาพแวดล้อมการใช้งานจริง



ภาพที่ 6 ภาพเปรียบเทียบระบบใหม่กับระบบเก่าที่ไม่มี RAG

วิธีการประเมินผล (Evaluation Method)

การประเมินมุ่งเน้นที่คุณภาพของคำตอบเชิงภาษาและประสบการณ์ผู้ใช้ เนื่องจากระบบเป็น Conversational AI ที่สร้างคำตอบในรูปแบบภาษาธรรมชาติ ซึ่งอาจมีความถูกต้องได้หลายรูปแบบ และไม่สามารถวัดผลด้วยตัวชี้วัดเชิงสถิติ เช่น Accuracy หรือ F1-score ได้โดยตรง ดังนั้น งานวิจัยนี้จึงใช้แนวทาง Hybrid Evaluation ประกอบด้วย

1. การประเมินคุณภาพเนื้อหาด้วย AI Judge (LLM-as-a-Judge)
2. การประเมินคุณภาพการนำเสนอด้วยกฎเชิงโครงสร้าง (Rule-Based Evaluation)

โดยเกณฑ์การให้คะแนนมีทั้งหมด 3 ประเภทดังนี้

1. คะแนนค่าเฉลี่ยในช่วง 1–5 คะแนน

การคำนวณคะแนนเฉลี่ยใช้สูตรดังนี้:

$$Average\ Score = \frac{\Sigma \frac{n}{i=1} Score_i}{n}$$

โดยที่

- $Score_i$ คือคะแนนของบทสนทนาแต่ละชุด
- n คือจำนวนชุดทดสอบทั้งหมด (50 ชุด)

2. คะแนนค่าเบี่ยงเบนมาตรฐาน

การคำนวณค่าเบี่ยงเบนมาตรฐานใช้สูตรดังนี้:

$$SD = \sqrt{\frac{\Sigma(Score_i - \overline{Score})^2}{n-1}}$$

โดยที่

- $Score_i$ คือ คะแนนของชุดทดสอบลำดับที่ i
- $\overline{Score}$ ค่าเฉลี่ยของคะแนนทั้งหมด
- n คือจำนวนชุดทดสอบทั้งหมด (50 ชุด)

3. เปรียบเทียบค่าเฉลี่ยกับเกณฑ์การแปลผลของ Likert Scale ในช่วง 1–5 คะแนน โดยเกณฑ์การแปลผลคะแนน แบ่งออกเป็น 5 ระดับ

ช่วงคะแนนเฉลี่ย	ระดับการประเมิน
4.51 – 5.00	มากที่สุด
3.51 – 4.50	มาก
2.51 – 3.50	ปานกลาง
1.51 – 2.50	น้อย
1.00 – 1.50	น้อยที่สุด

ตารางที่ 1 เกณฑ์การแปลผลของ Likert Scale

5.1 ผลการประเมินคุณภาพเนื้อหาด้วย AI Judge (G-Eval / Chain-of-Thought)

5.1.1 หลักการประเมิน

การประเมินในส่วนนี้อ้างอิงแนวคิด LLM-as-a-Judge ซึ่งเป็นแนวทางที่ใช้โมเดลภาษาขนาดใหญ่ทำหน้าที่ประเมินคุณภาพของข้อความ โดยพิจารณาบริบทของคำถามและคำตอบร่วมกัน วิธีการดังกล่าวเหมาะสมกับระบบที่ใช้ Large Language Model และ AI Agent ซึ่งไม่สามารถวัดผลด้วยค่า Accuracy หรือ Precision ได้โดยตรง โมเดล AI Judge ทำการให้คะแนนในช่วง 1–5 คะแนนตามมาตราส่วนแบบ Likert Scale โดยพิจารณา 3 มิติ ได้แก่

1. ความสุภาพ (Politeness) หมายถึง การตอบกลับของระบบใช้ภาษาที่สุภาพ เหมาะสมกับบริบทการให้บริการลูกค้า และมีลักษณะเป็นมิตร
2. ความตรงประเด็น (Relevance) หมายถึง ระบบสามารถเข้าใจเจตนาของลูกค้า (User Intent) ได้อย่างถูกต้อง และเลือกนำเสนอข้อมูลที่สอดคล้องกับคำถาม โดยเฉพาะการแยกประเภทบริการ เช่น อินเทอร์เน็ตบ้านและอินเทอร์เน็ตมือถือ
3. ความครบถ้วนสมบูรณ์ของข้อมูล (Completeness) หมายถึง พฤติกรรมของ AI Agent ที่มุ่งเน้นการให้ข้อมูลที่จำเป็นอย่างครบถ้วน โดยมีความกระชับ และลดภาระจากข้อมูลส่วนเกินที่ไม่เกี่ยวข้อง

5.1.2 สรุปผลการประเมินคุณภาพเนื้อหาด้วย AI Judge

จากการประเมินบทสนทนาจำนวน 50 ชุด ได้ผลคะแนนเฉลี่ยดังนี้

มิติการประเมิน	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมาย
ความสุภาพ (Politeness)	4.60	0.42	มากที่สุด
ความตรงประเด็น (Relevance)	4.28	1.07	มาก
ความสมบูรณ์ของข้อมูล (Completeness)	3.74	0.87	มาก

ตารางที่ 2 ผลการประเมินคุณภาพด้วย AI Judge

5.2 การประเมินคุณภาพการนำเสนอด้วยกฎเชิงโครงสร้าง (Rule-Based Evaluation)

5.2.1 หลักการประเมิน

การประเมินในส่วนนี้ใช้แนวคิด Heuristic-based Evaluation ร่วมกับหลักการ UX Writing เพื่อตรวจสอบคุณภาพการนำเสนอของข้อความโดยไม่พิจารณาความถูกต้องของเนื้อหาเชิงลึก แต่จะเน้นที่รูปแบบและประสบการณ์การอ่าน เป็นหลักการประเมินแบ่งออกเป็น 3 มิติ ได้แก่

1. ความง่ายในการอ่าน (Response Legibility) หมายถึง ความชัดเจนของรูปแบบข้อความ การเว้นวรรคและการแบ่งย่อหน้า เพื่อให้ผู้อ่านสามารถเข้าใจเนื้อหาได้อย่างรวดเร็ว
2. คุณภาพเชิงโครงสร้างของคำตอบ (Structural Quality) หมายถึง ความเหมาะสมของรูปแบบการนำเสนอ เช่น การใช้ตารางหรือรายการ เพื่อช่วยจัดระเบียบข้อมูลให้เข้าใจง่าย
3. ความรู้สึกและโทนอารมณ์ (Sentiment) หมายถึง ระดับความสุภาพและ ความเป็นมิตรของคำตอบ เพื่อให้การสื่อสารมีความเป็นมืออาชีพและสร้างความประทับใจแก่ผู้ใช้

5.2.2 สรุปผลการประเมินคุณภาพการนำเสนอด้วย Rule-Based Evaluation

จากการประเมินบทสนทนาจำนวน 50 ชุด ได้ผลคะแนนเฉลี่ยดังนี้

มิติการประเมิน	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความหมาย
ความง่ายในการอ่าน (Response Legibility)	4.98	0.14	มากที่สุด
คุณภาพของคำตอบ (Structural Quality)	4.25	1.06	มาก
ความรู้สึกและโทนอารมณ์ (Sentiment)	4.35	1.11	มาก

ตารางที่ 3 ผลการประเมินคุณภาพ Rule-Based Evaluation

6.อภิปราย

จากผลการศึกษา ระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะที่พัฒนาในรูปแบบ AI Agent ร่วมกับ Retrieval-Augmented Generation (RAG) สามารถให้คำตอบที่ถูกต้องและเหมาะสมกับบริบทงานบริการลูกค้า โดยตัวอย่างการใช้งานจริงแสดงให้เห็นว่าระบบสามารถดึงข้อมูลโปรโมชั่น เช่น ราคา ความเร็ว และเงื่อนไข มาใช้ในการตอบคำถามได้อย่างเป็นรูปธรรม และจัดรูปแบบคำตอบในลักษณะโครงสร้าง เช่น ตาราง ซึ่งช่วยให้ผู้ใช้เข้าใจและเปรียบเทียบข้อมูลได้ง่ายขึ้น เมื่อเปรียบเทียบกับระบบแชตบอทแบบเดิมที่ให้บริการผ่าน LINE ซึ่งแสดงผลเป็นข้อความต่อเนื่อง ผู้ใช้ต้องตีความข้อมูลด้วยตนเอง พบว่าระบบที่ใช้ RAG ช่วยลดความซับซ้อนในการอ่านและเพิ่มความชัดเจนของข้อมูล สอดคล้องกับผลการประเมินด้านความง่ายในการอ่าน (4.98) และคุณภาพเชิงโครงสร้าง (4.25) ที่อยู่ในระดับสูง ในด้านคุณภาพเนื้อหา ระบบมีคะแนนความสุภาพ (4.60) และความตรงประเด็น (4.28) อยู่ในระดับดีถึงดีมาก สะท้อนความสามารถในการเข้าใจเจตนาผู้ใช้และเลือกใช้ภาษาที่เหมาะสม อย่างไรก็ตาม คะแนนความครบถ้วนของข้อมูล (3.74) ที่ต่ำกว่าในมิติอื่น แสดงถึงข้อจำกัดในการสังเคราะห์ข้อมูลที่มีหลายเงื่อนไข เช่น การเปรียบเทียบหลายแพ็กเกจพร้อมกัน ทั้งนี้ งานวิจัยยังไม่มีการทดลองเชิงเปรียบเทียบกับระบบที่ไม่ใช้ RAG อย่างเป็นรูปธรรม โดยการเปรียบเทียบในงานนี้เป็นเชิงคุณภาพจากลักษณะของระบบและตัวอย่างการใช้งานจริง ซึ่งถือเป็นข้อจำกัด และควรได้รับการศึกษาเพิ่มเติมในอนาคต

7.สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะด้วย RAG และ AI Agent เพื่อให้ผู้ใช้สามารถเข้าถึงข้อมูลโปรโมชั่นได้อย่างถูกต้อง ชัดเจน และเหมาะสมกับการใช้งานจริง ผลการประเมินจากบทสนทนา 50 ชุด แสดงให้เห็นว่าระบบมีประสิทธิภาพในระดับดีถึงดีมาก โดยมีคะแนนความสุภาพ (4.60) ความตรงประเด็น (4.28) และความง่ายในการอ่าน (4.98) ซึ่งสะท้อนถึงความสามารถในการสื่อสารข้อมูลได้อย่างมีประสิทธิภาพและเป็นมิตรต่อผู้ใช้ นอกจากนี้ ตัวอย่างการใช้งานจริงแสดงให้เห็นว่าระบบสามารถนำข้อมูลจากคลังความรู้มาจัดรูปแบบเป็นคำตอบเชิงโครงสร้าง เช่น ตาราง ช่วยให้ผู้ใช้เข้าใจข้อมูลและเปรียบเทียบโปรโมชั่นได้ง่ายกว่าระบบแชตบอทแบบเดิม

8. ข้อเสนอแนะ

ระบบแนะนำโปรโมชั่นอินเทอร์เน็ตอัจฉริยะด้วย RAG และ AI Agent สามารถทำงานได้ตามวัตถุประสงค์ แต่ยังสามารถพัฒนาเพิ่มเติมเพื่อเพิ่มความครบถ้วนและความแม่นยำของคำตอบได้ ในอนาคตควรมีการเพิ่มการประเมินผลเชิงปริมาณ เช่น ความถูกต้องของคำตอบและความพึงพอใจของผู้ใช้ เพื่อใช้เปรียบเทียบประสิทธิภาพระหว่างระบบเก่าและระบบใหม่ ระบบนี้ยังสามารถต่อยอดไปใช้ในงาน

ในด้านอื่นๆ เช่น ระบบแนะนำสินค้า หรือระบบถาม–ตอบในองค์กร โดยปรับคลังความรู้ให้เหมาะสมกับบริบทการใช้งาน อย่างไรก็ตาม งานวิจัยนี้ยังมีข้อจำกัดด้านความครบถ้วนของข้อมูลในกรณีที่คำถามมีความซับซ้อน ในอนาคตควรพัฒนาคลังความรู้ให้มีความครอบคลุมมากขึ้น และดำเนินการทดลองเชิงเปรียบเทียบเพื่อยืนยันประสิทธิภาพของระบบในระดับที่ชัดเจนยิ่งขึ้น

9. กิตติกรรมประกาศ

บทความนี้สำเร็จลุล่วงได้ด้วยดี เนื่องจากได้รับความกรุณาอย่างสูงจาก ผู้ช่วยศาสตราจารย์ ก่องกาญจน์ ดุลยไชย์ ผู้ช่วยศาสตราจารย์ ดร.พาสน์ ปราโมกข์ชน และ ผู้ช่วยศาสตราจารย์ ดร. สมนึก สินธุปวน ที่ให้คำแนะนำและข้อคิดเห็นต่าง ๆ อันเป็นประโยชน์อย่างยิ่งต่องานวิจัย ขอขอบคุณบริษัท โทรคมนาคมแห่งชาติ จำกัด (มหาชน) ที่ให้โอกาสและการสนับสนุนรวมถึงการให้คำปรึกษา ขอขอบคุณบุคลากรสาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้ ที่ให้ความรู้และคำปรึกษาด้านการจัดทำเอกสาร รวมถึงการสนับสนุนตลอดจนการปรับปรุง แก้ไขข้อบกพร่องต่าง ๆ จนทำให้งานวิจัยครั้งนี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

10. เอกสารอ้างอิง

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *Artificial Intelligence Applications and Innovations*. https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31

Anthropic Team. (2025). *Building AI agents with tool calling*. https://www.anthropic.com/index/tool-use-in-ai-agents

Auth0 Team. (2026). *Understanding OAuth 2.0 and OpenID Connect*. https://auth0.com/intro-to-iam/what-is-oauth-2

Dataherald. (2024). *Building intelligent agents with RAG and FastAPI architecture*. https://medium.com/@dataherald/building-ai-agents-with-fastapi-and-rag-architecture-2024

Ganesh, et al. (2024). Unified recommendation system with chatbot support. https://thesai.org/Downloads/Volume15No3/Paper_79-Retrieval_Augmented_Generation_Approach.pdf

Google AI. (2026). *Gemini Developer API*. https://ai.google.dev/gemini-api/docs

Hugging Face Team. (2024). *Retrieval-augmented generation (RAG): Combining LLMs with your data*. https://huggingface.co/blog/rag

Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open-domain question answering. *EACL 2021*. https://arxiv.org/abs/2007.01282

Lewis, P. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *NeurIPS 2020*. https://arxiv.org/abs/2005.11401

Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. https://aclanthology.org/W04-1013.pdf

Luo, X. (2019). Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science, 38*(6), 937–947. https://pubsonline.informs.org/doi/10.1287/mksc.2019.1192

Madaan, et al. (2024). Retrieval-augmented generation for document QA. https://www.ijcrt.org/papers/IJCRT24A5161.pdf

Open-WebUI Team. (2026). *Open-WebUI: A web-based UI for open-source large language models*. https://github.com/open-webui/open-webui

Papineni, K. (2002). BLEU: A method for automatic evaluation of machine translation. *ACL 2002*. https://aclanthology.org/P02-1040.pdf

PostgreSQL Global Development Group. (2026). *Introduction to PostgreSQL*. https://www.postgresql.org/docs/current/index.html

Thomas, et al. (2022). Enhancing customer satisfaction with chatbots. https://pmc.ncbi.nlm.nih.gov/articles/PMC9355322/

Tiangolo. (2026). *FastAPI — The modern web framework for building APIs with Python*. https://fastapi.tiangolo.com/

Twilio Team. (2026). *Implementing two-factor authentication (2FA) in web applications*. https://www.twilio.com/docs/verify/authentication

vLLM Team. (2026). *vLLM: Easy, fast, and cheap LLM serving*. https://vllm.ai/

Zhang, T. (2020). BERTScore: Evaluating text generation with BERT. *ICLR 2020*. https://arxiv.org/abs/1904.09675

Zheng, L. (2023). Judging LLM-as-a-judge with MT-bench and Chatbot Arena. *NeurIPS 2023*. https://arxiv.org/abs/2306.05685