

Received: November 28, 2024; Revised: March 17, 2025; Accepted: March 24, 2025

การประเมินและเปรียบเทียบประสิทธิภาพการบริการตอบคำถามลูกค้า
ที่สร้างขึ้นจากปัญญาประดิษฐ์เชิงสร้างสรรค์
Evaluation and comparison of customer service performance
generated by Generative AI

ณัฐภา ฬิวมา¹ และธรรณพ อารีพรรค^{2*}
Nattha Phiwma¹ and Thannob Aribarg^{2*}

¹คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต กรุงเทพมหานคร

¹Faculty of Science and Technology Suan Dusit University Bangkok, Bangkok

²วิทยาลัยนวัตกรรมการดิจิทัลเทคโนโลยี มหาวิทยาลัยรังสิต ปทุมธานี

²College of Digital Innovation Technology Rangsit University Pathum Thani, Pathum Thani Province

*Corresponding Author E-mail Address : thannob.a@rsu.ac.th

บทคัดย่อ

ปัญญาประดิษฐ์เชิงสร้างสรรค์มีความสามารถในการโต้ตอบแบบสนทนาและเข้าใจภาษาธรรมชาติซึ่งมีหลากหลายโปรแกรม หากใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ที่มีประสิทธิภาพการตอบคำถามได้ใกล้เคียงกับมนุษย์ก็ยังสามารถสร้างความประทับใจให้กับผู้ใช้ได้และเป็นประโยชน์ต่อการใช้ในการบริการตอบคำถามลูกค้า ดังนั้นงานวิจัยนี้มีวัตถุประสงค์เพื่อประเมินและเปรียบเทียบประสิทธิภาพการบริการตอบคำถามลูกค้าที่สร้างขึ้นจาก Generative AI ได้แก่ แชทจีพีที คลอดด์ และ เจมีไน ขั้นตอนการดำเนินการวิจัย คือ เตรียมข้อมูลคำถามและคำตอบที่ถูกต้อง สร้างคำตอบโดย Generative AI คือ แชทจีพีที คลอดด์ และ เจมีไน ทำการประเมินคำตอบด้วยการใช้อัลกอริทึม ROUGE-L และผู้เชี่ยวชาญ เพื่อเปรียบเทียบประสิทธิภาพในการตอบคำถาม โดย แชทจีพีที คลอดด์ และ เจมีไน ทำการตอบคำถาม 2 รูปแบบ ได้แก่ รูปแบบปกติและรูปแบบบทบาทสมมติตอบคำถามรูปแบบละ 3 ครั้ง ผลการวิจัย พบว่า 1) ผลการประเมินการตอบคำถามด้วย Generative AI โดยใช้อัลกอริทึม ROUGE-L รูปแบบปกติ ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ แชทจีพีที มีค่าเฉลี่ยเท่ากับ 0.26547 ผลการประเมินการตอบคำถามด้วย Generative AI และรูปแบบบทบาทสมมติ ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ แชทจีพีที มีค่าเฉลี่ยเท่ากับ 0.29307 2) ผลการประเมินการตอบคำถามด้วย Generative AI โดยผู้เชี่ยวชาญในรูปแบบปกติ ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ แชทจีพีทีมีค่าเฉลี่ยเท่ากับ 4.1556 และผลการประเมินการตอบคำถามด้วย Generative AI รูปแบบบทบาทสมมติ ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ แชทจีพีที มีค่าเฉลี่ยเท่ากับ 4.2867

คำสำคัญ: ปัญญาประดิษฐ์เชิงสร้างสรรค์ ปัญญาประดิษฐ์ แชทจีพีที การบริการตอบคำถาม โมเดลภาษาขนาดใหญ่

Abstract

Generative AI has the ability to interact in a conversational manner and understand natural language, where there are a variety of programs available. If Generative AI is high-performance with the ability to answer questions similar to humans, it will impress users and be useful for customer service applications. Therefore, the objective of this research was to evaluate and compare the effectiveness of customer

questioning-answering services generated by Generative AI models, namely ChatGPT, Claude, and Gemini. The research procedure involved preparing a dataset of questions and accurate answers, generating responses using Generative AI models (ChatGPT, Claude, and Gemini), and assessing these responses using the ROUGE-L algorithm and the evaluation of experts to compare the efficiency of each model's answers. ChatGPT, Claude, and Gemini were used to answer questions in two formats: the standard prompt format and a roleplay prompt format, with each question answered three times per format. The results of this study were as follows: 1) According to the standard format evaluation using the ROUGE-L algorithm, the answers generated by ChatGPT were most similar to the correct answers, with an average score of 0.26547. In the roleplay format, ChatGPT also produced answers most similar to the correct ones, with an average score of 0.29307; 2) Regarding the expert evaluations of the standard format responses, ChatGPT's answers were rated the closest to the correct ones, with an average score of 4.1556. In the roleplay prompt format, ChatGPT again scored the highest similarity, with an average of 4.2867.

Keywords: Generative artificial intelligence, Artificial intelligence, ChatGPT, Customer service, Large language models

บทนำ

แนวโน้มการใช้ปัญญาประดิษฐ์ (Artificial Intelligence; AI) ในการทำงานมีอย่างต่อเนื่องโดยมีการผสมรวมปัญญาประดิษฐ์เข้ากับกระบวนการทำงาน มีการใช้ประโยชน์จาก AI ในรูปแบบต่าง ๆ เช่น รูปแบบแชทบอทเสมือนเป็นผู้ช่วยส่วนตัว แม้ว่าแชทบอทจะถูกนำไปใช้งานเป็นผู้ช่วยส่วนตัวเพื่อวัตถุประสงค์ต่าง ๆ แต่ก็มีบทบาทสำคัญในการให้ข้อมูลทั้งในด้านการศึกษา การวิจัย การตลาด และการให้บริการผ่านการตอบคำถาม ผ่านการโต้ตอบของผู้ใช้ ในช่วงไม่กี่ปีที่ผ่านมาการใช้งานแชทบอทเพิ่มมากขึ้น กลุ่มแชทบอทที่น่าสนใจที่ใช้อยู่ในปัจจุบัน ได้แก่ แชทบอทแบบโมเดลภาษาขนาดใหญ่ (Large Language Models; LLMs) ที่ได้รับการฝึกอบรมจากข้อมูลจำนวนมาก เป็นโมเดลการเรียนรู้ของเครื่องที่สามารถทำความเข้าใจและสร้างข้อความภาษาของมนุษย์ได้ ปัญญาประดิษฐ์เชิงสร้างสรรค์ (Generative AI) และปัญญาประดิษฐ์เชิงสนทนา (Conversational AI) ซึ่งเป็นรูปแบบของการนำเทคโนโลยี AI มาใช้งานที่ช่วยอำนวยความสะดวกให้คนกับคอมพิวเตอร์สามารถพูดคุยกันได้ (Lins dos Santos, 2023)

ปัจจุบันมีการใช้ Generative AI ซึ่งเป็นเทคโนโลยีที่ครอบคลุมโมเดลภาษาที่กว้างขวางและมีบทบาทอย่างยิ่งในการทำงาน Generative AI เป็นเทคโนโลยีที่มีความสามารถในการสร้างเนื้อหาหลากหลายประเภท ช่วยเพิ่มประสิทธิภาพและลดภาระงานของมนุษย์ได้อย่างมีนัยสำคัญ ในด้านการสร้างข้อความ เทคโนโลยีนี้สามารถสร้างเนื้อหาที่สอดคล้องกับบริบทต่าง ๆ นอกจากนี้ AI ยังสามารถสร้างภาพและวิดีโอที่มีความสมจริง ส่วนด้านเสียงและดนตรี AI สามารถสังเคราะห์เสียงธรรมชาติและสร้างเพลงที่เลียนแบบสไตล์มีโอชิฟได้อย่างแม่นยำ ในการพัฒนาซอฟต์แวร์ AI ช่วยสร้างและแก้ไขโค้ด อีกทั้ง AI ยังมีบทบาทในการสร้างงานออกแบบศิลปะและยังช่วยสร้างข้อมูลสังเคราะห์เพื่อใช้ในการวิจัย เทคโนโลยีนี้จึงกลายเป็นเครื่องมือสำคัญที่สนับสนุนการทำงานและสร้างเนื้อหาในหลากหลายสาขา ช่วยให้ผู้เชี่ยวชาญสามารถมุ่งเน้นไปที่งานที่ต้องใช้ความคิดสร้างสรรค์มากขึ้น (Stryker and Scapicchio, 2024) นอกจากนี้การใช้ Generative AI ซึ่งเป็นโมเดลภาษาขนาดใหญ่ที่มาพร้อมกับทักษะที่หลากหลายสามารถสร้างเนื้อหาใหม่เป็นผลลัพธ์ เนื้อหาใหม่นี้อาจมีลักษณะเป็นข้อความ รูปภาพ และเสียงคุณภาพสูงตามโมเดลภาษาขนาดใหญ่ที่ได้รับการฝึกอบรมมา อินเทอร์เน็ตแชทบอทกับ AI เชิงสร้างสรรค์สามารถจดจำ สรุป แปล คัดการณ์ และสร้างเนื้อหาเพื่อตอบสนองต่อคำถามของผู้ใช้โดยไม่ต้องมีปฏิสัมพันธ์กับมนุษย์ (IBM, 2023) ตัวอย่าง Generative AI ประเภท แชทบอท AI ที่ใช้ในปัจจุบัน เช่น ChatGPT พัฒนาโดย OpenAI มีฐานข้อมูลที่ใช้ในการตอบคำถามปรับปรุงล่าสุดเดือนตุลาคม ปี 2023 ส่วน Claude ที่พัฒนาโดย Anthropic ฐานข้อมูลปรับปรุงล่าสุดเดือนเมษายน 2024 และ Gemini ที่พัฒนาขึ้นโดย Google AI สามารถเชื่อมต่อกับแหล่งข้อมูลภายนอกที่เป็นแบบเรียลไทม์ได้ เช่น Google search เพื่อดึงข้อมูลล่าสุดมาใช้ในการตอบคำถาม

แชทบอท AI ใช้เทคโนโลยีความเข้าใจภาษาธรรมชาติ (Natural Language Processing; NLP) เพื่อเข้าใจความหมายของข้อความที่ผู้ใช้ป้อนเข้าไปและใช้ AI สร้างการตอบกลับที่เหมาะสม เทคโนโลยีเหล่านี้ใช้ทั้งการเรียนรู้ของเครื่อง (Machine learning) การเรียนรู้เชิงลึก (Deep Learning) เพื่อพัฒนาฐานความรู้จากการโต้ตอบกับผู้ใช้ และ Transformer

Models เช่น GPT และ BERT ทำให้สามารถสร้างบทสนทนาเชิงลึกได้ใกล้เคียงมนุษย์มากขึ้น เทคโนโลยีนี้จึงมีบทบาทสำคัญในธุรกิจและชีวิตประจำวัน เนื่องจากการเพิ่มประสิทธิภาพการทำงาน ลดภาระงานพนักงาน ช่วยในการทำงานซ้ำ ๆ ให้บริการที่เป็นส่วนตัวและตรงใจลูกค้ามากขึ้น เข้าถึงกลุ่มลูกค้าได้กว้างขึ้นสามารถให้บริการลูกค้าได้หลายภาษาและหลายช่องทาง ลูกค้าสามารถติดต่อได้ทุกที่ทุกเวลา บริการลูกค้าตอบคำถามและแก้ปัญหาได้รวดเร็วตลอด 24 ชั่วโมง เช่น ChatGPT, Claude หรือ Gemini เป็นต้น ซึ่งคุณลักษณะทางอารมณ์จะช่วยเพิ่มประสิทธิภาพของบริการแชทบอท รวมทั้งการโต้ตอบ ความเห็นอกเห็นใจจะช่วยเพิ่มความพึงพอใจของลูกค้า การใช้แชทบอทต้องปรับปรุงการตอบสนองเพื่อเสริมสร้างความสัมพันธ์กับลูกค้า แชทบอทที่มีค่าแสดงถึงอารมณ์จะช่วยเพิ่มประสิทธิภาพในการสื่อสารกับลูกค้า ความน่าเชื่อถือและความมั่นใจส่งผลในเชิงบวกต่อความพึงพอใจของลูกค้า (Yun and Park, 2022) การประยุกต์ใช้ Generative AI ในปัจจุบันเข้ามามีบทบาทในการทำงาน เช่น ด้านการแพทย์ ด้านการศึกษา ภาคธุรกิจและการตลาด ช่วยวิเคราะห์พฤติกรรมผู้บริโภค สร้างเนื้อหาการตลาด และให้บริการลูกค้าผ่านระบบอัตโนมัติ ด้านความบันเทิงและงานสร้างสรรค์ เปิดมิติใหม่ในการสร้างภาพ ดนตรี วิดีโอ และงานเขียน ด้านการเงิน ช่วยวิเคราะห์ความเสี่ยง ตรวจสอบการฉ้อโกง และให้คำแนะนำด้านการลงทุน อย่างไรก็ตาม ยังมีความท้าทายด้านจริยธรรม ความเป็นส่วนตัว และความปลอดภัยของข้อมูลที่ต้องคำนึงถึง ทั้งนี้ Generative AI ถือเป็นนวัตกรรมสำคัญที่จะช่วยยกระดับคุณภาพชีวิตและขับเคลื่อนเศรษฐกิจในยุคดิจิทัล (McKinsey and Company, 2023; MIT Technology Review, 2023; World Economic Forum, 2023; Gartner, 2024)

ในยุคดิจิทัลปัจจุบันการใช้เทคโนโลยีแชทบอทได้รับความนิยมมากขึ้นในกลุ่มธุรกิจขนาดย่อม ขนาดกลาง และขนาดเล็ก (MSME) เพื่อปรับปรุงการให้บริการลูกค้า ผลการศึกษาแสดงให้เห็นว่าการสำรวจศักยภาพของ ChatGPT ในการพัฒนาบริการลูกค้าใน MSMEs นั้นมีความหวังและมีประโยชน์มากมาย ในแวดวงขององค์กร MSME คุณภาพของการบริการลูกค้าเป็นสิ่งที่ต้องให้ความสำคัญสูงสุด ChatGPT ช่วยให้ MSMEs เพิ่มประสิทธิภาพและคุณภาพในการให้บริการลูกค้า ส่งผลให้ประสบการณ์โดยรวมของลูกค้าดีขึ้น (Subagja et al., 2023) อิทธิพลของเทคโนโลยีแชทบอทที่ขับเคลื่อนด้วย AI เช่น ChatGPT เป็นตัวแทนการสนทนาอัตโนมัติเหล่านี้มีผลกระทบเชิงบวกต่อการตลาดดิจิทัลและการดำเนินงานอีคอมเมิร์ซ ความสามารถในการประมวลผลคำถามของลูกค้าอย่างรวดเร็วและสร้างคำตอบที่ปรับให้เป็นรายบุคคลตามข้อมูลที่ได้รับจากลูกค้าทำให้ธุรกิจสามารถสร้างความสัมพันธ์ที่แข็งแกร่งขึ้นกับลูกค้าได้ ขณะเดียวกันก็ส่งเสริมการเพิ่มการเปลี่ยนแปลงยอดขายในระยะยาว แชทบอทยังช่วยปรับปรุงการดูแลลูกค้าและมีความพึงพอใจอย่างมีนัยสำคัญ ซึ่งระบบอัตโนมัติช่วยลดต้นทุนและเพิ่มประสิทธิภาพในการดำเนินงาน การตอบสนองที่เร็วขึ้นนำไปสู่ความภักดีของลูกค้าที่เพิ่มขึ้น (Chowdhury, 2023) จากงานวิจัยของ นันทน์ภัส (2560) ที่ได้ทำการศึกษาแนวทางการปรับใช้แชทบอทสำหรับงานบริการลูกค้า ผลการวิจัย พบว่า ความสำคัญของแชทบอทไม่ได้อยู่เพียงแค่ประโยชน์การใช้งานและสามารถช่วยสนับสนุนการทำงานของบุคลากรในธุรกิจได้เพียงเท่านั้น แต่ยังรวมถึงการที่แชทบอทสามารถทำให้ธุรกิจสามารถได้รับข้อมูลของผู้ใช้อีกด้วย เหตุผลหลักที่ทำให้ธุรกิจปรับใช้แชทบอทสำหรับงานบริการลูกค้ามาจากการที่ธุรกิจไม่มีบุคลากรที่ทำหน้าที่ตอบลูกค้าโดยเฉพาะธุรกิจไม่สามารถตอบกลับลูกค้าได้ตลอด 24 ชั่วโมง หรือผู้ดูแลของธุรกิจไม่สามารถตอบคำถามได้อย่างทั่วถึงและรวดเร็วจากการที่มีข้อความของลูกค้าที่ส่งเข้ามาจำนวนมากในแต่ละวัน Generative AI ช่วยเพิ่มปฏิสัมพันธ์กับลูกค้านอกเหนือจากการแก้ปัญหาอย่างง่าย และช่วยเพิ่มประสิทธิภาพในการดำเนินงานบริการลูกค้า และงานวิจัยของสมพงษ์ และรชพร (2566) ได้ทำการวิจัยเรื่องประสิทธิภาพระบบตอบรับอัตโนมัติที่ส่งผลต่อคุณภาพบริการด้านการสื่อสารของธุรกิจสายการบินที่ทำการบินภายในประเทศ ได้ทำการวิเคราะห์ประสิทธิภาพระบบตอบรับอัตโนมัติที่ส่งผลต่อคุณภาพการบริการด้านการสื่อสารของธุรกิจสายการบินที่ทำการบินในประเทศไทย ผลการวิจัย พบว่า ประสิทธิภาพระบบตอบรับอัตโนมัติส่งผลต่อคุณภาพการบริการด้านการสื่อสารของธุรกิจสายการบินที่ทำการบินภายในประเทศไทยโดยส่งผลร้อยละ 68 และมีระดับนัยสำคัญทางสถิติที่ระดับ 0.05 นอกจากนี้งานวิจัยของสุทธิภาส และวศิน (2563) ได้ทำการวิจัยเรื่องการประเมินประสิทธิภาพ ประสิทธิผล และความพึงพอใจในการใช้โปรแกรมแชทบอทธนาคารในประเทศไทย มีวัตถุประสงค์เพื่อประเมินประสิทธิภาพ ประสิทธิผล พฤติกรรมการใช้โปรแกรมแชทบอทและพัฒนาตัวแบบอิทธิพลของตัวแปรที่เกี่ยวข้องที่ส่งผลต่อความพึงพอใจในการใช้โปรแกรมแชทบอทของธนาคารผ่านมุมมองผู้ใช้โปรแกรม ผลการทดสอบสมมติฐาน พบว่า ประสิทธิภาพของแชทบอทมีผลต่อระดับการมีส่วนร่วมของลูกค้าธนาคารในการใช้แชทบอท ($R^2=.33$) ประสิทธิภาพของผลโปรแกรมแชทบอทมีผลต่อระดับความรู้สึกเป็นมิตรในขณะทำธุรกรรม ($R^2=.431$) และพฤติกรรมการใช้แชทบอทมีผลต่อระดับการรับรู้ทางอารมณ์ขณะใช้งาน ($R^2=.326$) การใช้แชทบอทช่วยเพิ่มประสิทธิภาพการบริการลูกค้าและการมีส่วนร่วมมีการลงทุนน้อยกว่าตัวแทนมนุษย์แต่มีข้อจำกัดในการจัดการสถานการณ์ทางอารมณ์ที่ซับซ้อน เพิ่มการมีส่วนร่วมเมื่อเทียบกับการตลาดทางอีเมลแบบดั้งเดิม แชทบอทสามารถจัดการกับลูกค้าหลายคนพร้อมกัน (Boruah, 2024) การประมวลผลภาษาธรรมชาติ (NLP) มีความสำคัญสำหรับการใช้งาน AI แบบสร้างสรรค์

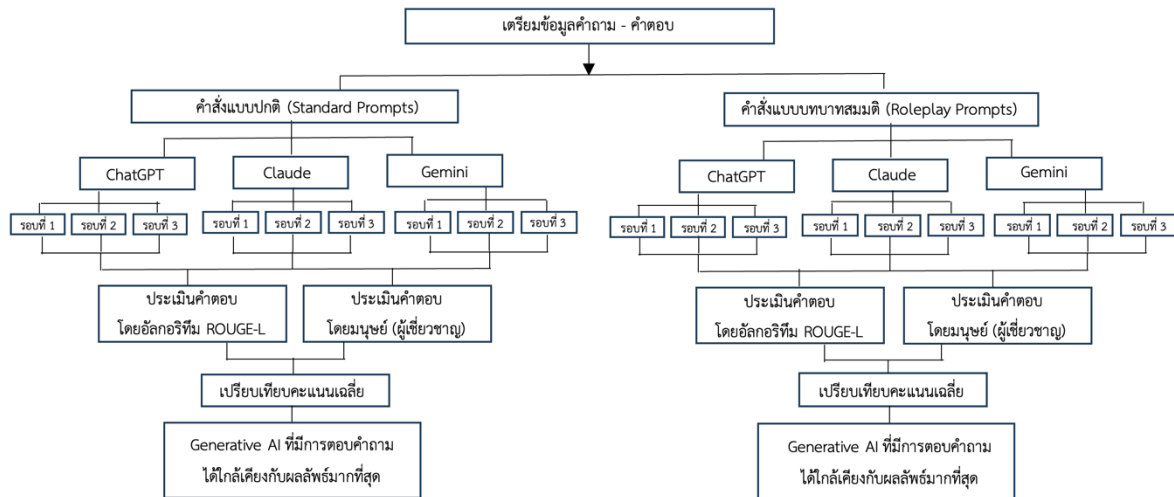
สามารถเปลี่ยนแปลงประสบการณ์การบริการลูกค้าได้อย่างมีนัยสำคัญ (ศรุต, 2567) แต่ยังมีข้อจำกัดสำคัญในบริบทการบริการลูกค้า ได้แก่ ความยากในการเข้าใจบริบทที่ซับซ้อนและเชื่อมโยงข้อมูลที่หลากหลาย โดยเฉพาะกรณีที่ต้องอ้างอิงเหตุการณ์ล่าสุดหรือข้อมูลเฉพาะทาง อีกทั้งยังมีข้อจำกัดด้านความแม่นยำ เนื่องจาก AI สร้างข้อมูลใหม่โดยไม่มีการตรวจสอบความถูกต้อง และการที่ AI ไม่สามารถอัปเดตข้อมูลได้ทันทีหลังการฝึกอบรม ส่งผลให้ข้อมูลบางส่วนล้าสมัย ขาดความละเอียดอ่อนในการตอบคำถามซึ่งอาจนำไปสู่ข้อความที่ไม่เหมาะสมในบางบริบท ดังนั้นการใช้งาน Generative AI ควรทำควบคู่กับการตรวจสอบและความเชี่ยวชาญของมนุษย์เพื่อให้ได้ผลลัพธ์ที่ถูกต้องและเหมาะสมที่สุด

จากข้อมูลที่ทำการศึกษาจะเห็นได้ว่าการใช้แชทบอทในงานด้านการบริการตอบคำถามให้กับลูกค้ามีประโยชน์ในการใช้งานบริการ นอกจากนี้มีการนำมาใช้โมเดลภาษาขนาดใหญ่ร่วมกับวิธีการผสานระบบดึงข้อมูลสำหรับการตอบคำถามทางการศึกษาโดยมีผลลัพธ์คะแนน ROGUE สูงขึ้น 67.8% เมื่อเทียบกับ ChatGPT แบบมาตรฐาน (Wu et al., 2023) จากงานวิจัยของ Poonprasaeetporn et al. (2023) พบว่า การส่งคำสั่งแบบบทบาทสมมติ (Roleplay prompt) กับ ChatGPT มีความถูกต้องของคำตอบด้านสมุนไพรอยู่ที่ 86.67% มากกว่าแบบไร้ตัวอย่างอ้างอิง (Zero-shot prompt) ที่ 77.48% ในการศึกษา 6 แบบการใช้คำสั่งสำหรับแบบจำลอง GPT-4 สำหรับคำอธิบายเกี่ยวกับปรากฏการณ์ทางฟิสิกส์และเคมี พบว่า Chain of Thought prompts (CoT prompts) รวมกับโจทย์และเกณฑ์การให้คะแนน จะช่วยเพิ่มความแม่นยำของผลลัพธ์ได้อย่างมีนัยสำคัญโดยเพิ่มขึ้น 13.44% และ 3.7% เมื่อเทียบกับ Zero-shot prompts และ Few-shot prompts (Lee, G.-G. et.al., 2024) ดังนั้นงานวิจัยนี้จึงมุ่งเน้นไปที่ Generative AI สำหรับการโต้ตอบแบบสนทนาและเข้าใจภาษาธรรมชาติ ที่สร้างความประทับใจคล้ายคลึงกับการสนทนาของมนุษย์ สำหรับช่วยตอบคำถามที่พบบ่อยในอุตสาหกรรมบริการ โดยทำการเปรียบเทียบประสิทธิภาพการบริการตอบคำถามลูกค้าโดยใช้ Generative AI โดยมีวัตถุประสงค์เพื่อประเมินและเปรียบเทียบประสิทธิภาพการบริการตอบคำถามลูกค้าที่สร้างขึ้นจาก ChatGPT, Claude และ Gemini ซึ่งเป็นแชทบอทที่ได้รับความนิยมในปัจจุบัน อีกทั้งการวิจัยนี้มุ่งประเมินและเปรียบเทียบประสิทธิภาพรูปแบบการคำสั่งแบบปกติกับคำสั่งแบบบทบาทสมมติเพื่อศึกษารูปแบบการคำสั่งที่เหมาะสมสำหรับการบริการตอบคำถามให้กับลูกค้า เพื่อให้ นักพัฒนาสามารถนำจากการวิจัยนี้ไปต่อยอดในการพัฒนาแชทบอทเพื่อคัดเลือกแบบจำลองจาก Generative AI ที่เหมาะสมกับรูปแบบการคำสั่งสำหรับการตอบคำถามเพื่อบริการให้กับลูกค้าได้อย่างมีประสิทธิภาพ

วัสดุ อุปกรณ์ และวิธีการ

วิธีการดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยแบบเชิงปริมาณ โดยทำประเมินการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งคำสั่งแบบปกติ (Standard prompt) เป็นการสั่งงานโดยตรง โดยระบุสิ่งที่ต้องการให้ Generative AI ทำอย่างชัดเจน และคำสั่งแบบบทบาทสมมติ (Roleplay prompt) เป็นการกำหนดบทบาทหรือหน้าที่ให้กับ Generative AI เพื่อให้ตอบสนองตามบริบทที่กำหนด การใช้คำสั่งแบบปกติและคำสั่งแบบบทบาทสมมติเพื่อเปรียบเทียบประสิทธิภาพในการตอบคำถามว่าการใช้ Prompt รูปแบบใดจะมีความเหมาะสมและใกล้เคียงกับคำตอบต้นฉบับ โดยใช้อัลกอริทึม ROUGE-L และมนุษย์



รูปที่ 1 ขั้นตอนการดำเนินการประเมินการตอบคำถามด้วย Generative AI

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ ได้แก่ ผู้เชี่ยวชาญจำนวน 5 ท่าน โดยวิธีการสุ่มแบบเจาะจง (Purposive sampling) เป็นการเลือกกลุ่มตัวอย่างเพื่อความเหมาะสมในการวิจัยและเป็นผู้ที่มีประสบการณ์ในการสอบถามเกี่ยวกับข้อมูลโรงแรมผ่านเว็บไซต์ที่มีการตอบกลับแบบอัตโนมัติ

เครื่องมือที่ใช้ในการวิจัย

1. คำถามและคำตอบของโรงแรมแห่งหนึ่งซึ่งเป็นแหล่งข้อมูลที่เผยแพร่สู่สาธารณะ ชุดข้อมูลนี้ประกอบด้วยคำถามที่พบบ่อย (FAQ) จำนวน 50 ข้อ พร้อมคำตอบที่จัดเตรียมไว้จากเว็บไซต์ของโรงแรม ใช้สำหรับการประเมินและเปรียบเทียบประสิทธิภาพการตอบคำถามระหว่าง Generative AI และผู้เชี่ยวชาญกับคำตอบต้นฉบับจากเว็บไซต์ของโรงแรม ตัวอย่างคำถามและคำตอบที่ใช้ในการทดลอง เช่น

คำถาม โรงแรมมีสระว่ายน้ำของโรงแรมหรือไม่

คำตอบ โรงแรมมีสระว่ายน้ำตั้งอยู่ที่ชั้น 5 และชั้น 17

คำถาม โรงแรมรับชำระเงินแบบใดบ้าง?

คำตอบ โรงแรมรับชำระเงินด้วยบัตรเครดิตและเงินสดเท่านั้น

2. แบบประเมินการให้คะแนนการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) ด้วยมนุษย์ (ผู้เชี่ยวชาญ)
3. Generative AI โมเดลที่ใช้ในการสร้างคำตอบ ได้แก่ ChatGPT, Claude และ Gemini
4. Python Programming สำหรับการคำนวณ ROUGE-L Score

การเก็บรวบรวมข้อมูล

1. ทำการประเมินประสิทธิภาพการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT (GPT4o-mini), Claude (3.5 Sonnet) และ Gemini (1.5 Flash) ทำการประเมิน 2 รูปแบบ แบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) โดยใช้อัลกอริทึม ROUGE-L

1.1 การเตรียมข้อมูลและกำหนดคำถาม

1) สร้างชุดคำถามโดยใช้ชุดคำถามที่พบบ่อยจากจากโรงแรมแห่งหนึ่งจำนวน 50 คำถาม จากเว็บไซต์ของโรงแรม

2) กำหนดคำตอบอ้างอิงสำหรับแต่ละคำถามซึ่งเป็นคำตอบต้นฉบับจากเว็บไซต์ของโรงแรม เพื่อใช้เป็นเกณฑ์ในการเปรียบเทียบ

3) การสร้างคำตอบด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) โดยใช้คำสั่งดังตารางต่อไปนี้

ตารางที่ 1 คำสั่งสำหรับการสร้างคำตอบด้วย Generative AI แบบปกติ และบทบาทสมมติ

รูปแบบ	คำสั่ง
ปกติ	ตอบคำถามโดยใช้รายละเอียดใน INFO แท็ก <INFO>{ข้อมูลของโรงแรม}</INFO> คำถาม คือ {คำถาม}
บทบาทสมมติ	สวมบทบาทเป็นพนักงานต้อนรับ ตอบคำถามโดยใช้รายละเอียดใน INFO แท็ก <INFO>{ข้อมูลของโรงแรม}</INFO> คำถาม คือ {คำถาม}

1.2 ใช้ Generative AI แต่ละโปรแกรม สร้างคำตอบสำหรับคำถามเดียวกัน 3 ครั้ง (Iteration 1, Iteration 2, และ Iteration 3) เพื่อให้ได้คำตอบหลายรูปแบบจากคำถามเดียวกัน และทำการบันทึกคำตอบที่สร้างขึ้นทั้งหมด

1.3 การประเมินคำตอบด้วยอัลกอริทึม ROUGE-L

ประเมินคำตอบโดยใช้อัลกอริทึม ROUGE-L F1 เพื่อเปรียบเทียบคำตอบที่ได้จาก Generative AI กับคำตอบอ้างอิง ROUGE จะวัดความสอดคล้องของคำตอบตามลำดับประโยคและการจับคู่คำตามบริบทที่ใกล้เคียงมากที่สุด (Longest Common Subsequence; LCS) ทำการคำนวณ ROUGE สำหรับคำตอบแต่ละรอบที่สร้างจาก Generative AI โดยทำการคำนวณจาก ROUGE-L recall และ ROUGE-L precision ดังต่อไปนี้

$$ROUGE - L recall = \frac{\text{ข้อความร่วมกันที่ยาวที่สุด (LCS)}}{\text{จำนวนคำในคำตอบอ้างอิง}} \quad (1)$$

$$ROUGE - L precision = \frac{\text{ข้อความร่วมกันที่ยาวที่สุด (LCS)}}{\text{จำนวนคำในคำตอบที่ได้จาก Generative AI}} \quad (2)$$

$$ROUGE - L F1 = \frac{2 * ROUGE - L recall * ROUGE - L precision}{ROUGE - L recall + ROUGE - L precision} \quad (3)$$

งานวิจัยนี้ใช้ ROUGE-L เป็นตัววัดที่ใช้พิจารณาความเหมือนของลำดับคำ ซึ่งจะช่วยให้สามารถประเมินความคล้ายคลึงของคำตอบที่ได้จาก Generative AI กับคำตอบอ้างอิงได้อย่างยืดหยุ่นมากกว่าการเปรียบเทียบแบบคำต่อคำหรือกลุ่มของคำ อีกทั้งในบทความเดียวกันอาจมีรูปแบบของคำตอบที่ถูกต้องหลากหลายรูปแบบ ดังนั้นการใช้ ROUGE-L จึงช่วยให้สามารถวัดความคล้ายคลึงเชิงโครงสร้างได้โดยไม่ต้องอาศัยการเรียงลำดับคำที่แน่นอนได้อย่างยืดหยุ่น โดยค่า ROUGE-L จะอยู่ระหว่าง 0 ถึง 1 โดยคะแนนที่สูงกว่า (ใกล้ 1) บ่งชี้ว่าคำตอบที่สร้างขึ้นและคำตอบที่อ้างอิงมีความคล้ายคลึงกัน (Lin, 2004)

1.4 การวิเคราะห์ผลเปรียบเทียบคะแนนที่ได้จากอัลกอริทึม ROUGE ของคำตอบที่สร้างขึ้นจาก Generative AI แต่ละโปรแกรมทั้ง 3 ครั้ง บันทึกผลลัพธ์และวิเคราะห์ความแตกต่างระหว่าง Iterations เพื่อหาว่ารูปแบบใดของการสร้างคำตอบสอดคล้องกับมาตรฐานที่สุด โดยพิจารณาจากผลการให้คะแนน ROUGE ของแต่ละ Iteration จากทั้ง 3 รอบ แล้วนำมาหาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐาน

2. ประเมินการให้คะแนนการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT (GPT4o-mini), Claude (3.5 Sonnet) และ Gemini (1.5 Flash) ทำการประเมิน 2 รูปแบบ แบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) ด้วยมนุษย์ (ผู้เชี่ยวชาญ)

ในการเก็บรวบรวมข้อมูลการวิจัย ได้ดำเนินการตามลำดับขั้นตอน ดังนี้

2.1 สร้างแบบประเมินการให้คะแนนการตอบคำถามจากคำถามที่พบบ่อยของโรงแรม ด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ และแบบบทบาทสมมติ

2.2 นำคำตอบที่สร้างโดย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติและแบบบทบาทสมมติ มาใช้ในการประเมินคำตอบทั้งหมดโดยผู้เชี่ยวชาญ

2.3 ประเมินคำตอบที่สร้างจาก Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ และแบบบทบาทสมมติ ด้วยแบบประเมินที่ผู้วิจัยได้พัฒนาขึ้น โดยผู้เชี่ยวชาญจะทำการให้คะแนนคำตอบทุกข้อ โดยมีคะแนน ตั้งแต่ 1–5 คะแนน โดยที่ค่า 1 แสดงว่าคำตอบมีความใกล้เคียงกับคำตอบอ้างอิงที่น้อย จากนั้นไล่ระดับไปจนถึง 5 ซึ่งแสดงถึงว่าคำตอบมีความใกล้เคียงของคำตอบอ้างอิงที่มาก

2.4 ทำการวิเคราะห์ผลเปรียบเทียบคะแนนที่ได้จากผู้เชี่ยวชาญโดยใช้ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน

3. เปรียบเทียบคะแนนเฉลี่ยที่ได้จากการประเมินโดยแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน คือ 1) ประเมินโดยอัลกอริทึม ROUGE-L และ 2) ประเมินโดยผู้เชี่ยวชาญ โดยคะแนนที่ได้จะถูกนำมาเปรียบเทียบเฉพาะภายในกลุ่มเดียวกันเท่านั้นเพื่อศึกษาความแตกต่างของการตอบคำถาม

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเปรียบเทียบประสิทธิภาพในการตอบคำถามของ Generative AI ที่ประเมินการให้คะแนนโดยใช้ อัลกอริทึม ROUGE-L และมนุษย์ (ผู้เชี่ยวชาญ) โดยใช้คะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

ผลการวิจัย

การตอบคำถามของ Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) ได้ผลลัพธ์ดังตารางที่ 2 และตารางที่ 3

ตารางที่ 2 ตัวอย่างการตอบคำถามของ Generative AI ได้แก่ ChatGPT, Claude และ Gemini แบบปกติ

คำถาม	คำตอบ		
	ChatGPT	Claude	Gemini
โรงแรมมีสระว่ายน้ำของโรงแรมหรือไม่?	มีค่ะ โรงแรมมีสระว่ายน้ำ ตั้งอยู่ที่ชั้น 5 และชั้น 17 ของโรงแรม	มีค่ะ โรงแรมมีสระว่ายน้ำ 2 แห่ง ตั้งอยู่ที่ชั้น 5 และชั้น 17	มีค่ะ สระว่ายน้ำของโรงแรมตั้งอยู่ที่ชั้น 5 และชั้น 17
โรงแรมรับชำระเงินแบบใดบ้าง	โรงแรมรับการชำระเงินด้วย: บัตรเครดิต เงินสด ไม่สามารถใช้ค่าเงินอื่นในการชำระเงินได้ และราคาห้องพักที่แสดงบนเว็บไซต์รวมภาษี และค่าบริการแล้ว	โรงแรมรับชำระเงินเฉพาะ 2 วิธีคือ: บัตรเครดิต เงินสด โรงแรมรับเฉพาะเงินไทยบาทเท่านั้น ไม่สามารถจ่ายค่าห้องพักด้วยสกุลเงินอื่นได้	โรงแรมรับชำระเงินด้วยบัตรเครดิตและเงินสดเท่านั้น

ตารางที่ 3 ตัวอย่างการตอบคำถามของ Generative AI ได้แก่ ChatGPT, Claude และ Gemini แบบบทบาทสมมติ

คำถาม	คำตอบ		
	ChatGPT	Claude	Gemini
โรงแรมมีสระว่ายน้ำของโรงแรมหรือไม่?	โรงแรมมีสระว่ายน้ำให้บริการ คะ สระว่ายน้ำของเราตั้งอยู่บน ชั้น 5 และชั้น 17 คะ ท่าน สามารถเลือกใช้บริการตาม สะดวกเลยนะคะ หากต้องการ ข้อมูลเพิ่มเติมหรือความ ช่วยเหลืออื่น ๆ สามารถแจ้ง ดิฉันได้ค่ะ	ครับ โรงแรมของเรามีสระ ว่ายน้ำให้บริการ 2 แห่งครับ ตั้งอยู่ที่: ชั้น 5 ชั้น 17 ท่านสามารถเลือกใช้บริการ สระว่ายน้ำได้ตามความ สะดวก ซึ่งแต่ละสระก็จะมี บรรยากาศและทัศนียภาพที่ แตกต่างกันไปครับ มีอะไร เพิ่มเติมที่ต้องการทราบ เกี่ยวกับสระว่ายน้ำหรือสิ่ง อำนวยความสะดวกอื่น ๆ ของโรงแรมไหมครับ?	มีค่ะ โรงแรมของเรามีสระว่ายน้ำ ให้บริการ 2 สระด้วยกันนะคะ สระแรกอยู่ที่ชั้น 5 และอีกสระอยู่ ที่ชั้น 17 คะ ท่านสามารถเลือกใช้ บริการได้ตามความสะดวกเลยคะ
โรงแรมรับชำระเงินแบบใดบ้าง	ทางโรงแรมรับชำระเงินด้วย บัตรเครดิตและเงินสดเท่านั้นคะ สำหรับบัตรเครดิตสามารถใช้ เพื่อยืนยันการจองและชำระ ค่าใช้จ่ายได้ ส่วนสกุลเงินที่รับ เฉพาะเงินบาทเท่านั้นคะ ไม่ สามารถชำระด้วยสกุลเงินอื่นได้ หากมีคำถามเพิ่มเติมหรือข้อ สงสัยใด ๆ สามารถสอบถามได้ เลยนะคะ ยินดีให้บริการคะ	โรงแรมของเรารับชำระเงิน เป็นบัตรเครดิตและเงินสด เท่านั้นครับ โดยเรารับ เฉพาะสกุลเงินไทยบาทนะ ครับ ไม่สามารถชำระด้วย สกุลเงินอื่นได้ครับ หากคุณ มีคำถามเพิ่มเติมเกี่ยวกับ การชำระเงิน ผมยินดีให้ ข้อมูลเพิ่มเติมครับ	โรงแรมรับชำระเงินแบบบัตร เครดิตและเงินสดเท่านั้นคะ

ตารางที่ 4 จำนวนคำเฉลี่ยจากคำตอบทั้งหมดที่ได้จาก Generative AI จากแบบปกติ และแบบบทบาทสมมติ

Generative AI	Standard prompt	Roleplay prompt	เปลี่ยนแปลง(%)
ChatGPT	41.28	58.11	40.77
Claude	77.76	79.38	2.08
Gemini	110.83	123.39	11.33

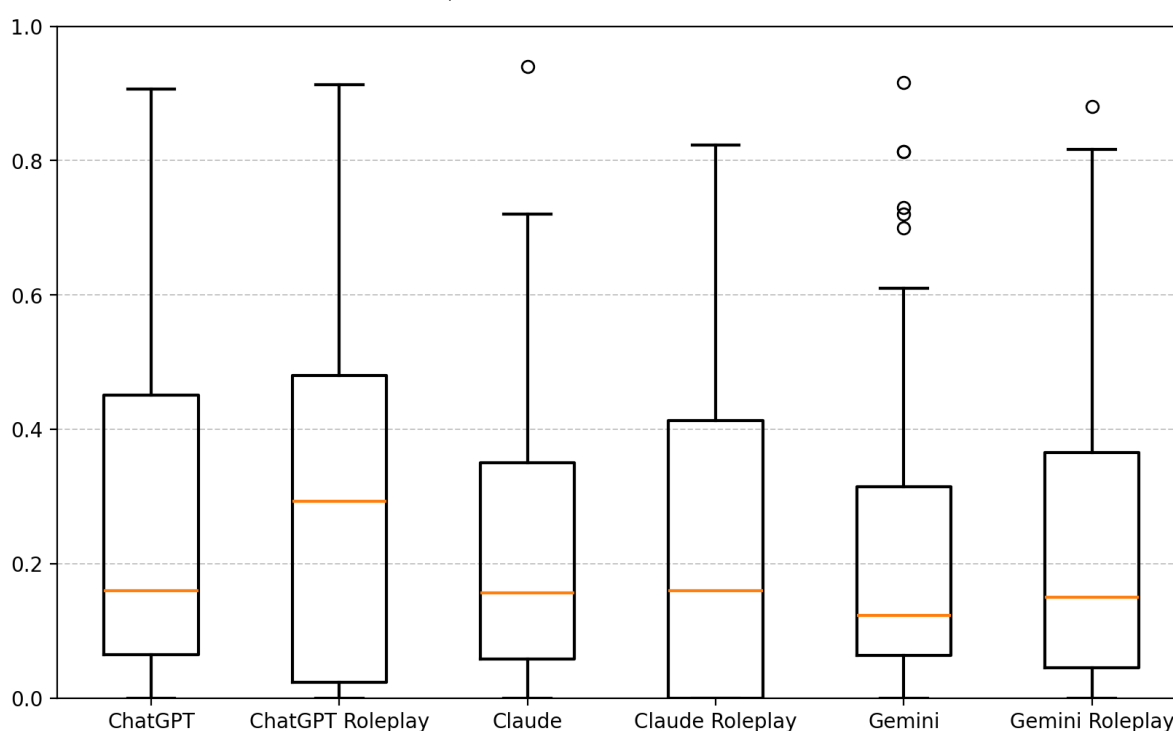
จากตารางที่ 2 และ 3 จะพบว่า การคำสั่งแบบบทบาทสมมติส่งผลให้คำตอบจาก Generative AI มีจำนวนข้อความที่ยาวขึ้น ซึ่งสอดคล้องกับตารางที่ 4 ที่แสดงให้เห็นว่าคำตอบจากคำสั่งแบบบทบาทสมมติมีจำนวนคำเฉลี่ยมากกว่าคำตอบจากคำสั่งแบบปกติ โดย ChatGPT มีอัตราการเพิ่มขึ้นสูงสุด 40.77% จาก 41.28 คำเป็น 58.11 คำ รองลงมาคือ Gemini เพิ่มขึ้น 11.33% จาก 110.83 คำเป็น 123.39 คำ และ Claude เพิ่มขึ้นเพียง 2.08% จาก 77.76 คำเป็น 79.38 คำ จากผลลัพธ์นี้แสดงให้เห็นว่าคำสั่งแบบบทบาทสมมติมีส่วนช่วยกระตุ้นให้ AI สร้างคำตอบที่ละเอียดและยาวขึ้น โดยเฉพาะในกรณีของ ChatGPT ซึ่งมีการเพิ่มขึ้นของจำนวนคำมากที่สุด

ผู้วิจัยได้รวบรวมข้อมูลจากการประเมินการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) โดยใช้อัลกอริทึม ROUGE-L และมนุษย์ (ผู้เชี่ยวชาญ) สรุปผลการประเมินการตอบคำถามด้วย Generative AI ดังนี้

ตารางที่ 5 ผลการประเมินการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) โดยใช้อัลกอริทึม ROUGE-L

Generative AI	Standard prompt		Roleplay prompt	
	$\bar{x}$	S.D.	$\bar{x}$	S.D.
ChatGPT	0.26547	0.23861	0.29307	0.27855
Claude	0.22400	0.20983	0.22993	0.23479
Gemini	0.23213	0.24761	0.23747	0.23288

จากตารางที่ 5 พบว่า ผลการประเมินการตอบคำถามด้วย Generative AI ทั้งในแบบปกติและแบบบทบาทสมมติ พบว่า ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ ChatGPT มีค่าเฉลี่ยเท่ากับ 0.26547 รองลงมาคือ Gemini มีค่าเฉลี่ยเท่ากับ 0.23213 และ Claude มีค่าเฉลี่ยเท่ากับ 0.22400 สำหรับผลการประเมินการตอบคำถามด้วย Generative AI แบบบทบาทสมมติ พบว่า ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุดยังคงเป็น ChatGPT ซึ่งมีค่าเฉลี่ยเท่ากับ 0.29307 รองลงมาคือ Gemini มีค่าเฉลี่ยเท่ากับ 0.23747 และคะแนนประเมินต่ำที่สุด ได้แก่ Claude ซึ่งมีค่าเฉลี่ยเท่ากับ 0.22993 ทั้งนี้เมื่อพิจารณาจำนวนค่าเฉลี่ยของคำตอบจาก Generative AI ในตารางที่ 4 โดยเปรียบเทียบกับคำสั่งแบบปกติและแบบบทบาทสมมติ พบว่า การมีจำนวนข้อความที่ยาวขึ้นอาจส่งผลต่อการได้ค่า ROUGE-L ให้ที่สูงขึ้น แต่เมื่อพิจารณา Generative AI ทั้ง 3 พบว่า ChatGPT จะให้คำตอบที่มีค่าเฉลี่ยจำนวนคำที่น้อยกว่า Claude และ ChatGPT แต่กลับได้ค่า ROUGE-L ที่สูงกว่า ซึ่งแสดงให้เห็นว่าคำตอบของ ChatGPT มีความสอดคล้องกับคำตอบอ้างอิงมากกว่า ดังนั้นจำนวนคำอาจไม่ใช่ปัจจัยที่ส่งผลโดยตรงต่อคุณภาพของคำตอบ



รูปที่ 2 เปรียบเทียบการกระจายของผลการประเมินการตอบคำถามของ ChatGPT Claude และ Gemini แบบปกติ และแบบบทบาทสมมติด้วยแผนภาพกล่อง

จากรูปที่ 2 แสดงการกระจายของผลการประเมินการตอบคำถามของ ChatGPT Claude และ Gemini แบบปกติ และแบบบทบาทสมมติด้วยแผนภาพกล่อง พบว่า ChatGPT แบบบทบาทสมมติมีประสิทธิภาพการตอบคำถามสูงที่สุดโดยมีค่ามัธยฐาน (Median) และควอไทล์ที่สาม (Q3) สูงกว่า Generative AI อื่นอย่างชัดเจน สะท้อนถึงความสามารถในการตอบคำถามได้ใกล้เคียงกับคำตอบที่ถูกต้อง Claude มีประสิทธิภาพรองลงมา ในขณะที่ Gemini มีค่ามัธยฐานที่ต่ำที่สุด อีกทั้งยังพบว่าการตอบคำถามแบบบทบาทสมมติมีค่ามัธยฐานที่สูงกว่าแบบปกติในทุก Generative AI นอกจากนี้เมื่อพิจารณาส่วนของค่าสุดโต่ง (Outlier) จะพบมากที่สุดที่ Gemini และรองลงมา คือ Claude และ Gemini Roleplay

ตารางที่ 6 ผลการประเมินการตอบคำถามด้วย Generative AI ได้แก่ ChatGPT, Claude และ Gemini ทั้งแบบปกติ (Standard prompt) และแบบบทบาทสมมติ (Roleplay prompt) จากผู้เชี่ยวชาญ

Generative AI	Standard prompt		Roleplay prompt	
	$\bar{x}$	S.D.	$\bar{x}$	S.D.
ChatGPT	4.1556	0.32126	4.2867	0.21670
Claude	3.8733	0.32887	4.0178	0.31534
Gemini	3.9622	0.43125	4.0222	0.31190

จากตารางที่ 6 พบว่า ผลการประเมินการตอบคำถามด้วย Generative AI แบบปกติและแบบบทบาทสมมติจากผู้เชี่ยวชาญ ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ ChatGPT มีค่าเฉลี่ยเท่ากับ 4.1556 รองลงมาได้แก่ Gemini มีค่าเฉลี่ยเท่ากับ 3.9622 และ Claude มีค่าเฉลี่ยเท่ากับ 3.8733 และผลการประเมินการตอบคำถามด้วย Generative AI แบบบทบาทสมมติ (Roleplay prompt) ผลการประเมินคำตอบที่คล้ายคลึงกับคำตอบที่ถูกต้องมากที่สุด ได้แก่ ChatGPT มีค่าเฉลี่ยเท่ากับ 4.2867 รองลงมาได้แก่ Gemini มีค่าเฉลี่ยเท่ากับ 4.0222 และคะแนนประเมินน้อยที่สุด ได้แก่ Claude มีค่าเฉลี่ยเท่ากับ 4.0178

การอภิปรายผล

ผลการวิจัยแสดงให้เห็นว่า ChatGPT มีประสิทธิภาพในการตอบคำถามที่ดีที่สุดในคำสั่ง แบบปกติ และ บทบาทสมมติ ซึ่งสะท้อนถึงความสามารถของ ChatGPT ในการตีความของคำถามและสร้างคำตอบได้ใกล้เคียงกับคำตอบที่ถูกต้องมากที่สุด โดยเฉพาะการตอบคำถามในรูปแบบ บทบาทสมมติ พบว่า ChatGPT มีค่าเฉลี่ย ROUGE-L สูงที่สุดที่ 0.29307 และสอดคล้องกับคะแนนจากผู้เชี่ยวชาญสูงสุดที่ 4.2867 ขณะที่ Gemini และ Claude มีประสิทธิภาพรองลงมาตามลำดับ นอกจากนี้พบว่า เมื่อใช้คำสั่งแบบบทบาทสมมติ ทำให้ ChatGPT มีค่าเฉลี่ยของจำนวนคำในคำตอบเพิ่มขึ้น 40.77% เมื่อเทียบกับคำสั่งแบบปกติ ซึ่งเป็นอัตราการเพิ่มขึ้นที่สูงกว่าทั้ง Gemini (11.33%) และ Claude (2.08%) แม้ว่าคำสั่งแบบบทบาทสมมติจะทำให้คำตอบยาวขึ้น แต่ ChatGPT ยังคงมีจำนวนคำเฉลี่ยน้อยที่สุดที่ 58.11 คำ เมื่อเทียบกับ Claude (79.38 คำ) และ Gemini (123.39 คำ) แสดงให้เห็นว่า ChatGPT สามารถตอบคำถามได้ กระชับและตรงประเด็นมากกว่า Claude และ Gemini ดังนั้นจำนวนคำที่มากกว่าอาจไม่ใช่ปัจจัยที่ส่งผลโดยตรงต่อคุณภาพของคำตอบ แต่ความถูกต้องและความสอดคล้องกับคำตอบอ้างอิงจะเป็นปัจจัยที่มีความสำคัญมากกว่า จากนั้นเมื่อพิจารณาค่าสุดโต่ง (Outlier) พบว่า มีจำนวนมากใน Gemini ซึ่งอาจเป็นผลมาจากการที่ Gemini มีความสามารถในการเชื่อมต่อหรือสืบค้นข้อมูลจากแหล่งข้อมูลภายนอกหรือเว็บไซต์ต่าง ๆ ทำให้มีโอกาสให้คำตอบนอกเหนือเนื้อหาที่ระบุไว้ เช่น มีการให้คำตอบที่นอกเหนือจากบริบทของโรงแรม หรือ ข้อมูลของโรงแรมแต่จากเว็บไซต์อื่นที่มีข้อมูลไม่ตรงกัน เป็นต้น นอกจากนี้ ยังพบว่า การกำหนดบทบาทสมมติ หรือการสร้างบทบาทที่ชัดเจนให้กับ Generative AI ทำให้การตีความและตอบคำถามมีความถูกต้องและแม่นยำยิ่งขึ้นมากกว่าการถามตอบแบบปกติ

อย่างไรก็ตาม การวิจัยนี้ยังมีข้อจำกัดในการตอบคำถามจากบริบทของโรงแรมอย่างเดียว ซึ่งอาจจะไม่ครอบคลุมในการตอบคำถามในบริบทอื่น ๆ อีกทั้งยังทำการศึกษาเฉพาะภาษาไทยเท่านั้น ซึ่งในอนาคตผู้วิจัยจะมีการศึกษาความสามารถในบริบทอื่น ๆ รวมทั้งเพิ่มเติมรูปแบบของคำสั่ง และรูปแบบของคำถามที่ซับซ้อนกว่านี้

บทสรุป

งานวิจัยนี้เป็นการประเมินและเปรียบเทียบประสิทธิภาพการบริการตอบคำถามลูกค้าที่สร้างขึ้นจาก Generative AI ได้แก่ ChatGPT, Claude และ Gemini โดยทำการประเมินคำตอบจากการใช้อัลกอริทึม ROUGE-L และมนุษย์ เพื่อวิเคราะห์ถึงความแม่นยำของการตอบสนองในบริบทของการให้บริการลูกค้า ซึ่งทำการทดสอบโดยใช้คำถาม 50 คำถาม และทำการทดสอบโดยการเปรียบเทียบค่าคะแนนเฉลี่ยของการตอบคำถามในเรื่องความถูกต้องในการตอบคำถาม ผลการวิจัยพบว่า ChatGPT ให้ค่าคะแนนเฉลี่ยมากที่สุดในการให้คำตอบที่ใกล้เคียงกับคำตอบที่ถูกต้องมากที่สุดทั้ง 2 รูปแบบคือการประมวลผลคำตอบในรูปแบบปกติ (Standard prompt) และรูปแบบบทบาทสมมติ (Roleplay prompt) สามารถนำไปเป็นทางเลือกเพื่อนำไปประยุกต์ใช้ในการเป็นแชทบอทหรือเป็นผู้ช่วยในการตอบคำถามอัตโนมัติในธุรกิจหรือหน่วยงานต่าง ๆ ได้ เพื่อสนับสนุนงานบริการลูกค้า ให้การบริการตอบคำถามได้แม่นยำใกล้เคียงกับคำตอบที่ถูกต้อง ซึ่งการประยุกต์ใช้ Generative AI ในการตอบคำถามลูกค้าช่วยลดเวลาที่ใช้ในการให้บริการและเพิ่มความพึงพอใจของลูกค้าได้ มีความยืดหยุ่นในการตอบคำถามที่ซับซ้อนและมีรายละเอียดมากขึ้น

เอกสารอ้างอิง

- นันทน์ภัส ประจกการ. (2560). แนวทางการปรับใช้แชทบอทสำหรับงานบริการลูกค้า. การค้นคว้าอิสระสาขาวิชาวิทยาศาสตร์มหาบัณฑิต สาขาวิชาการบริหารการตลาด คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์.
- ศรุต จรุงเมธี. (2567). 5 เหตุผลที่ช่วย Generative AI จำเป็นกับงานบริการลูกค้า. ค้นเมื่อ 10 กันยายน 2567.
<https://www.amitysolutions.com/th/blogs/5-ways-Generative-ai-assistants-master-customer-service>
- สมพงษ์ อัครวิธิบัติ และรชพร จันทร์สว่าง. (2566). ประสิทธิภาพระบบตอบรับอัตโนมัติที่ส่งผลกระทบต่อคุณภาพบริการด้านการสื่อสารของธุรกิจสายการบินที่ทำการบินภายในประเทศ. วารสารวิจัยรำไพพรรณี. 17(3): 77-89.
- สุทธิภาส ชาญชัยประสิทธิ์ และวศิน ชูประยูร. (2563). การประเมินประสิทธิภาพ ประสิทธิผล และความพึงพอใจในการใช้โปรแกรม Chatbot ธนาคารในประเทศไทย. วารสารรังสิตสารสนเทศ. 26(1): 117-136.
- Boruah M. (2024). Chatbots vs. Humans: Which one should you choose and why?. Accessed 11 Oct. 2024.
<https://www.kommunicate.io/blog/chatbots-vs-humans/>
- Chowdhury N.M. (2023). Improving customer care with ChatGPT: A case study. Bachelor of Science Computer Science and Technology. Chongqing University of Posts and Telecommunications Chongqing China.
- Gartner. (2024). Top strategic technology trends 2024: Advancing business value with AI trust and sustainability. Accessed 11 Oct. 2024. <https://www.gartner.com/en/articles/gartner-top-10-strategic-technology-trends-for-2024>
- IBM. (2023). 5 types of chatbot and how to choose the right one for your business. Accessed 6 Jul. 2024.
<https://www.ibm.com/think/topics/chatbot-types>
- Lee, G.-G., Latif, E., Wu, X., Liu, N., and Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. Computers and Education: Artificial Intelligence, 6, 100213.
- Lin C.Y. (2004). ROUGE: A Package for automatic evaluation of summaries. Text Summarization Branches Out. Barcelona. Spain.
- Lins dos Santos, L.F. (2023). Evaluating and comparing generative-based chatbots based on process requirements. Master of Mathematics Thesis. Computer Science, University of Waterloo.
- McKinsey and Company. (2023). The state of AI in 2023: Generative AI's breakout year. McKinsey Digital. Accessed 11 Oct. 2024. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-Generative-ais-breakout-year>

- MIT Technology Review. (2019). The AI issue: The state of artificial intelligence. Massachusetts Institute of Technology. Accessed 6 Sep. 2024. <https://www.technologyreview.com/2019/01/08/137912/the-state-of-artificial-intelligence/>
- Poonpresartporn A., Uttarakorn P. and Chuaybamroong R. (2023). "Comparing the accuracy and referencing of ChatGPT's responses to herbal medicine queries: A zero-shot versus roleplay prompting approach. *The Thai Journal of Pharmaceutical Sciences*. 47(4): 1-11.
- Stryker C. and Scapicchio M. (2024). What is Generative AI. Accessed 6 Sep. 2024. <https://www.ibm.com/topics/Generative-ai>
- Subagja A.D., Ausat A.MA., Sari A.R., Wanof M.I and Suherlan. (2023). Improving customer service quality in MSMEs through the use of ChatGPT. *Jurnal Minfo Polgan*. 12(2): 380-386.
- World Economic Forum. (2023). The future of jobs report 2023. World Economic Forum. Accessed 6 Sep. 2024. <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>
- Wu Y., Henriksson A., Duneld M. and Nouri J. (2023). Towards improving the reliability and transparency of chatgpt for educational question answering. In *European Conference on Technology Enhanced Learning* 28 August 2023. Switzerland. 475-488.
- Yun J. and Park J. (2022). The effects of chatbot service recovery with emotion words on customer satisfaction, repurchase intention, and positive word-of-mouth. Accessed 6 Sep. 2024. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.922503/full>