

ผลกระทบของการเปลี่ยนแปลงคุณลักษณะของคำศัพท์ในเนื้อหา
ที่สร้างโดย ChatGPT โดยใช้แนวทาง N-Gram

The Effects of Lexical-Features Changes in ChatGPT-Generated
Content Using N-Gram Approach

¹ชัชพงศ์ หาเชิงสูงหอม, ²มุกชดา สุขธาราจารย์, ³ธนศักดิ์ ศิริคะเนรัตน์

⁴ภิรมพลีทฐ์ เตชะราชันย์ และ ⁵ลีนา มะลูลีม

¹Chatpong Hacherngsoonghom, ²Mukda Suktarachan, ³Thanasak Sirikanerat

⁴Bhimbasistha Tejarajanya, and ⁵Lena Maluleem

ภาควิชาภาษาศาสตร์ คณะมนุษยศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

Department of Linguistics, Faculty of Humanities, Kasetsart University

E-mail: ¹chatpong.h@ku.th, ²mukda.su@ku.th, ³thanasak.si@ku.th

⁴bhimbasistha.t@ku.th and ⁵lena.m@ku.th

Received December 12, 2024; Revised January 21, 2025; Accepted February 15, 2025

บทคัดย่อ

บทความนี้มีวัตถุประสงค์ 1) เพื่อวิเคราะห์การเปลี่ยนแปลงชนิดของคำศัพท์ที่ปรากฏในเนื้อหาที่สร้างโดย ChatGPT ผ่านการวิเคราะห์เชิง n-gram และ 2) เพื่อวิเคราะห์คุณลักษณะเชิงศัพท์ของงานเขียนที่ปรับคำโดย ChatGPT ในตำแหน่งแกรมที่ 3 รูปแบบการวิจัยเป็นการวิจัยเชิงปริมาณ ใช้แนวคิดการวิเคราะห์เชิง n-gram โดยเน้นตำแหน่งแกรมที่ 3 ของเนื้อหาเป็นกรอบการวิจัยเพื่อประเมินการเปลี่ยนแปลงชนิดคำและลักษณะเชิงศัพท์จากการปรับคำ พื้นที่วิจัยเป็นการวิจัยเอกสาร กลุ่มตัวอย่างคือ นิทาน จำนวน 8 เรื่อง ใช้วิธีคัดเลือกแบบสุ่มเจาะจง เครื่องมือที่ใช้ในการวิจัยมี 2 ชนิด คือ 1) โปรแกรม ChatGPT 2) โปรแกรมไพธอน ผ่าน Google Collaboratory วิเคราะห์ข้อมูลโดยใช้สถิติพื้นฐาน ผลการวิจัยพบว่า 1. การเปลี่ยนแปลงของเนื้อหาในแกรมที่ 3 ส่งผลต่อความสมจริงและความสอดคล้องของเนื้อหา โดยคำนามเป็นคำที่มีความถี่ในการเปลี่ยนแปลงสูงสุด ถัดไปเป็น คำกริยา คำบ่งชี้ คำสรรพนาม คำคุณศัพท์ คำสันธาน คำวิเศษณ์ ตัวเลข และคำอุทาน ตามลำดับ 2. ลักษณะเชิงศัพท์ที่มีรูปแบบความสัมพันธ์ 8 ประเภท เช่น 1) Antonym 2) Co-hyponym 3) Co-hypernym 4) Hyponym 5) Hypernym 6) Synonym 7) Meronym และ 8) Holonym รวมถึงรูปแบบใหม่ที่เกิดจาก

การใช้ Word Embeddings Similarity ซึ่งมีอัตราการเกิดสูงสุด สะท้อนถึงศักยภาพของ AI ในการรักษาบริบทและความเหมาะสมของเนื้อหา องค์ความรู้จากงานวิจัยนี้ช่วยเสริมสร้างความเข้าใจการบูรณาการภาษาศาสตร์กับเทคโนโลยีปัญญาประดิษฐ์ โดยองค์ความรู้ที่ได้สามารถนำไปประยุกต์ใช้ในการพัฒนาวิธีการประเมินเนื้อหาที่สร้างโดย AI การปรับปรุงคำสั่ง (Prompts) ให้มีประสิทธิภาพ และออกแบบสื่อการสอนที่เหมาะสม นอกจากนี้ ยังสนับสนุนการใช้งาน AI อย่างรับผิดชอบ โปร่งใส และมีจริยธรรม เพื่อประโยชน์ต่อการวิชาการ อุตสาหกรรม และสังคมโดยรวม

คำสำคัญ: คุณลักษณะของคำศัพท์; แฉัตจีพีที; N-gram

Abstract

This article aims 1) to analyze changes in word types appearing in content generated by ChatGPT through n-gram analysis and 2) to examine the lexical features of revised content modified by ChatGPT at the third-gram position. The research adopts a quantitative approach, employing n-gram analysis focusing on the third-gram position as the framework for evaluating changes in word types and lexical features resulting from word modifications. The study is document-based, with the sample comprising eight children's stories selected through purposive sampling. The research utilized two tools: 1) the ChatGPT program and 2) Python via Google Collaboratory. The data were analyzed using basic statistical methods. The findings revealed that 1. Regarding the first objective, modifications at the third-gram position influenced the realism and coherence of the content. Nouns were the most frequently affected word type, followed by verbs, determiners, pronouns, adjectives, conjunctions, adverbs, numerals, and interjections, respectively. 2. For the second objective, eight types of lexical relationships were identified: 1) Antonym, 2) Co-hyponym, 3) Co-hypernym, 4) Hyponym, 5) Hypernym, 6) Synonym, 7) Meronym, and 8) Holonym. A new pattern, Word embedding similarity, was also observed as the most frequently occurring relationship, reflecting the AI's capability to maintain contextual relevance and content appropriateness. The knowledge derived from this research enhances the integration of linguistics and artificial intelligence (AI). The user can apply this model to develop improved methods for evaluating AI-generated content, refining prompts for better performance, and designing suitable teaching materials. Furthermore, this study advocates for responsible, transparent, and ethical AI use, contributing to advancements in academia, industry, and society.

Keywords: Lexical-Features Changes; ChatGPT; N-gram

บทนำ

ในยุคที่ปัญญาประดิษฐ์ (AI) มีบทบาทสำคัญในการสร้างเนื้อหาในบริบทต่าง ๆ การศึกษาผลกระทบของการเปลี่ยนแปลงคุณลักษณะของคำศัพท์ในเนื้อหาที่สร้างโดยเอไอ เช่น ChatGPT จึงมีความสำคัญ งานวิจัยที่ผ่านมาแสดงให้เห็นถึงการเปลี่ยนแปลงคำศัพท์ที่ส่งผลต่อคุณภาพและความสอดคล้องของข้อความที่สร้างโดยโมเดลภาษาหรือเอไอ ในขณะที่ปัญญาประดิษฐ์ยังคงพัฒนาต่อไป โดยเฉพาะอย่างยิ่งในงานวิจัยด้านการประมวลผลภาษาธรรมชาติ (NLP) ซึ่งมีความพยายามในการตรวจสอบลักษณะเฉพาะของเนื้อหาที่สร้างโดยเอไอ โดยใช้ประโยชน์จากระเบียบวิธี n-gram ซึ่งวิเคราะห์ลำดับคำที่ต่อเนื่องกัน ทำให้สามารถระบุรูปแบบที่แสดงถึงการเปลี่ยนแปลงคำศัพท์และนัยยะของการเปลี่ยนแปลงเหล่านี้ต่อการสร้างเนื้อหาและความถูกต้องได้ (GeekforGeeks, 2020, 2021; Arts2Survive, 2022) ความสำคัญของงานวิจัยนี้เกิดจากการที่ข้อความที่สร้างโดยเอไอแพร่หลายและได้รับความนิยมมากขึ้น ความกังวลเกี่ยวกับความน่าเชื่อถือ ความชัดเจน และความสมบูรณ์ของเนื้อหาที่สร้างขึ้นก็กลายมาเป็นประเด็นสำคัญ (Javapoint, n.d.; Ramponi, 2022) นอกจากนี้การปรับเปลี่ยนคำศัพท์ ความซับซ้อน และความเป็นทางการสามารถส่งผลโดยตรงต่อความเข้าใจและการมีส่วนร่วมของผู้ใช้ ส่งผลให้ประสิทธิภาพโดยรวมของการสื่อสารได้รับผลกระทบ (Patel, 2024) การศึกษาประเด็นต่าง ๆ เหล่านี้มีความเกี่ยวข้องอย่างยิ่งกับการเรียนการสอน ซึ่งการแยกแยะเนื้อหาระหว่างเนื้อหาที่สร้างโดยมนุษย์และโดยเอไอเป็นความท้าทายในการประเมินการเขียนของนักเรียนและความซื่อสัตย์ทางวิชาการ ทำให้เกิดคำถามเกี่ยวกับความสามารถในการเขียนของแต่ละบุคคลและความเสี่ยงของทักษะการคิดวิเคราะห์ที่ลดน้อยลงในหมู่ผู้เรียน (GeekforGeeks, 2021; Javapoint, n.d.)

อย่างไรก็ตาม จากการสำรวจข้อจำกัดของโมเดล n-gram รวมถึงปัญหาที่เกี่ยวข้องกับความไม่เพียงพอของข้อมูล ความซับซ้อนในการคำนวณ และความสามารถในการเชื่อมโยงข้อความที่มีการอ้างอิงระยะไกลในข้อความ ซึ่งอาจเป็นอุปสรรคในการสร้างเนื้อหาที่มีความสอดคล้องกัน (Imani & Habil, 2014; Arts2Survive, 2022; Martínez et al., 2024) การแก้ไขข้อกังวลเหล่านี้ถือเป็นสิ่งสำคัญต่อการเพิ่มประสิทธิภาพและการประยุกต์ใช้เอไอ ในการสร้างเนื้อหาที่เขียนขึ้นให้มีคุณภาพ ซึ่งเป็นประเด็นสำคัญในวงการวิชาการ การตลาด และสื่อสารมวลชน

พื้นที่วิจัยมุ่งเน้นการวิเคราะห์ข้อความที่สร้างโดย ChatGPT ซึ่งได้รับความนิยมมากขึ้นเรื่อย ๆ ในแวดวงการศึกษาและการตลาด เนื่องจากสามารถตอบโต้การสร้างเนื้อหาที่หลากหลาย นักวิจัยได้เลือกนิทานสำหรับเด็กจำนวน 8 เรื่องจากเว็บไซต์ American Literature ซึ่งเป็นแหล่งข้อมูลที่มีเนื้อหาเรียบง่ายและเหมาะสมสำหรับการวิเคราะห์คุณลักษณะเชิงศัพท์ กลุ่มเป้าหมายของการศึกษานี้ ได้แก่ ผู้พัฒนาเอไอ และผู้ใช้งานทั่วไปที่ต้องการปรับปรุงคุณภาพเนื้อหาด้วย ChatGPT

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์การเปลี่ยนแปลงชนิดคำในเนื้อหาที่สร้างโดย ChatGPT ผ่าน n-gram และศึกษาคุณลักษณะเชิงศัพท์ของคำที่ปรับในตำแหน่งแกรมที่ 3 การวิจัยใช้ระเบียบวิธี

เชิงปริมาณ โดยอาศัยโปรแกรม ChatGPT และ Python ผ่าน Google Collaboratory เพื่อวิเคราะห์ข้อมูล และประเมินผลลัพธ์ และนำเสนอผลการวิเคราะห์การเปลี่ยนแปลงคำและความสัมพันธ์เชิงศัพท์ รวมถึง Word Embeddings Similarity ที่สะท้อนศักยภาพของเอไอ ในการรักษาบริบทของเนื้อหา ซึ่งผล การศึกษาจะช่วยพัฒนาวิธีการประเมินเนื้อหาเอไอ ให้แม่นยำและเหมาะสมยิ่งขึ้น สนับสนุนการสร้าง เนื้อหาที่โปร่งใสและรับผิดชอบต่อสังคม อีกทั้งยังสามารถนำไปปรับใช้ในการศึกษา การตลาด และการพัฒนาสื่อสารมวลชนเพื่อเพิ่มประสิทธิภาพการสื่อสารในระดับต่าง ๆ อย่างยั่งยืน

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์การเปลี่ยนแปลงชนิดของคำศัพท์ที่ปรากฏในเนื้อหาที่สร้างโดย ChatGPT ผ่าน การวิเคราะห์เชิง n-gram
2. เพื่อวิเคราะห์คุณลักษณะเชิงศัพท์ของงานเขียนที่ปรับคำโดย ChatGPT ในตำแหน่งแกรมที่ 3

การทบทวนวรรณกรรม

การทบทวนวรรณกรรมแบ่งเป็น 2 ส่วน ได้แก่แนวคิดและทฤษฎีที่เกี่ยวข้องกับเทคโนโลยี ปัญญาประดิษฐ์และแนวคิดและทฤษฎีด้านภาษาศาสตร์ โดยมีรายละเอียดดังนี้

เทคโนโลยีปัญญาประดิษฐ์

ตั้งแต่ปัญญาประดิษฐ์ (AI) ถือกำเนิดขึ้นมาก็ได้เปลี่ยนแปลงสาขาต่าง ๆ อย่างมีนัยสำคัญ รวมถึงอาณาจักรของเรื่องของการสร้างเนื้อหาด้วย ส่วนที่โดดเด่นในการเปลี่ยนแปลงนี้คือการ ประมวลผลภาษาธรรมชาติ (Natural Language Processing หรือ NLP) ซึ่งเป็นสาขาย่อยของเอไอ ที่เน้น การทำให้คอมพิวเตอร์เข้าใจและโต้ตอบกับภาษาของมนุษย์ได้อย่างมีประสิทธิภาพ (GeekforGeeks, 2020; Arts2Survive, 2022) NLP ครอบคลุมขั้นตอนที่สำคัญหลายขั้นตอน รวมถึงการวิเคราะห์ คำศัพท์ ซึ่งเกี่ยวข้องกับการแบ่งข้อความออกเป็นหน่วยที่จัดการได้ และการวิเคราะห์เชิงความหมาย ซึ่งมุ่งเน้นที่จะดึงความหมายจากภาษา (GeekforGeeks, 2020; Aggarwal, 2023)

ในอดีตระบบ NLP พึ่งพากฎที่สร้างขึ้นด้วยมืออย่างมาก จนถึงช่วงทศวรรษ 1980 หลังจากนั้น อัลกอริทึมการเรียนรู้ของเครื่องจักรก็เริ่มเข้ามามีบทบาทมากขึ้น ทำให้มีความสามารถในการ ประมวลผลภาษาที่ดีขึ้น การเปลี่ยนแปลงครั้งนี้ได้วางรากฐานให้กับแอปพลิเคชัน NLP สมัยใหม่ ซึ่งขณะนี้มีการใช้งานต่าง ๆ เช่น การจดจำเสียงพูด การแปลด้วยเครื่อง และการสร้างข้อความ (Jagarlapoodi, 2023) ความก้าวหน้าเหล่านี้รวมถึงการพัฒนาโมเดลเชิงสร้างสรรค์ เช่น ChatGPT ซึ่งช่วยให้สามารถสร้างข้อความที่สอดคล้องและเกี่ยวข้องกับบริบทได้อย่างรวดเร็ว จึงเกิดคำถาม เกี่ยวกับลักษณะเฉพาะและคุณภาพของเนื้อหาที่สร้างโดยเอไอ (Javapoint, n.d.) การใช้ n-gram ซึ่งเป็นวิธีการวิเคราะห์ลำดับที่ต่อเนื่องกันของรายการ n รายการจากตัวอย่างข้อความที่กำหนด

ถือเป็นแนวทางที่มีค่าสำหรับการตรวจสอบคุณลักษณะของคำศัพท์ในการสร้างข้อความ (Arts2Survive, 2022) การใช้ n-gram ช่วยให้นักวิจัยค้นพบรูปแบบและการเปลี่ยนแปลงในคุณลักษณะของคำศัพท์ในเนื้อหาที่สร้างโดยเอไอ ให้ข้อมูลเชิงลึกเกี่ยวกับการทำงานของโมเดลและผลกระทบของผลลัพธ์ การทำความเข้าใจผลกระทบของการเปลี่ยนแปลงคำศัพท์ในโมเดล เช่น ChatGPT มีความสำคัญต่อการประเมินความถูกต้องและความน่าเชื่อถือของข้อความ เนื่องจากเอไวมักมีข้อจำกัดในการระบุความแตกต่างในภาษาของมนุษย์ จึงจำเป็นต้องมีวิธีการประเมินคุณภาพที่แข็งแกร่งยิ่งขึ้น (GeekforGeeks, 2021; Ramponi, 2022)

การใช้ n-gram ในการสร้างเนื้อหาโดย GPT

N-grams คือชุดลำดับของคำหรืออักขระที่ต่อเนื่องกัน ซึ่งเป็นพื้นฐานสำคัญของการประมวลผลภาษาธรรมชาติ (NLP) และภาษาศาสตร์เชิงคำนวณ แนวคิดนี้เริ่มจากการพัฒนาแบบจำลองภาษาในช่วงแรก เพื่อทำความเข้าใจและคาดการณ์ลำดับข้อความ โดย N-grams มีจุดเด่นในการจับความสัมพันธ์ในท้องถิ่นและบริบท ทำให้ถูกนำไปใช้ในแอปพลิเคชัน NLP หลากหลายประเภท (Lang, 2024; Gomed, 2023) การประยุกต์ใช้ n-grams ในยุคแรกๆ สามารถสืบย้อนไปถึงโมเดลภาษาสถิติในช่วงทศวรรษ 1980 ซึ่งพยายามทำนายความน่าจะเป็นของคำโดยพิจารณาจากบริบทก่อนหน้า แนวทางนี้ได้รับการริเริ่มในด้านการจดจำเสียงพูดและการแปลด้วยเครื่อง โดยโมเดล n-gram ช่วยในระบบสร้างเอนต์พุดที่เกี่ยวข้องกับบริบทได้โดยการวิเคราะห์รูปแบบในข้อความขนาดใหญ่ (Madala, 2023; Nguyen, 2020) เมื่อทรัพยากรการคำนวณขยายตัวและข้อมูลสามารถเข้าถึงได้มากขึ้น การใช้ n-gram ก็เพิ่มมากขึ้น ทำให้ผู้วิจัยและนักพัฒนาสามารถสร้างแบบจำลองที่ซับซ้อนมากขึ้น ซึ่งสามารถจัดการชุดข้อมูลที่มีขนาดใหญ่และซับซ้อนยิ่งขึ้นได้ (Nithyashree, 2024) ในช่วงทศวรรษ 1990 ความก้าวหน้าในเทคนิคการปรับให้เรียบกลายมาเป็นสิ่งสำคัญสำหรับการปรับปรุงประสิทธิภาพของโมเดล n-gram เทคนิคเหล่านี้ช่วยแก้ปัญหาที่เกิดจากข้อมูลเบาบาง เนื่องจากจำนวน n-gram ที่เป็นไปได้จะเพิ่มขึ้นแบบทวีคูณตามขนาดของ 'n' วิธีการปรับให้เรียบที่โดดเด่น ได้แก่ การปรับให้เรียบแบบบวกรหรือแบบลาปลาซ และการลดแบบกัวด-ทัวริง ซึ่งปรับการแจกแจงความน่าจะเป็นเพื่อจัดการกับ n-gram ที่มองไม่เห็นได้ดีขึ้น (DeepAI, n.d.; Dey, 2023) แม้จะมีประสิทธิภาพ แต่ n-gram ก็ยังมีข้อจำกัด เช่น หน้าต่างบริบทคงที่ที่อาจไม่สามารถจับความสัมพันธ์ระยะไกลภายในข้อความได้ ข้อจำกัดเหล่านี้กระตุ้นให้มีการวิจัยเพิ่มเติมเกี่ยวกับเทคนิคการสร้างแบบจำลองขั้นสูง เช่น แบบจำลองมาร์คอฟและเครือข่ายประสาทเทียม ซึ่งมุ่งหวังที่จะปรับปรุงความเข้าใจบริบทให้เกินขีดความสามารถของแบบจำลอง n-gram (Mljourney, 2024; Dey, 2023) อย่างไรก็ตาม n-grams ยังคงเป็นองค์ประกอบสำคัญของ NLP และยังคงมีอิทธิพลต่อการพัฒนาโมเดลภาษา รวมถึงแอปพลิเคชันร่วมสมัยในเทคโนโลยีการสร้างเนื้อหาที่ขับเคลื่อนด้วยเอไอ เช่น GPT (Research Graph, 2024)

การสร้างเนื้อหา N-grams มีต้นกำเนิดมาจากการพัฒนาในช่วงแรกของโมเดลภาษาสถิติ และมีบทบาทสำคัญในการปรับปรุงแอปพลิเคชันต่างๆ รวมถึงการจดจำเสียง การแปลด้วยเครื่อง และการสร้างข้อความอัตโนมัติ ความสามารถในการจับการอ้างอิงภายในช่วยให้สร้างข้อความที่มีความสอดคล้องและเกี่ยวข้องกับบริบท ทำให้ N-grams มีความสำคัญอย่างยิ่งในการพัฒนาเทคโนโลยีการสร้างเนื้อหาที่ขับเคลื่อนด้วยเอไอ เช่น ซีรีส์ Generative Pre-trained Transformer (GPT) (Lang, 2024; Gomed, 2023; Madala, 2023) อยู่ที่ยุคของการพัฒนาความสามารถในการคาดเดาข้อความและการสร้างแบบจำลองภาษา ช่วยให้ระบบสร้างเนื้อหาเลียนแบบการเขียนของมนุษย์ ค้นหาสำนวนทั่วไป และจำแนกความรู้สึก ส่งเสริมประสบการณ์ผู้ใช้ในแพลตฟอร์มดิจิทัล (Nguyen, 2020; Nithyashree, 2024; DeepAI, n.d.) อย่างไรก็ตามแม้จะมีข้อได้เปรียบ แต่ n-grams ยังต้องเผชิญกับความท้าทาย เช่น ข้อมูลไม่ทั่วถึงและหน้าต่างบริบทที่จำกัดความสามารถในการจับการอ้างอิงระยะไกลซึ่งจำเป็นสำหรับเนื้อหาที่ซับซ้อนมากขึ้น (Dey, 2023; Mljourney, 2024) ยิ่งไปกว่านั้นการผสมรวม n-gram เข้ากับเทคนิคการเรียนรู้ของเครื่องขั้นสูง รวมถึงโมเดลการเรียนรู้เชิงลึกยังคงมีศักยภาพในการแก้ไขข้อจำกัดบางประการเหล่านี้ได้ พร้อมทั้งเพิ่มความแม่นยำในการทำนายอีกด้วย (Research Graph, 2024; Jain, 2024) อย่างไรก็ตาม การพิจารณาทางจริยธรรมเกี่ยวกับอคติในข้อมูลการฝึกอบรมนั้นสร้างความกังวลอย่างมาก เนื่องจากโมเดลที่พึ่งพาชุดข้อมูลที่มีข้อบกพร่องอาจสร้างผลลัพธ์ที่เข้าใจผิดหรือเป็นอันตรายได้ ดังนั้น การทำให้แน่ใจว่า n-grams และแอปพลิเคชันในการสร้างเนื้อหาได้รับการพัฒนาด้วยข้อมูลตัวแทน และมาตรฐานทางจริยธรรมจึงมีความสำคัญต่อการส่งเสริมความไว้วางใจและประสิทธิภาพในเทคโนโลยีเอไอ (Kalotra, 2023; Team Experts, 2023) โดยสรุป n-grams เป็นองค์ประกอบสำคัญของการสร้างเนื้อหา NLP และเอไอ สมัยใหม่ ซึ่งให้ข้อมูลเชิงลึกอันมีค่าเกี่ยวกับโครงสร้างข้อความและการโต้ตอบของผู้ใช้ แม้ว่าบทบาทพื้นฐานของ n-grams จะนำไปสู่ความก้าวหน้าที่สำคัญ แต่การวิจัยอย่างต่อเนื่องและการพิจารณาทางจริยธรรมจะเป็นสิ่งสำคัญในการเพิ่มประสิทธิภาพการใช้งานและรับรองการใช้งานอย่างมีความรับผิดชอบในบริบทที่หลากหลาย (GeeksforGeeks, 2020; Mljourney, 2024)

แนวคิดและทฤษฎีด้านภาษาศาสตร์

ชนิดของคำ (Word Classes)

คำ ถือได้ว่าเป็นหน่วยที่เล็กที่สุดที่มีความหมายในตัวเอง สามารถเป็นได้ทั้งเสียงที่เปล่งออกมาหรือตัวอักษรที่สร้างขึ้น (ราชบัณฑิตยสถาน, ม.ป.ป.) คำสามารถจัดหมวดหมู่ได้ขึ้นอยู่กับลักษณะของการให้ความหมายในประโยค ซึ่งสามารถแบ่งออกได้สองประเภทหลัก ๆ ประเภทแรกเป็นกลุ่มของคำที่ให้รายละเอียดของประโยค ที่เรียกว่า Major Class, Form Class, Open Class หรือเรียกว่า Content Word (Cambridge Dictionary, n.d.) ประกอบไปด้วยชนิดของคำดังต่อไปนี้

Nouns คำนาม เป็นคำแทนคน สัตว์ หรือสิ่งของ เช่น Cat, Tree, Book

Verbs คำกริยา ใช้แสดงอาการต่าง ๆ ของประธานของประโยค เช่น Walk, Fly, Run

Adjectives คำคุณศัพท์ เป็นคำที่ใช้ทำหน้าที่ขยายคำนามหรือคำสรรพนาม เช่น Smart, Worst
 Adverbs คำวิเศษณ์ ทำหน้าที่ขยายความ เพื่อให้คำที่ต้องการขยายได้แก่ คำนาม คำกริยา และ
 คำวิเศษณ์กันเองมีความชัดเจนมากยิ่งขึ้น เช่น Lovely, Always, Tomorrow

และอีกประเภทหนึ่งเป็นกลุ่มของคำที่ทำให้โครงสร้างของประโยคนั้นถูกต้องตามหลักของ
 ไวยากรณ์ โดยมีชื่อเรียกว่า Minor Class, Structure Class, Close Class หรือ Function Word
 (Cambridge Dictionary, n.d.) ประกอบไปด้วยชนิดของคำดังต่อไปนี้

Prepositions คำบุพบท เป็นคำที่ใช้บอกตำแหน่ง หน้าที่ หรือแสดงความสัมพันธ์ระหว่างคำหรือ
 ประโยค โดยใช้วางนำหน้าคำนาม สรรพนาม เช่น in, on, at

Pronouns คำสรรพนาม เป็นคำที่สามารถใช้แทนคำนามเพื่อไม่ต้องกล่าวซ้ำ เช่น I, She, They

Determiners คำบ่งชี้ เป็นชื่อที่ใช้เรียกคำบ่ง หรือบอก หรือกำหนดนาม เช่น this, that, those

Conjunctions คำสันธาน หรือคำเชื่อม ใช้เชื่อมความเพื่อให้เกิดความสละสลวย สามารถเชื่อม
 ได้ทั้งคำ ข้อความ หรือประโยค เช่น and, or, but

Interjections คำอุทาน คำที่เปล่งออกมาเพื่อแสดงความรู้สึก หรือเสริมให้เนื้อความมีความ
 สมบูรณ์ยิ่งขึ้น เช่น Ouch!, Hooray!, Wow!

ตัวอย่างของคำทั้ง 9 ชนิดที่ยกมานี้ เป็นเพียงตัวอย่างของชนิดของคำที่พบได้บ่อย และเป็น
 ส่วนประกอบในการสร้างประโยคตั้งแต่ประโยคพื้นฐาน จนถึงประโยคที่ซับซ้อนขึ้น ผู้วิจัยวิเคราะห์
 ชนิดของคำโดยการปรับคำของ ChatGPT ว่าชนิดของคำที่ถูกปรับ เป็นชนิดเดียวกันกับคำก่อนหน้า
 หรือไม่ มีการเปลี่ยนแปลงเกิดขึ้นในชนิดของคำประเภทใดมากที่สุด ที่เป็นนัยสำคัญที่ทำให้คำ
 ความน่าจะเป็นเทียบเคียงความเป็นมนุษย์เกิดการเปลี่ยนแปลง

ความสัมพันธ์ทางความหมายของคำ (Lexical Relations)

Lexical Relation ความสัมพันธ์ทางความหมาย เป็นการศึกษาความหมายของคำศัพท์ในภาษา
 หนึ่ง ๆ ในแง่ของความหมายระหว่างคำ เพื่อให้เข้าใจรูปแบบความสัมพันธ์ที่เกิดขึ้นระหว่างคำศัพท์
 (Langeek, n.d.) ซึ่งสามารถแบ่งได้ตามประเภทดังต่อไปนี้

– Synonymy เป็นความสัมพันธ์ของคำศัพท์ที่มีความหมายเหมือนหรือคล้ายคลึงกันและ
 สามารถใช้ทดแทนกันได้ เช่น คำว่า terrible กับ awful ในประโยคต่อไปนี้ The party was terrible = the
 party was awful.

– Antonymy เป็นความสัมพันธ์ของคำศัพท์ที่มีความหมายตรงกันข้าม เช่น hot ≠ cold, never ≠
 always

– Meronymy เป็นความสัมพันธ์ของคำศัพท์ที่มีความหมายเป็นส่วนประกอบของอีกคำศัพท์
 เช่น Roof, Floor, Wall เป็นส่วนประกอบของ House

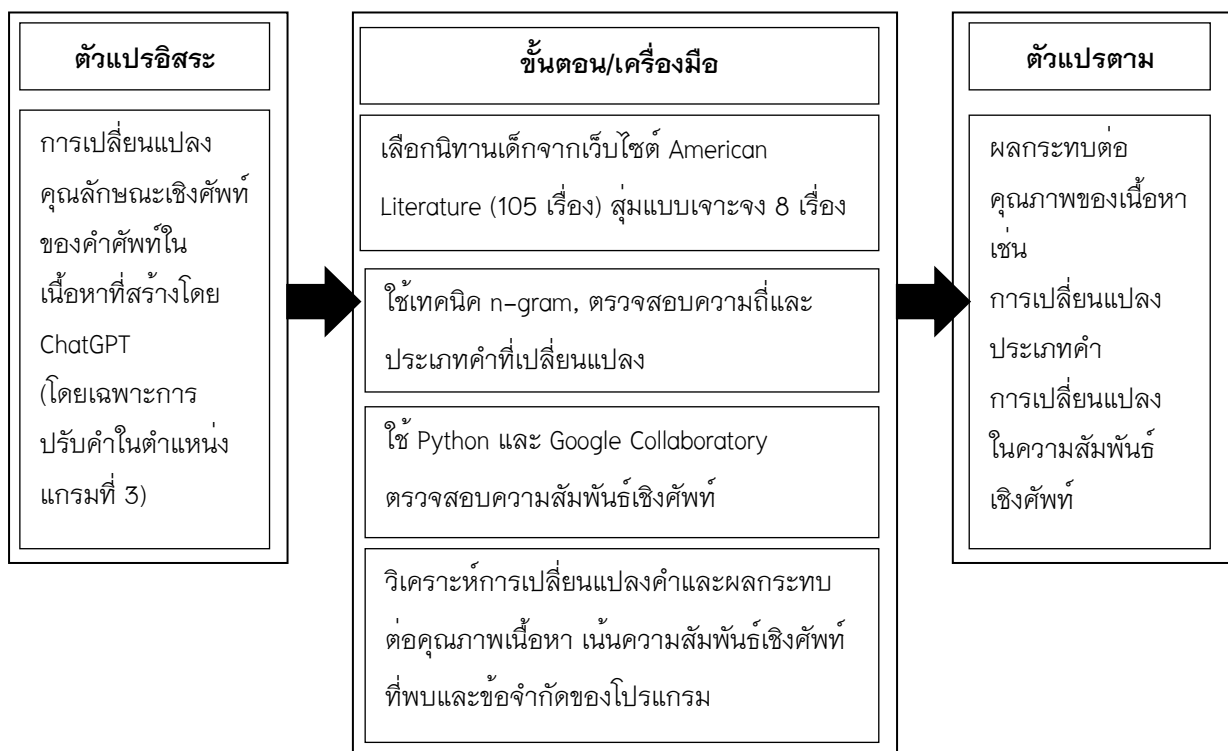
– Hypernymy เป็นความสัมพันธ์ของความหมายในแนวดิ่งที่คำศัพท์หนึ่งมีความหมายเป็น
 หน่วยใหญ่กว่าคำศัพท์อีกคำหนึ่ง เช่น Country เป็นคำศัพท์หน่วยใหญ่ของ District, Subdistrict, Village

- Hyponymy เป็นความสัมพันธ์ของความหมายในแนวดิ่งที่คำศัพท์หนึ่งมีความหมายเป็นหน่วยเล็กกว่าคำศัพท์อีกคำหนึ่ง เช่น Green, white and blue เป็นหน่วยย่อยของ Color

คำในแต่ละภาษามีความสัมพันธ์กันไม่ทางใดก็ทางหนึ่งขึ้นอยู่กับเกณฑ์ที่เลือกในการจัดคำหนึ่งคำสามารถมีความสัมพันธ์กับคำอื่น ๆ ได้หลายรูปแบบ ผู้วิจัยวิเคราะห์ความสัมพันธ์ของคำศัพท์ที่ถูกปรับโดย ChatGPT ว่ามีความสัมพันธ์ประเภทใดที่เกิดขึ้นในระหว่างก่อนปรับคำและหลังปรับคำโดย ChatGPT

กรอบแนวคิดการวิจัย

งานวิจัยนี้เป็นการวิจัยแบบผสม ผู้วิจัยกำหนดกรอบแนวคิดการวิจัยตามทฤษฎี ดังต่อไปนี้



ภาพที่ 1 กรอบแนวคิดวิจัย

ระเบียบวิธีวิจัย

งานวิจัยนี้ตรวจสอบผลกระทบของการเปลี่ยนแปลงคุณลักษณะของคำศัพท์ในเนื้อหาที่สร้างโดย ChatGPT ใช้การออกแบบการวิจัยแบบตัดขวาง แนวทางนี้ช่วยให้สามารถรวบรวมข้อมูลได้ในช่วงเวลาเดียวจากตัวอย่างที่เป็นตัวแทน ทำให้ผู้วิจัยสามารถวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะของคำศัพท์ต่างๆ และผลกระทบที่มีต่อคุณภาพของเนื้อหาที่สร้างขึ้นได้

1. การรวบรวมข้อมูล

ในงานวิจัยนี้ ผู้วิจัยได้เลือกนิทานสำหรับเด็กจากเว็บไซต์ AmericanLiterature ซึ่งมีทั้งหมด 105 เรื่อง (105,000 คำ) โดยใช้สูตร Taro Yamane คำนวณกลุ่มตัวอย่างที่มีค่าความคลาดเคลื่อน 0.02 ได้จำนวน 2,441.86 คำ จากนั้นแบ่งเป็นนิทานเรื่องละ 300 คำต่อเรื่อง ทำให้ได้จำนวนกลุ่มตัวอย่างทั้งหมด 8 เรื่อง โดยใช้วิธีการสุ่มแบบเจาะจงเพื่อให้มีเนื้อหา ตัวละคร และข้อคิดที่หลากหลาย ประกอบด้วยนิทาน 8 เรื่องดังนี้: The Three Little Pigs, Little Red Riding Hood, Jack and the Beanstalk, The Tortoise & the Hare, The Ugly Duckling, The Pied Piper of Hamelin, The Little Match Girl และ The Princess and the Pea

2. การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเป็นการวิเคราะห์ข้อมูลเชิงปริมาณ โดยใช้วิธีทางสถิติเพื่อระบุรูปแบบและความสัมพันธ์เชิงศัพท์

3. การวิเคราะห์ N-gram

ส่วนประกอบหลักของวิธีการนี้คือการนำการวิเคราะห์ n-gram มาใช้เพื่อประเมินคุณลักษณะทางศัพท์ของเนื้อหาที่สร้างโดย ChatGPT n-gram ซึ่งเป็นลำดับต่อเนื่องของรายการ n รายการจากข้อความจะถูกใช้เพื่อจับลักษณะโครงสร้างและบริบทของเนื้อหาที่สร้างขึ้น โดยงานวิจัยนี้ได้ทำการทดลองแล้วพบว่า การเปลี่ยนคำศัพท์ในแกรมที่ 3 จะการเปลี่ยนแปลงที่ส่งผลต่อเนื้อหา ซึ่งส่งผลกระทบต่อการเปลี่ยนแปลงคำศัพท์ในเนื้อหาที่สร้างโดย ChatGPT ซึ่งช่วยให้เข้าใจได้ดียิ่งขึ้นว่าการเปลี่ยนแปลงเหล่านี้สามารถเพิ่มประสิทธิภาพของข้อความที่สร้างโดยเอไอ ในแอปพลิเคชันต่างๆ ได้

ผลการวิจัย

วัตถุประสงค์ที่ 1. การเปลี่ยนแปลงคำศัพท์ที่ปรากฏในเนื้อหาที่สร้างโดย ChatGPT ผ่านการวิเคราะห์เชิง n-gram เกิดจากการแทนที่คำศัพท์ของ ChatGPT ซึ่งผลการแทนที่คำทำให้ชนิดของคำที่ปรากฏเปลี่ยนไป การเปลี่ยนแปลงคำศัพท์ในเนื้อหาโดย ChatGPT ผ่านการวิเคราะห์ n-gram

การศึกษานี้มุ่งวิเคราะห์ผลการเปลี่ยนแปลงคำศัพท์ที่ปรากฏในเนื้อหาที่สร้างโดย ChatGPT ผ่านการแทนที่คำในตำแหน่ง n-grams ที่กำหนด (ตำแหน่ง 3, 4, 5, 6, และ 7) โดยเริ่มนับใหม่ทุกครั้งตามตำแหน่งที่กำหนด พบว่าตำแหน่งที่ 3 มีความถี่และผลลัพธ์ที่เหมาะสมที่สุด เนื่องจากคำในตำแหน่งนี้มีการเปลี่ยนแปลงหลากหลาย ส่งผลให้ประโยคมีความถูกต้องตามไวยากรณ์และเหมาะสมกับบริบทมากขึ้น

จากการทดลองโดยใช้ ChatGPT และ Python วิเคราะห์ พบการเปลี่ยนแปลงในประเภทคำ เช่น คำนาม (N), คำกริยา (V), คำคุณศัพท์ (ADJ), คำวิเศษณ์ (ADV), และคำประเภทอื่น ๆ การปรับแต่งดังกล่าวส่งผลต่อโครงสร้างและเนื้อหาอย่างมีนัยสำคัญ โดยเฉพาะตำแหน่ง 3-grams ซึ่งแสดงถึงแนวโน้มการเปลี่ยนแปลงที่ชัดเจนในชนิดและความถี่ของคำที่ปรากฏ คำนาม (Nouns): คำนามเป็น

ประเภทคำที่พบบ่อยที่สุดในงานเขียน โดยในการปรับคำครั้งที่ 1 ปรากฏ 633 ครั้ง และเพิ่มขึ้นในการปรับคำครั้งที่ 2 เป็น 693 ครั้ง การใช้คำนามเพิ่มขึ้นในเรื่อง Jack and the Beanstalk (No. 1 revised) อย่างเด่นชัด จาก 27 ครั้ง เป็น 63 ครั้ง การเพิ่มขึ้นนี้แสดงถึงการเสริมรายละเอียดของตัวละครและวัตถุในเรื่อง การเพิ่มคำนามสะท้อนถึงความพยายามในการทำให้งานเขียนมีความสมจริงและมีภาพจินตนาการที่ชัดเจนยิ่งขึ้น

คำกริยา (Verbs): คำกริยาที่บอกการกระทำหรือสถานะของตัวละคร ปรากฏ 412 ครั้ง ในการปรับคำครั้งที่ 1 และเพิ่มขึ้นเล็กน้อยเป็น 413 ครั้ง ในการปรับคำครั้งที่ 2 การเปลี่ยนแปลงของคำกริยาไม่เด่นชัด แต่ในบางเรื่อง เช่น The Princess and the Pea (No. 2 revised) คำกริยาเพิ่มขึ้นจาก 21 ครั้ง เป็น 27 ครั้ง แสดงถึงการปรับปรุงเนื้อหาเพื่อเพิ่มความชัดเจนของการกระทำ คำกริยายังคงเป็นเครื่องมือสำคัญในการขับเคลื่อนเนื้อหางานเขียน แม้ว่าการเปลี่ยนแปลงจะมีขนาดเล็ก

คำคุณศัพท์ (Adjectives): คำคุณศัพท์ปรากฏ 224 ครั้ง ในการปรับคำครั้งที่ 1 และเพิ่มขึ้นเป็น 245 ครั้ง ในการปรับคำครั้งที่ 2 ใน Jack and the Beanstalk (No. 1 revised) คำคุณศัพท์เพิ่มขึ้นจาก 9 ครั้ง เป็น 27 ครั้ง ซึ่งให้เห็นถึงการเพิ่มรายละเอียดของตัวละครและวัตถุ คำคุณศัพท์ช่วยเสริมให้เนื้อหางานเขียนมีความสมบูรณ์และน่าสนใจมากขึ้น โดยเฉพาะในเรื่องที่มีการบรรยายเชิงลึก

คำวิเศษณ์ (Adverbs): คำวิเศษณ์ที่ขยายคำกริยา ปรากฏ 106 ครั้ง ในการปรับคำครั้งที่ 1 และเพิ่มขึ้นเป็น 115 ครั้ง ในการปรับคำครั้งที่ 2 ในเรื่อง Little Red Riding Hood (No. 1 revised) คำวิเศษณ์เพิ่มขึ้นจาก 5 ครั้ง เป็น 11 ครั้ง แสดงถึงการเพิ่มรายละเอียดของการกระทำ การใช้คำวิเศษณ์สะท้อนถึงความพยายามในการเพิ่มมิติให้การบรรยายลักษณะของการกระทำในเรื่อง

คำสรรพนาม (Pronouns): คำสรรพนามช่วยหลีกเลี่ยงการใช้คำนามซ้ำ โดยปรากฏ 232 ครั้ง ในการปรับคำครั้งที่ 1 และลดลงเป็น 229 ครั้ง ในการปรับคำครั้งที่ 2 ใน Little Red Riding Hood (No. 3 revised) ไม่มีการใช้คำสรรพนามเลย ซึ่งอาจสะท้อนถึงความพยายามในการระบุดีรายละเอียดด้วยคำนามแทน การลดคำสรรพนามในบางกรณีอาจเพิ่มความชัดเจนของตัวละครและเนื้อหา

คำสันธาน (Conjunctions): คำสันธานปรากฏ 145 ครั้ง ในการปรับคำครั้งที่ 1 และลดลงเป็น 108 ครั้ง ในการปรับคำครั้งที่ 2 การลดคำสันธานเกิดขึ้นในหลายเรื่อง เช่น The Tortoise and the Hare (No. 3 revised) ซึ่งลดลงจาก 8 ครั้ง เป็น 6 ครั้ง การลดคำสันธานสะท้อนถึงความพยายามในการทำให้อรรถกถากระชับและลดความซ้ำซ้อน

คำบ่งชี้ (Determiners): คำบ่งชี้ช่วยเพิ่มความเจาะจงในเนื้อหา โดยปรากฏ 294 ครั้ง ในการปรับคำครั้งที่ 1 และลดลงเป็น 250 ครั้ง ในการปรับคำครั้งที่ 2 ในเรื่อง Jack and the Beanstalk คำบ่งชี้ลดลงอย่างชัดเจนในทุกเวอร์ชัน การลดคำบ่งชี้ช่วยลดความซ้ำซ้อนและทำให้อรรถกถากระชับมากขึ้น

ตัวเลข (Numerals): ตัวเลขปรากฏ 4 ครั้ง ในการปรับคำครั้งที่ 1 และลดลงเหลือ 2 ครั้ง ในการปรับคำครั้งที่ 2 การลดลงของตัวเลขเกิดขึ้นในเรื่องที่ไม่จำเป็นต้องอ้างอิงจำนวน ตัวเลขถูกใช้เมื่อจำเป็นต่อเนื้อหาเท่านั้น

คำอุทาน (Interjections): คำอุทานปรากฏ 1 ครั้ง ในการปรับคำครั้งที่ 1 และไม่มีการใช้ในการปรับคำครั้งที่ 2 การไม่มีคำอุทานในการปรับคำครั้งที่ 2 แสดงถึงการเล่าเรื่องที่เน้นความเป็นกลาง การเล่าเรื่องที่ไม่มีคำอุทานช่วยคงอารมณ์ของเรื่องให้นิ่งและไม่กระตุ้นอารมณ์มากเกินไป

การเปลี่ยนแปลงคำในตำแหน่งที่ 3 โดย ChatGPT แสดงให้เห็นความแตกต่างในชนิดและความถี่ของคำ เช่น คำนาม 693 ครั้ง คำกริยา 413 ครั้ง และคำบ่งชี้ 270 ครั้ง แม้ว่าคำนามจะยังคงเป็นคำที่พบบ่อยที่สุด แต่ความถี่ของคำแต่ละประเภทมีการเพิ่มขึ้นหรือลดลง รวมถึงการเปลี่ยนแปลงชนิดของคำบางคำ การแก้ไขคำที่เหมาะสมในตำแหน่งเหล่านี้ช่วยเพิ่มประสิทธิภาพและลดเวลาสร้างงานได้อย่างมีนัยสำคัญ

ตำแหน่งแกรมที่ 3 แสดงความเปลี่ยนแปลงเด่นชัดจากการปรับคำในทุก ๆ คำที่ 3 ส่งผลให้ค่าความน่าจะเป็นเปลี่ยนแปลงในแนวโน้มที่สูงขึ้นมากกว่าตำแหน่งอื่น ๆ โดยเฉพาะการเปลี่ยนแปลงคำนามซึ่งเกิดบ่อยที่สุด แสดงให้เห็นว่าความถี่ของการปรับคำส่งผลต่อค่าความน่าจะเป็น อย่างไรก็ตาม ผลลัพธ์ยังขึ้นอยู่กับประเภทของคำที่ถูกปรับ หากทุกคำในตำแหน่งที่ 3 เป็นคำสันธานหรือคำอุทาน อาจส่งผลต่อค่าความน่าจะเป็นแตกต่างไปจากที่พบในงานวิจัยนี้

วัตถุประสงค์ที่ 2. จากการวิเคราะห์คุณลักษณะเชิงศัพท์ของงานเขียนที่ถูกปรับคำในตำแหน่งแกรมที่ 3 โดย ChatGPT ผู้วิจัยได้ค้นพบผลลัพธ์การเปลี่ยนแปลงคุณลักษณะเชิงศัพท์ที่ปรากฏ 8 รูปแบบ ซึ่งเป็นรูปแบบที่สามารถพบได้จากการนำเอางานเขียนไปวิเคราะห์โดยใช้ฐานข้อมูลภาษาจาก ไพธอน ผ่าน Google Collaboratory ซึ่งได้ผลลัพธ์ดังต่อไปนี้

1) Antonym ความสัมพันธ์แบบคำคู่ตรงข้าม คุณลักษณะเชิงศัพท์ประเภทนี้ หากมองตามคำสั่งที่ใช้สั่งให้ ChatGPT ปรับคำแล้วนั้น เป็นความสัมพันธ์ที่ไม่น่าจะเกิดขึ้น เนื่องจากการหาคำที่มีคุณลักษณะคล้ายกัน ความหมายคล้ายกัน หน้าที่คล้ายกัน และสามารถใช้ทดแทนกันได้ ไม่ควรเป็นคำที่เป็นคู่ตรงข้ามกัน แต่ความสัมพันธ์ในรูปแบบนี้ เกิดขึ้น 7 ครั้งจากจำนวนการเกิดทั้งหมด ซึ่งถือได้ว่าเป็นจำนวนการเกิดเพียงเล็กน้อย แต่อย่างไรก็ตามเป็นรูปแบบความสัมพันธ์ที่อาจเกิดจากความผิดพลาดของคำสั่งและตัวโปรแกรมเอง

2) Co-hyponym หรือแบบเต็มเรียกว่า Co-hyponym: shares common hypernyms (X) เป็นความสัมพันธ์ที่ไม่ได้เป็น Hyponym ซึ่งกันและกัน แต่เป็นความสัมพันธ์ที่มี Hypernym เป็นคำเดียวกัน จึงทำให้ความสัมพันธ์ในรูปแบบนี้ การใช้คำศัพท์ร่วมกันเป็นการเกิดขึ้นของคู่คำศัพท์ที่ไม่ได้มีความสัมพันธ์ซึ่งกันและกันในระดับที่ใกล้ชิดจนสามารถทำให้เกิดความสัมพันธ์หนึ่ง ๆ ได้ โปรแกรมจึงจัดหมวดหมู่เพิ่มเติมขึ้นมาเพื่อรองรับการเกิดขึ้นของความสัมพันธ์รูปแบบนี้ ซึ่งเกิดขึ้นถึง 103 ครั้ง และสามารถนับได้ว่าเป็นการเกิดขึ้นในอันดับที่ 3 จากการเกิดขึ้นทั้งหมด แสดงให้เห็นว่า คู่ของคำศัพท์บางคู่ที่ไม่ได้มี

ความสัมพันธ์กันโดยตรง แต่มีความสัมพันธ์ร่วมกันเป็นความสัมพันธ์ที่เกิดขึ้นมากจากการวิเคราะห์ของโปรแกรม

3) Co-hypernym หรือแบบเต็มเรียกว่า Co-hypernym: shares common hyponyms (X) เป็นความสัมพันธ์ที่ไม่ได้เป็น Hypernym ซึ่งกันและกัน แต่เป็นความสัมพันธ์ที่มี Hyponym เป็นคำเดียวกัน จึงทำให้ความสัมพันธ์ในรูปนี้ ซึ่งความสัมพันธ์ในรูปแบบนี้เป็นความสัมพันธ์ชั่วคราวข้ามของความสัมพันธ์แบบ Co-hyponym แต่อย่างไรก็ตาม ความสัมพันธ์รูปแบบนี้โปรแกรมไม่ได้วิเคราะห์ว่าเป็นความสัมพันธ์ชั่วคราวอย่างชัดเจน ยังมีการเกิดขึ้นของตัวความสัมพันธ์นี้เองถึง 3 ครั้ง ซึ่งถึงแม้จะมีค่าการเกิดของความสัมพันธ์ในระดับที่น้อยมาก แต่จากผลการวิเคราะห์ของโปรแกรม ยังคงให้ค่าว่าความสัมพันธ์ที่ไม่ได้สัมพันธ์กันโดยตรงของคู่ศัพท์แต่ใช้ Hyponym ร่วมกันเป็นอีกหนึ่งความสัมพันธ์ที่สามารถแบ่งได้

4) Hyponym ความสัมพันธ์แบบคำหนึ่งเล็กกว่าอีกคำหนึ่ง การวิเคราะห์คู่คำศัพท์เพื่อหาความสัมพันธ์ทางคุณลักษณะเชิงศัพท์ที่ทำหลักมีค่าเล็กกว่าอีกคำที่ถูกหามาทดแทน เป็นการใช้ความสามารถของ ChatGPT ในการสรรหาคำศัพท์ที่อยู่ในวงศ์เดียวกัน เป็นการทำให้ความหมายของคำกว้างขึ้น ความสัมพันธ์ในรูปแบบนี้เกิดทั้งสิ้น 40 ครั้ง ซึ่งการทำให้ความหมายของคำกว้างขึ้นนี้เองเป็นการสรรหาคำมาแทนที่เพื่อทำให้บริบทของเนื้อหา ยังคงอยู่ ไม่ถูกเปลี่ยนไปจากเดิมมาก

5) Hypernym ความสัมพันธ์แบบคำหนึ่งใหญ่กว่าอีกคำหนึ่ง เช่นเดียวกับ Hyponym แต่ Hypernym เป็นการวิเคราะห์ที่มองว่าคำหลักเป็นคำที่มีค่าใหญ่กว่าอีกคำที่ถูกหามาทดแทน การเกิดความสัมพันธ์แบบนี้เป็นการทำให้ความหมายของคำแคบลง ทำให้สามารถแสดงเนื้อหาที่เฉพาะเจาะจงกว่าเดิมได้ ซึ่งมีอัตราการเกิดอยู่ที่ 70 ครั้ง สะท้อนให้เห็นว่า ChatGPT พยายามจำกัดความหมายของคำให้แคบลง เพื่อที่จะให้ความหมายและบริบทต่าง ๆ ของเรื่องราวมีความชัดเจนมากยิ่งขึ้น

6) Synonym ความสัมพันธ์ของคู่คำที่เหมือนกัน สามารถใช้ทดแทนกันได้เหมาะสม หากไม่นับความสัมพันธ์แบบ Word to Vector ความสัมพันธ์นี้ถือได้ว่าเป็นความสัมพันธ์ที่มีอัตราการเกิดสูงที่สุด ซึ่งอัตราการเกิดอยู่ที่ 473 ครั้ง ซึ่งเรียงอันดับที่ 2 รองจากความสัมพันธ์ที่กล่าวมาข้างต้น Synonym เป็นคู่ของความสัมพันธ์ที่ควรเกิดขึ้นมากที่สุดหากมองตามคำสั่งที่ผู้วิจัยใช้สั่งให้ ChatGPT สรรหาคำมาเพื่อทำการปรับคำ เนื่องจากความสัมพันธ์ประเภทนี้เป็นความสัมพันธ์ที่เรียกได้ว่าแทบจะใช้คำศัพท์คำใหม่แทนคำศัพท์คำเดิมได้โดยไม่มีอะไรผิดเพี้ยนไปจากเดิม อัตราการเกิดที่สูงของความสัมพันธ์นี้แสดงให้เห็นถึงศักยภาพของ ChatGPT และความสามารถในการวิเคราะห์ประเภทความสัมพันธ์ของฐานข้อมูลจาก ไพธอน ว่ามีประสิทธิภาพและมีความน่าเชื่อถือ

7) Meronym ความสัมพันธ์แบบคำหนึ่งเป็นส่วนประกอบของคำหนึ่ง อีกหนึ่งความสัมพันธ์ขนาดเล็กที่พบได้ค่อนข้างน้อยในการวิเคราะห์ครั้งนี้ ด้วยอัตราการเกิดเพียง 11 ครั้ง แต่ไม่ได้หมายความว่าความสัมพันธ์นี้ไม่สำคัญ เพราะการวิเคราะห์ของโปรแกรมไม่ได้มองว่าความสัมพันธ์ของคู่คำศัพท์เป็นรูปแบบอื่น แต่เป็นรูปแบบของคำที่เป็นส่วนประกอบของคำที่ถูกนำมาแทนที่

ความสัมพันธ์ชนิดนี้เหมือนจะทำให้ความหมายของคำกว้างขึ้น หากแต่จริง ๆ แล้ว เป็นการแทนที่ของคำใหม่ที่เป็นคำหลักที่คำนี้ไปเป็นส่วนประกอบ จึงไม่สามารถจำแนกเป็นคำประเภทอื่นได้

8) Holonym ความสัมพันธ์แบบคำหนึ่งเป็นส่วนหลักของคำหนึ่ง คำคู่ตรงข้ามของ Meronym คำประเภทนี้คำหลักจะเป็นคำหลักที่ ChatGPT ได้หาคำที่เป็นส่วนประกอบมาทดแทน หากมองภาพอาจจะมองว่าเป็นการทำให้ความหมายแคบลง แต่ความเป็นจริงแล้ว คำที่เป็นส่วนประกอบจะเป็นส่วนใดก็ได้ของคำหลักที่ใช้ในการวิเคราะห์

การเปลี่ยนแปลงในรูปแบบอื่นที่ปรากฏ

การเปลี่ยนแปลงในรูปแบบนี้ทางโปรแกรมใช้ชื่อว่า Related through word embeddings similarity เป็นการเปลี่ยนแปลงที่เกิดขึ้นจากการทำงานของโปรแกรมคอมพิวเตอร์ที่คำนวณระยะความเหมือนของคำศัพท์ผ่านรูปแบบที่เรียกว่า Word to Vector โดยอาศัยการแปลงค่าคำเป็นตัวเลขจาก Word Embedding

Word Embedding เป็นการแปลงค่าของคำให้กลายเป็นตัวเลข ซึ่งผลลัพธ์จะอยู่ในรูปแบบของ Vector ซึ่งจะนำไปสู่การทำงานของ Word to Vector ซึ่งเป็นการนำโมเดลของ Word Embedding หลาย ๆ ชิ้นมาประกอบกันเพื่อทำการฝึกในการหาค่าความใกล้เคียงของคำศัพท์ในแต่ละคำที่ถูกแปลงให้เป็น Vector ในที่นี้มีการใช้งานหลัก ๆ อยู่ 2 รูปแบบ คือ CBOW และ Skip-gram ซึ่งประเภทแรก CBOW เป็นการหาคำจากบริบทเหมือนหรือคำถัดไปที่สามารถเกิดขึ้นได้มาเป็นตัวกำหนดค่าของคำที่เป็น Vector ส่วนอีกประเภทเป็นการหาคำโดยให้ Vector เป็นตัวกลางในการกระจายค่าเพื่อหาคำที่มีค่าความเหมือนที่ใกล้เคียงที่สุด

หลักการนี้เองเป็นสาเหตุที่ทำให้เกิดการเลือกคำมาปรับของ ChatGPT ที่อยู่ในรูปแบบของ Related through word embeddings similarity ซึ่งหากไม่นับคู่ของคำที่สามารถกำหนดคุณลักษณะเชิงศัพท์ได้ การเกิดขึ้นของรูปแบบนี้จึงถือได้ว่าเป็นรูปแบบที่เกิดขึ้นในระดับสูงที่สุดถึง 584 ครั้ง ซึ่งสะท้อนให้เห็นว่า การทำงานของ ChatGPT จริง ๆ แล้วนั้น ไม่ได้อาศัยเพียงการทำงานที่เกิดจากคุณลักษณะเชิงศัพท์ หรือลักษณะทางภาษาอื่น ๆ เพียงอย่างเดียว แต่เนื่องด้วย ChatGPT เป็นหนึ่งในโปรแกรมแชทบอทที่มีความสามารถและศักยภาพที่สามารถทำงานได้อย่างหลากหลายนี้เอง จึงเกิดการนำหลักการทางคณิตศาสตร์และคอมพิวเตอร์เข้ามาเป็นตัวช่วยในการฝึกฝนเพื่อให้ ChatGPT เรียนรู้และสามารถตอบสนองความต้องการของผู้ใช้งานมากที่สุด

การเกิดรูปแบบ Related through word embeddings ในงานวิจัยนี้ไม่ได้ถือเป็นข้อผิดพลาดของโปรแกรม แต่สะท้อนถึงกระบวนการทำงานที่ช่วยให้ ChatGPT ทำงานได้เต็มประสิทธิภาพ การวิเคราะห์ด้วยฐานข้อมูลภาษา Python ผ่าน Google Collaboratory ช่วยยืนยันว่ารูปแบบนี้เกิดจากการประมวลผลของโปรแกรมที่ผสานหลักการต่าง ๆ เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด ผู้วิจัยจึงให้ความสำคัญกับรูปแบบนี้ในฐานะส่วนหนึ่งของการวิเคราะห์คุณลักษณะเชิงศัพท์ในตำแหน่งแกรมที่ 3 ซึ่งแสดงถึงประสิทธิภาพของ ChatGPT ในการปรับคำ

ความไม่เข้าคู่จากการวิเคราะห์คุณลักษณะเชิงศัพท์ของโปรแกรม

จากผลการวิเคราะห์เพื่อหาคุณลักษณะเชิงศัพท์ที่เกิดขึ้นระหว่างคู่ของคำศัพท์ที่ถูกปรับโดย ChatGPT นั้น ถึงแม้จะใช้เงื่อนไขต่าง ๆ หรือคุณสมบัติทางคณิตศาสตร์เพื่อช่วยในการคำนวณหาค่าคุณลักษณะเชิงศัพท์ของคู่ศัพท์แล้วนั้น ปฏิเสธไม่ได้เลยว่าอีกหนึ่งผลลัพธ์ที่จะต้องแสดงออกมานั้นก็คือ ไม่มีความสัมพันธ์ทางคุณลักษณะเชิงศัพท์ หรือที่ ไพธอน เรียกว่า No lexical relation found ซึ่งในส่วนนี้สามารถแบ่งการค้นพบออกได้ 2 ลักษณะ ตามประเภทของการเกิด ดังต่อไปนี้

1) No lexical relation found แบบที่คู่ของคำศัพท์ที่ถูกปรับไม่มีความสัมพันธ์ของคุณลักษณะเชิงศัพท์ต่อกัน ผลลัพธ์ในข้อนี้เป็นผลลัพธ์ทั่วไปที่สามารถเกิดขึ้นได้ คู่ของคำในแต่ละคู่ไม่จำเป็นต้องมีความสัมพันธ์ใด ๆ ซึ่งกันและกันก็ได้ แต่ตามหลักแล้วควรมีความสัมพันธ์กันในรูปแบบใดรูปแบบหนึ่ง เนื่องจากการใช้พรมต์ในการสั่งการให้ ChatGPT เลือกคำที่มีความสัมพันธ์ใกล้เคียงกันมาปรับ แต่อย่างไรก็ตามความผิดพลาดที่เกิดขึ้นอาจเกิดขึ้นตั้งแต่การใช้พรมต์ที่ไม่ชัดเจนพอจึงทำให้ ChatGPT ทำงานได้ไม่ตรงตามความต้องการอย่างเต็มประสิทธิภาพ การเกิดขึ้นของผลลัพธ์ในลักษณะนี้จึงถือว่าเป็นสิ่งที่ยอมรับได้ และสามารถเกิดขึ้นได้ทั่วไป

2) No lexical relation found แบบที่คู่ของคำศัพท์ที่ถูกปรับมีความสัมพันธ์ของคุณลักษณะต่อกัน รูปแบบของผลลัพธ์นี้ เป็นรูปแบบของผลลัพธ์ที่ไม่ควรเกิดขึ้น เนื่องจากหากคู่ของคำมีความสัมพันธ์ในคุณลักษณะเชิงศัพท์รูปแบบใดรูปแบบหนึ่งแล้วนั้น ย่อมควรถูกจัดหมวดหมู่ให้ถูกต้องตามรูปแบบความสัมพันธ์ที่เกิดขึ้น แต่อย่างไรก็ตาม ผลลัพธ์นี้อาจเกิดขึ้นได้จากความผิดพลาดของคำสั่งของผู้วิจัย และฐานข้อมูลภาษาที่มีไม่มากพอ จึงทำให้ผลลัพธ์ไม่ถูกนับว่าเป็นผลลัพธ์ที่มีความสัมพันธ์ซึ่งกันและกัน

ข้อจำกัดของการวิเคราะห์คุณลักษณะเชิงศัพท์ของคู่ของคำศัพท์

ผลลัพธ์ "No lexical relation found" ที่พบว่าคู่คำมีความสัมพันธ์เชิงคุณลักษณะ แต่ระบบวิเคราะห์หาไม่มีความสัมพันธ์ สะท้อนข้อจำกัดของการใช้ Python และ Google Collaboratory ในการวิเคราะห์คำศัพท์ โดยพบว่าผลลัพธ์นี้เกิดขึ้นถึง 1,053 ครั้ง คิดเป็นครึ่งหนึ่งของคู่คำทั้งหมด ผู้วิจัยตรวจสอบด้วยตนเองพบว่าหลายคู่คำยังมีความสัมพันธ์เชิงศัพท์ แต่ไม่ได้รับการจำแนกอย่างถูกต้อง

ข้อจำกัดนี้อาจเกิดจากคำสั่งที่ไม่ชัดเจนในการให้ ChatGPT เลือกคำที่สัมพันธ์กัน หรือจากฐานข้อมูลที่ไม่เพียงพอ อย่างไรก็ตาม ผู้วิจัยยังคงยึดผลลัพธ์จากการวิเคราะห์โปรแกรมโดยตรง เพื่อคงความเป็นกลางและไม่แทรกแซงด้วยความคิดเห็นส่วนตัว ทั้งนี้ จำเป็นต้องพัฒนาศักยภาพการวิเคราะห์ของโปรแกรมให้แม่นยำยิ่งขึ้น

คุณลักษณะเชิงศัพท์กับการแปลผลลัพธ์จากการเกิด

การวิเคราะห์หาคุณลักษณะเชิงศัพท์จากคู่ของคำศัพท์ที่ถูกปรับในตำแหน่งแกรมที่ 3 โดย ChatGPT นั้น สามารถแสดงออกมาให้เห็นได้หลายลักษณะความสัมพันธ์แตกต่างกันไปตามคุณสมบัติของคู่คำศัพท์นั้น ๆ แต่ยังมีบางส่วนที่ยังเป็นข้อกังขาอยู่เช่นเดียวกันว่าผลจากการวิเคราะห์ที่ออกมา

เป็นลักษณะแบบนี้ขึ้น เกิดจากปัจจัยใดกันแน่ ในแต่ละรูปแบบการปรากฏมีความแตกต่างกันน้อย สลับกันไป แต่ผลการวิเคราะห์ได้สร้างข้อสรุปที่สามารถแบ่งออกเป็นประเด็นได้ดังต่อไปนี้

- 1) No lexical relation found เป็นความสัมพันธ์ที่เกิดขึ้นมากที่สุด
- 2) หากไม่นับข้อ 1) No lexical relation ความสัมพันธ์แบบ Related through word embeddings similarity จะเป็นความสัมพันธ์ที่เกิดขึ้นมากที่สุดเป็นอันดับที่ 1 โดยมีอัตราการเกิดอยู่ที่ 584 ครั้ง
- 3) Synonym เป็นความสัมพันธ์ที่เกิดขึ้นมากถึง 473 ครั้ง ซึ่งอยู่ที่อันดับที่ 2 ไม่ต่างจาก สมมติฐานที่ผู้วิจัยได้คาดไว้
- 4) Co-hyponym เป็นความสัมพันธ์ที่เกิดขึ้นในระดับมากกว่าที่สมมติฐานที่ผู้วิจัยได้คาดไว้ถึง 103 ครั้ง ซึ่งถือได้ว่าเป็นอันดับ 3
- 5) มีคุณลักษณะเพิ่มเติมอื่น ๆ ที่เกิดขึ้นมากกว่าที่ผู้วิจัยได้ทำการทดลองนำร่องไว้ถึง 6 คุณลักษณะความสัมพันธ์

อย่างไรก็ตาม อัตราในการเกิดเป็นผลมาจากปัจจัยหลายส่วน ไม่ว่าจะเป็นคำสั่งของผู้วิจัย ความสามารถของ ChatGPT หรือแม้กระทั่งความสามารถในการวิเคราะห์จากฐานข้อมูลของ ไพธอน ซึ่งส่งผลทำให้คุณลักษณะความสัมพันธ์เหล่านี้มีอัตราและอันดับการเกิดที่ไม่เท่ากัน ซึ่งทำให้ได้มาซึ่งผลลัพธ์ตามข้างต้นว่าคุณลักษณะเชิงศัพท์ใดบ้างที่ส่งผลต่อค่าความน่าจะเป็นเทียบเคียง ความเป็นมนุษย์ของงานเขียนที่สร้างขึ้นโดย ChatGPT และถูกปรับคำโดย ChatGPT ในตำแหน่งแกรม ที่ 3

อภิปรายผลการวิจัย

การวิจัยนี้แบ่งการวิเคราะห์ข้อมูลเป็น 2 ส่วน คือการเปลี่ยนแปลงคำศัพท์ในเนื้อหาที่สร้างโดย ChatGPT ผ่านการวิเคราะห์ n-gram โดยใช้ตำแหน่งแกรมที่ 3 และการวิจัยการวิเคราะห์คุณลักษณะเชิงศัพท์ของงานเขียนที่ถูกปรับโดย ChatGPT ในตำแหน่งแกรมที่ 3

การเปลี่ยนแปลงคำศัพท์ในเนื้อหาที่สร้างโดย ChatGPT ผ่านการวิเคราะห์ n-gram โดยใช้ตำแหน่งแกรมที่ 3 ในการปรับเปลี่ยนคำศัพท์ ผลลัพธ์แสดงให้เห็นถึงลักษณะและความถี่ของคำ ประเภทต่าง ๆ ในเนื้อหา 3 รุ่น (GEN1, GEN2, GEN3) ซึ่งสะท้อนถึงการเปลี่ยนแปลงโครงสร้างและรูปแบบการเล่าเรื่องของเอไอ โดยสรุปได้ดังนี้

คำนาม (Nouns) เป็นประเภทคำที่พบบ่อยที่สุด (633 ครั้ง) ใช้ในการระบุตัวละคร สถานที่ และ วัตถุ การ GEN ครั้งที่ 3 มีการเพิ่มจำนวนคำนามในเรื่องบางเรื่อง เช่น *Jack and the Beanstalk* และ *The Princess and the Pea* แสดงถึงการเพิ่มรายละเอียดในเนื้อหาเพื่อความสมจริง

คำกริยา (Verbs) ปรากฏ 412 ครั้ง แต่มีแนวโน้มลดลงในการ GEN ครั้งที่ 3 เช่น ใน *The Three Little Pigs* และ *Little Red Riding Hood* การลดลงนี้สะท้อนถึงการเปลี่ยนโฟกัสจากการเล่าเรื่องเชิงการกระทำไปสู่การบรรยาย

คำคุณศัพท์ (Adjectives) พบ 224 ครั้ง โดย GEN1 ของ *Little Red Riding Hood* มีการใช้คำคุณศัพท์สูงสุด การ GEN ครั้งที่ 3 มีการลดจำนวนคำคุณศัพท์ลง ซึ่งอาจเป็นการปรับเนื้อหาให้กระชับและตรงประเด็นมากขึ้น

คำวิเศษณ์ (Adverbs) ปรากฏ 106 ครั้ง การใช้คำวิเศษณ์มีจำนวนจำกัด แสดงถึงการเล่าเรื่องแบบตรงไปตรงมา GEN3 มีการใช้คำวิเศษณ์น้อยลง ซึ่งสะท้อนถึงความพยายามสร้างเนื้อหาที่เข้าใจง่าย

คำสรรพนาม (Pronouns) ปรากฏ 232 ครั้ง ช่วยลดการใช้คำนามซ้ำและทำให้การเล่าเรื่องสั้นไหล มีการลดลงเล็กน้อยใน GEN3 สะท้อนถึงการเปลี่ยนแปลงมุมมองการเล่าเรื่อง

คำสันธาน (Conjunctions) ปรากฏ 145 ครั้ง GEN1 มีการใช้คำสันธานสูงกว่า GEN3 ซึ่งอาจแสดงถึงความซับซ้อนของประโยคในช่วงแรก

คำบุพบท (Prepositions) ปรากฏ 291 ครั้ง ถูกใช้ในปริมาณที่สม่ำเสมอในทุก GEN แสดงถึงบทบาทสำคัญในการเพิ่มความชัดเจนของบริบท

คำบ่งชี้ (Determiners) ปรากฏ 294 ครั้ง ช่วยให้เนื้อหามีความเจาะจงมากขึ้น

ผลการวิเคราะห์แสดงให้เห็นว่า ChatGPT มีการปรับเปลี่ยนคำในแต่ละประเภทเพื่อสร้างความสมดุลระหว่างการเล่าเรื่อง การบรรยาย และการเพิ่มรายละเอียดของเนื้อหา การเปลี่ยนแปลงนี้สะท้อนถึงความพยายามของเอไอ ในการพัฒนาความสมจริงและความน่าสนใจของเนื้อหาที่สร้างขึ้น

การวิเคราะห์คุณลักษณะเชิงศัพท์ของงานเขียนที่ถูกปรับโดย ChatGPT ในตำแหน่งแกรมที่ 3

1. การวิเคราะห์คุณลักษณะเชิงศัพท์

จากการศึกษาคุณลักษณะเชิงศัพท์ที่เกิดจากการปรับคำในตำแหน่งแกรมที่ 3 โดย ChatGPT พบว่ามีรูปแบบความสัมพันธ์เชิงศัพท์ 8 ประเภทหลัก ได้แก่

Antonym (คำคู่ตรงข้าม) เกิดขึ้นเพียง 7 ครั้ง (ความถี่ต่ำ) อาจมาจากความผิดพลาดของคำสั่งหรือโปรแกรม

Co-hyponym คำคำที่มี hypernym เดียวกัน เกิดขึ้น 103 ครั้ง (อันดับที่ 3) สะท้อนถึงการจับคู่คำที่ไม่สัมพันธ์กันโดยตรง

Co-hypernym คำคำที่มี hyponym เดียวกัน เกิดขึ้น 3 ครั้ง (ความถี่ต่ำมาก) แต่ยังนับเป็นความสัมพันธ์เชิงศัพท์

Hyponym คำที่ความหมายแคบกว่า เกิดขึ้น 40 ครั้ง แสดงถึงการทำให้บริบทกว้างขึ้น

Hypernym คำที่ความหมายกว้างกว่า เกิดขึ้น 70 ครั้ง ช่วยทำให้เนื้อหาชัดเจนมากขึ้น

Synonym (คำเหมือน) เกิดขึ้นมากถึง 473 ครั้ง (อันดับที่ 2) แสดงศักยภาพของโปรแกรมในการจับคู่คำที่แทนกันได้

Meronym คำที่เป็นส่วนประกอบของอีกคำหนึ่ง เกิดขึ้น 11 ครั้ง (ความถี่ต่ำ)

Holonym คำที่เป็นองค์รวมของอีกคำหนึ่ง พบได้น้อย แต่ยังเป็นความสัมพันธ์เชิงศัพท์สำคัญ

2. การเปลี่ยนแปลงเพิ่มเติมผ่าน Word Embeddings Similarity

การเปลี่ยนแปลงส่วนใหญ่ (584 ครั้ง) เกิดจากการใช้โมเดล Word Embeddings เช่น CBOW และ Skip-gram ซึ่งช่วยระบุคำที่มีค่าความเหมือนสูงที่สุดในเชิงเวกเตอร์ การเปลี่ยนแปลงในรูปแบบนี้สะท้อนความสามารถเชิงคณิตศาสตร์และการประมวลผลของ ChatGPT ในการเลือกคำที่เหมาะสม แม้จะไม่สามารถจัดอยู่ในกลุ่มความสัมพันธ์เชิงศัพท์แบบดั้งเดิม

3. ความสำคัญและข้อค้นพบหลัก

Synonym และ Word Embeddings Similarity เป็นรูปแบบที่พบมากที่สุด แสดงถึงความสามารถของ ChatGPT ในการปรับคำให้คงบริบทเดิมได้อย่างมีประสิทธิภาพ แม้ความสัมพันธ์ย่อย เช่น Antonym และ Meronym จะพบได้น้อย แต่ยังมีบทบาทสำคัญในการอธิบายการทำงานของโปรแกรม การผสมผสานความรู้ทางภาษาศาสตร์กับการประมวลผลเชิงคอมพิวเตอร์ช่วยเพิ่มความแม่นยำในการเลือกคำที่เหมาะสม ผลการวิเคราะห์ชี้ว่า ChatGPT สามารถรักษาบริบทของข้อความได้ดีโดยอาศัยความสัมพันธ์เชิงศัพท์และการวิเคราะห์เวกเตอร์ อย่างไรก็ตาม ความถี่ที่ต่ำของบางรูปแบบอาจสะท้อนข้อจำกัดหรือข้อผิดพลาดในการสั่งงานหรือการประมวลผล

องค์ความรู้ใหม่

องค์ความรู้ใหม่ที่ได้จากงานวิจัยนี้ ได้แก่



ภาพที่ 2 องค์ความรู้ใหม่ที่ได้จากงานวิจัย

การเปลี่ยนแปลงคำในตำแหน่งแกรมที่ 3 ส่งผลต่อคุณภาพและความสอดคล้องของเนื้อหาที่สร้างโดย ChatGPT โดยช่วยปรับสมดุลการใช้คำนาม, คำกริยา, คำคุณศัพท์ และคำวิเศษณ์ ให้เนื้อหาสมจริงและเข้าใจง่ายขึ้น งานวิจัยนี้ใช้วิธี n-gram และเทคนิค Word Embeddings Similarity เพื่อวิเคราะห์คำและความสัมพันธ์เชิงศัพท์ 8 ประเภท เช่น Synonym, Antonym, และ Hypernym ช่วยให้ ChatGPT เลือกคำที่เหมาะสมและรักษาความสอดคล้องของเนื้อหาได้ดีขึ้น

ข้อเสนอแนะ

1. การนำผลวิจัยไปใช้สามารถช่วยพัฒนา AI สร้างเนื้อหาให้เหมาะสมกับบริบทต่าง ๆ เช่น การตลาด การศึกษา และงานวิชาการ โดยเน้นความสมจริงและคุณภาพของเนื้อหา นอกจากนี้ การพัฒนาเครื่องมือช่วยตรวจสอบและปรับปรุงเนื้อหาจะช่วยให้ผู้ใช้งานสามารถสร้างข้อความที่สอดคล้องและมีประสิทธิภาพมากขึ้นในบริบทที่หลากหลาย

2. การวิจัยในอนาคตควรขยายการวิเคราะห์ไปยังตำแหน่งแกรมอื่น ๆ และประเภทเนื้อหาที่หลากหลาย เช่น บทความข่าวหรือโฆษณา รวมถึงเปรียบเทียบผลลัพธ์จากโมเดล AI ต่าง ๆ เพื่อต่อยอดการพัฒนา ควบคู่การศึกษาเชิงจริยธรรมเพื่อแก้ไขปัญหาอคติและความเสี่ยงในการใช้งาน เพื่อให้เทคโนโลยีมีความน่าเชื่อถือและเหมาะสมในระยะยาว

เอกสารอ้างอิง

ราชบัณฑิตยสถาน. (ม.ป.ป.). *พจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2554*. สืบค้นเมื่อ 28 ตุลาคม 2567, จาก <https://dictionary.orst.go.th/>

Aggarwal, T. (20 December 2023). *Demystifying NLP: exploring lexical, syntactic, and semantic processing for powerful natural language understanding*. Retrieved November 14, 2024, from <https://techladder.in/article/demystifying-nlp-exploring-lexical-syntactic-and-semantic-processing-powerful-natural>

Arts2Survive. (4 December 2022). *NLP Unlocked: N-Grams#006*. Retrieved November 14, 2024, from <https://medium.com/@pankajchandravanshi/nlp-unlocked-n-grams-006-ceab1bc56bf4>

Cambridge Dictionary. (n.d.). *Word classes and phrase classes*. Retrieved November 14, 2024, from <https://dictionary.cambridge.org/grammar/british-grammar/word-classes-and-phrase-classes>

DeepAI. (n.d.). *N-Grams*. Retrieved November 14, 2024, from <https://deepai.org/machine-learning-glossary-and-terms/n-gram#:~:text=What%20are%20N%2Dgrams%3F,computational%20linguistics%20and%20text%20analysis>.

Dey, R. (2 October 2023). *Understanding language modeling: from N-grams to transformer-based neural models*. Retrieved November 14, 2024, from <https://medium.com/@roshmitadey/understanding-language-modeling-from-n-grams-to-transformer-based-neural-models-d2bdf1532c6d>

- Gomede, E. (19 August 2023). *Exploring N-gram Models in Natural Language Processing*. Retrieved November 14, 2024, from <https://medium.com/@evertongomede/exploring-n-gram-models-in-natural-language-processing-bf5852b32050>
- GeekforGeeks. (24 May 2020). *N-gram language models*. Retrieved November 14, 2024, from <https://medium.com/mti-technology/n-gram-language-models-70af02e742ad>.
- GeekforGeeks. (28 November 2021). *Understanding Semantic Analysis – NLP*. Retrieved November 14, 2024, from <https://www.geeksforgeeks.org/understanding-semantic-analysis-nlp/>
- Imani, A., & Habil, H. (2014). Lexical features of academic writing. *LSP International Journal*, 1(1).
- Jagarlapoodi, S. (19 May 2023). *Lexical analysis: unveiling the language's structural foundations*. Retrieved November 14, 2024, from <https://www.linkedin.com/pulse/lexical-analysis-unveiling-languages-structural-jagarlapoodi/>
- Jain, A. (5 February 2024). *N-grams in NLP*. Retrieved November 14, 2024, from <https://medium.com/@abhishekjainindore24/n-grams-in-nlp-a7c05c1aff12>.
- Javapoint. (n.d.). *NLP Tutorial*. Retrieved November 14, 2024, from <https://www.javatpoint.com/nlp>
- Kalotra, S. (22 November 2023). *A deep dive into openai's GPT models: architectural insights*. Retrieved November 14, 2024, from <https://www.signitysolutions.com/tech-insights/openais-gpt-models-architectural-insights>
- Lang, N. (19 October 2024). *What are N-grams?*. Retrieved November 14, 2024, from <https://databasecamp.de/en/ml/n-grams>
- Langeek. (n.d.). *Lexical relations*. Retrieved November 14, 2024, from <https://langeek.co/en/grammar/course/1633/lexical-relations>
- Madala, S. (4 May 2023). *Introduction to N-grams in NLP*. Retrieved November 14, 2024, from <https://www.scaler.com/topics/nlp/ngrams-in-nlp/>
- Martínez, G., Hernández, J. A., Conde, J., Reviriego, P. & Merino, E. (2024). Beware of words: evaluating the lexical diversity of conversational LLMs using ChatGPT as case study. *ACM Transactions on Intelligent Systems and Technology*. DOI:10.1145/3696459
- Mljourney. (3 July 2024). *Understanding N-Gram language models*. Retrieved November 14, 2024, from <https://mljourney.com/understanding-n-gram-language-models/>
- Nguyen, K. (19 May 2020). *N-gram language models Part 1: Unigram model*. Retrieved November 14, 2024, from <https://medium.com/mti-technology/n-gram-language-model-b7c2fc322799>

Nithyashree. (12 December 2024). *What are N-Grams and how to implement them in python?*.

Retrieved November 14, 2024, from <https://www.analyticsvidhya.com/blog/2021/09/what-are-n-grams-and-how-to-implement-them-in-python/>

Patel, R. (27 September 2024). *Evaluating content quality: a guide to verify AI-generated*

text. Retrieved November 14, 2024, from <https://texta.ai/blog/ai-content/evaluating-content-quality-a-guide-to-verify-ai-generated-text>

Ramponi, M. (23 December 2022). *How ChatGPT actually works*. Retrieved November 14,

2024, from <https://www.assemblyai.com/blog/how-chatgpt-actually-works/>

Research Graph. (19 March 2024). *The journey of large language models: evolution, application, and limitations*. Retrieved November 14, 2024, from

<https://medium.com/@researchgraph/the-journey-of-large-language-models-evolution-application-and-limitations-c72461bf3a6f>

Team Experts. (2 July 2023). *N-grams: AI (Brace for these Hidden GPT Dangers)*. Retrieved

November 14, 2024, from <https://predictivethought.com/n-grams-ai-brace-for-these-hidden-gpt-dangers/>