

ระบบแชทบอทสำหรับองค์กร

Beary GPT: Generative AI for SET



สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2566

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Beary GPT: Generative AI for SET



PIYAPORN THANGJITSIRISIN

PRUEKSA KHUMPAEN

A COOPERATIVE EDUCATION SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE, SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2023

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา ระบบแซทบอทสำหรับองค์กร
Bery GPT: Generative AI for SET

ชื่อนักศึกษา นางสาวปิยาภรณ์ ตั้งจิตสิริสิน รหัสนักศึกษา 63050157
นางสาวพุกษา ขำแป้น รหัสนักศึกษา 63050160

ปริญญา วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา วิทยาการคอมพิวเตอร์

ปีการศึกษา 2566

อาจารย์ที่ปรึกษา ดร.จักรพันธ์ เตไชยา

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตร์บัณฑิต สาขาวิชา วิทยาการคอมพิวเตอร์ ประจำปีการศึกษา 2566

คณะกรรมการสอบ	ลายมือชื่อ
ดร. รุ่งรัตน์ เวียงศรีพนาวัลย์ ประธานกรรมการ	รุ่งรัตน์ เวียงศรีพนาวัลย์
ดร.จักรพันธ์ เตไชยา กรรมการและอาจารย์ที่ปรึกษา	จักรพันธ์ เตไชยา

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	ระบบแชทบอทสำหรับองค์กร Bery GPT: Generative AI for SET
ชื่อนักศึกษา	นางสาวปิยาภรณ์ ตั้งจิตสิริสิน รหัสนักศึกษา 63050157 นางสาวพุกษา ข้าแป้น รหัสนักศึกษา 63050160
ปริญญา	วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ดร.จักรพันธ์ เตไชยา

บทคัดย่อ

โปรแกรมสนทนาอัตโนมัติ หรือแชทบอทสำหรับองค์กร พัฒนาขึ้นเพื่อสนับสนุนการจัดการความรู้และกระบวนการเรียนรู้ของบุคลากรในองค์กรขนาดใหญ่ โดยเน้นให้มีการเรียนรู้งานที่มีประสิทธิภาพและลดปัญหาที่เกี่ยวข้องกับการรับบุคลากรเข้ามาทำงานใหม่ในองค์กร เช่น ระบบการทำงานภายในองค์กร การศึกษาจากงานการเรียนรู้ของบุคลากรเก่า ข้อจำกัดในการเข้าถึงข้อมูล และทรัพยากร การสื่อสารและความร่วมมือ และความซับซ้อนของงานที่อาจเป็นอุปสรรคในการเรียนรู้ของบุคลากรใหม่ โครงการนี้มุ่งเน้นพัฒนาระบบแชทบอทที่สามารถรวบรวมและจัดการความรู้ในองค์กรได้อย่างมีประสิทธิภาพ โดยอำนวยความสะดวกให้กับบุคลากรใหม่ในการเข้าถึงข้อมูลและเรียนรู้เกี่ยวกับงาน ระบบนี้สามารถวิเคราะห์ผลลัพธ์ของกิจกรรมต่างๆ ในองค์กรเพื่อช่วยลดความผิดพลาด ลดเวลาและความยุ่งยากในการดำเนินงาน และเพิ่มคุณภาพของงานที่ทำให้องค์กรดำเนินงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้น โครงการนี้เป็นเครื่องมือที่สำคัญในการพัฒนาบุคลากรและองค์กรให้มีประสิทธิภาพสูงสุด โดยช่วยในการรวบรวมและจัดการข้อมูล ใช้งานระบบ Knowledge Management เพื่อให้บุคลากรสามารถค้นหาข้อมูลและเรียนรู้ได้อย่างสะดวก และใช้ปัญญาประดิษฐ์เพื่อให้ระบบสามารถโต้ตอบได้เหมือนมนุษย์และมีประสิทธิภาพมากขึ้น ซึ่งพัฒนาระบบในรูปแบบของ Web Application เพื่อใช้ในการพูดคุยโต้ตอบกับบุคลากรภายในองค์กร และช่วยลดเวลาที่บุคลากรใช้ในการอบรมและสอนงานแก่บุคลากรใหม่ ทำให้องค์กรสามารถดำเนินการได้อย่างมีประสิทธิภาพและลดความผิดพลาดในการทำงาน โดยการทดสอบประสิทธิภาพจะใช้ cosine similarity ได้ผลลัพธ์ที่ 0.90 และ Euclidean Distance 0.54 ซึ่งหมายความว่าระบบสามารถดึงข้อมูลมาตอบกลับได้อย่างถูกต้องและสอดคล้องตามคำถามของผู้ใช้งานและข้อมูลที่ถูกฝึกสอน

คำสำคัญ: อัตโนมัติ, แชทบอท, การจัดการความรู้, ปัญญาประดิษฐ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น เมื่อผู้มาติดต่อให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Beary GPT: Generative AI for SET
Students	Miss Piyaporn Thangjitsirisin Student ID 63050157 Miss Prueksa Khumpaen Student ID 63050160
Degree	Bachelor of Science (Computer Science)
Department	Computer Science
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2023
Advisor	Dr. Jakapun Tachaiya

Abstract

Beary GPT, Chatbot for Enterprise Systems, aims to support proprietary knowledge management and employee learning within the private corporation. The focus is on efficient learning and addressing issues related to onboarding new employees, such as internal workflows, learning from experienced staff, data accessibility, communication, and task complexity. The chatbots effectively collect and manage organizational knowledge, providing convenience for new employees to access information and learn. They analyze activities, reducing errors, saving time, streamlining operations, and improving work quality. This project is a valuable tool for developing efficient personnel and organizations, utilizing Knowledge Management systems, artificial intelligence and web-based chatbot interactions, while also reducing training time and minimizing errors. The performance test uses cosine similarity and the result is 0.90 and the Euclidean Distance is 0.54, which means the system can retrieve information to respond accurately and in accordance with the user's questions and trained data.

Keywords: Automatic, Chatbot, Knowledge Management, Artificial Intelligence

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานสหกิจศึกษานี้สำเร็จลุล่วงไปได้อย่างสมบูรณ์ด้วยความกรุณาอย่างยิ่งจาก ดร.จักรพันธ์ เตไชยา ผู้เป็นอาจารย์ที่ปรึกษาและกรรมการของงานสหกิจศึกษาครั้งนี้ ที่ได้สละเวลาเพื่อให้ความกรุณาในการช่วยเหลือ ให้คำปรึกษา และแนะนำสิ่งต่างๆ ที่เป็นประโยชน์ในการทำงาน สหกิจศึกษาครั้งนี้ รวมถึงช่วยตรวจทานและแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่ เพื่อให้คณะผู้จัดทำมีความรู้เพิ่มมากขึ้นและมีแนวทางในการแก้ไขปัญหาเพิ่มเติมตลอดการทำงาน จนงานสหกิจศึกษาครั้งนี้สำเร็จลุล่วงได้ด้วยดี คณะผู้จัดทำจึงขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์ ประธานกรรมการของงานสหกิจศึกษา ที่ช่วยตรวจทาน ให้คำปรึกษา และให้ข้อเสนอแนะในการทำงาน ทำให้ได้เห็นถึงข้อที่ควรแก้ไข ข้อดีและข้อเสียต่างๆ ในการทำงานสหกิจศึกษาในครั้งนี้

ขอขอบพระคุณ คุณคณิต เศวตสุทธิสิริกุล และ คุณอรรคมัญญ์ ปฐมวงศ์ษา บุคลากรฝ่ายพัฒนาระบบสารสนเทศองค์กร และ คุณสิรินดา จรัสสุนทรวงศ์ บุคลากรฝ่ายการบริหารจัดการทรัพยากรมนุษย์ ตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) ที่ได้คอยดูแล ช่วยอำนวยความสะดวก ให้คำแนะนำ ให้คำปรึกษาและคอยช่วยเหลืออย่างดีตลอดการทำงานตั้งแต่เริ่มทำงานสหกิจศึกษาจนงานสำเร็จไปได้ด้วยดี

สุดท้ายนี้ ขอขอบคุณเพื่อนๆ ทั้งจากภาควิชาวิทยาการคอมพิวเตอร์ เพื่อนจากภาควิชาอื่นๆ รวมทั้งเพื่อนในช่วงมัธยมศึกษา ที่ได้คอยช่วยเหลือและเป็นกำลังใจให้กับคณะผู้จัดทำมาโดยเสมอ

ปิยาภรณ์ ตั้งจิตสิริสิน
พฤษภา ขำแป้น

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ช
สารบัญรูป.....	ฉ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานสหกิจศึกษา.....	2
1.3 ขอบเขตของงานสหกิจศึกษา.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 เครื่องมือที่ใช้ในการพัฒนา.....	3
บทที่ 2 ทฤษฎีและเทคโนโลยีที่เกี่ยวข้อง.....	4
2.1 การจัดการองค์ความรู้.....	4
2.2 เทคโนโลยีที่ใช้ในการพัฒนา.....	4
2.2.1 Python.....	4
2.2.2 React Typescript.....	5
2.2.3 CSS.....	5
2.2.4 Azure OpenAI.....	6
2.2.4.1 GPT-3.5-turbo-16k.....	7
2.2.4.2 OpenAI Text Embedding.....	8
2.2.4.3 Tiktoken.....	8
2.2.5 Azure Cognitive Search.....	8
2.2.6 Azure Blob Storage.....	9
2.2.7 โปรแกรมที่ใช้ในการพัฒนา.....	9
2.2.7.1 Visual Studio Code.....	9
2.2.7.2 Azure Portal.....	10
2.2.7.3 Figma.....	10
2.2.7.4 Gitlab.....	11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.2.7.5 GitHub Desktop.....	11
2.2.7.6 Jira Software.....	11
2.2.7.7 Postman.....	12
2.2.7.8 Draw.io.....	12
2.3 ทฤษฎีที่ใช้ในการเปรียบเทียบภาษาธรรมชาติ.....	13
2.3.1 ROUGE Score.....	13
2.3.2 BLEU Score.....	13
2.3.3 Cosine Similarity.....	14
2.3.4 Euclidean distance.....	15
2.4 เทคโนโลยีที่เกี่ยวข้อง.....	16
2.4.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing).....	16
2.4.2 โมเดลภาษาขนาดใหญ่ (Large Language Model).....	17
2.4.3 เทคโนโลยี Cloud Computing.....	17
2.5 Prompt.....	17
2.6 งานวิจัยที่เกี่ยวข้อง.....	17
2.6.1 azure-search-openai-demo.....	17
2.6.2 sample-app-aoai-chatGPT.....	19
2.6.3 Retrieval Augmented Generation (RAG).....	20
บทที่ 3 วิธีการดำเนินงานสหกิจศึกษา.....	22
3.1 Hardware Architecture.....	22
3.2 Software Architecture.....	23
3.2.1 Chat Flow.....	23
3.2.2 Index flow.....	24
3.2.3 ลำดับการเรียกใช้ service.....	25
3.3 System Requirement.....	26
3.3.1 Functional Requirement.....	26
3.3.2 Non-Functional Requirement.....	26

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.4 การออกแบบระบบด้วยแผนภาพ UML (Unified Modeling Language).....	27
3.4.1 Use Case Diagram.....	27
3.4.2 Use Case Description.....	27
3.4.3 Sequence Diagram.....	29
3.4.3.1 Sequence Diagram Create Search Index.....	29
3.4.3.2 Sequence Diagram Upload Document to Blob Storage...	30
3.4.3.3 Sequence Diagram Generate Answer.....	31
3.4.3.4 Sequence Diagram Get Citation File.....	32
3.5 Engine.....	33
3.5.1 การสร้าง Application.....	33
3.5.2 การกำหนด Prompt ของระบบ.....	33
3.5.3 การปรับแต่ง Prompt ให้มีประสิทธิภาพและเหมาะสมกับระบบ.....	35
3.5.4 การสร้าง Keyword และการสรุปเนื้อหาเพื่อเตรียมข้อมูลคำตอบ.....	38
3.5.5 การแปลงข้อมูลให้อยู่ในรูปแบบเวกเตอร์.....	38
3.5.6 การค้นหาข้อมูลและเอกสารที่เกี่ยวข้อง.....	39
3.5.7 การดูข้อมูลเอกสารอ้างอิง.....	40
3.5.8 การตรวจสอบความถูกต้องของคำตอบ.....	40
3.6 ค่าใช้จ่ายสำหรับการใช้งานบริการของ Azure Service.....	41
3.7 การออกแบบหน้าจอปฏิสัมพันธ์ (UI Design).....	42
3.7.1 หน้าจอเริ่มต้นของแชทบอท.....	42
3.7.2 หน้าจอการสนทนาของระบบ.....	43
3.7.3 หน้าจอการแสดงผลไฟล์เอกสารอ้างอิง.....	43
บทที่ 4 ผลสหกิจศึกษาและการอภิปรายผล.....	44
4.1 ชุดข้อมูลที่ใช้ในการทดสอบระบบ.....	44
4.2 การประเมินประสิทธิภาพของระบบ.....	44
4.3 หน้าจอปฏิสัมพันธ์ของระบบ (User Interface).....	48
4.3.1 หน้าหลักของระบบแชทบอท.....	48
4.3.2 หน้าจอการสนทนากับระบบแชทบอท.....	48
4.3.3 หน้าจอการแสดงผลข้อมูลเอกสารอ้างอิง.....	49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4.4 การทดสอบการทำงานของระบบ	49
4.5 การประเมินความพึงพอใจในการใช้งานระบบ	50
บทที่ 5 สรุปผลสทหกิจศึกษาและข้อเสนอแนะ.....	55
5.1 สรุปผลสทหกิจศึกษา	55
5.2 ปัญหาที่พบ	56
5.3 ข้อเสนอแนะ	56
เอกสารอ้างอิง	57
ภาคผนวก.....	60
ภาคผนวก ก	61
ภาคผนวก ข	65
ภาคผนวก ค	68



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางแสดงโมเดลของ GPT 3.5	7
3.1 Use Case Ask Question	27
3.2 Use Case View Answer	28
3.3 Use Case Delete Chat History	28
3.4 ค่าจ่ายของ Azure OpenAI Language Model	41
3.5 ค่าจ่ายของ Azure OpenAI Embedding Model.....	41
3.6 ค่าใช้จ่ายของ Azure Cognitive Search Service.....	41
3.7 ค่าใช้จ่ายของ Azure Blob Storage.....	42
4.1 สรุปผลการทดสอบการตอบคำถามของระบบที่ใช้ในการกำหนดเกณฑ์.....	45
4.2 ตัวอย่างชุดคำถาม-คำตอบเกี่ยวกับข้อมูลทั่วไปที่ใช้ในการทดสอบระบบ.....	45
4.3 สรุปผลการทดสอบการตอบคำถามของระบบ.....	47
4.4 รายละเอียดการทดสอบการส่ง API	49
4.5 รายละเอียดการทดสอบหน้าจอการใช้งาน.....	50
4.6 ผลการประเมินความพึงพอใจในการใช้งานระบบ.....	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 สัญลักษณ์ Python.....	4
2.2 สัญลักษณ์ React	5
2.3 สัญลักษณ์ TypeScript	5
2.4 สัญลักษณ์ CSS.....	6
2.5 สัญลักษณ์ OpenAI	6
2.6 สัญลักษณ์ GPT	7
2.7 สัญลักษณ์ Azure Cognitive Search.....	9
2.8 สัญลักษณ์ Azure Blob Storage.....	9
2.9 สัญลักษณ์ Visual Studio Code.....	10
2.10 สัญลักษณ์ Azure.....	10
2.11 สัญลักษณ์ Figma.....	10
2.12 สัญลักษณ์ Gitlab.....	11
2.13 สัญลักษณ์ GitHub Desktop.....	11
2.14 สัญลักษณ์ Jira Software	12
2.15 สัญลักษณ์ Postman	12
2.16 สัญลักษณ์ Draw.io	12
2.17 ตัวอย่างการคำนวณความแม่นยำด้วย ROUGE Score	13
2.18 Cosine Similarity.....	14
2.19 ตัวอย่างการเปรียบเทียบประโยคด้วย Cosine Similarity.....	15
2.20 การคำนวณ Euclidean Distance.....	16
2.21 ลำดับการทำงานและการเรียกใช้ service ในโปรแกรม.....	18
2.22 หน้าจอการสนทนาของแอปพลิเคชันตัวอย่างที่ 2.6.1.....	19
2.23 หน้าจอการสนทนาของแอปพลิเคชันตัวอย่างที่ 2.6.2.....	20
2.24 ภาพรวมการทำงานของ RAG	20
3.1 Hardware Architecture	22
3.2 Chat Flow	23
3.3 Index flow.....	24
3.4 ลำดับการเรียกใช้ service ภายในระบบเซทบอท.....	25
3.5 Use Case Diagram.....	27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.6 Sequence Diagram Create Search Index	29
3.7 Sequence Diagram Upload Document to Blob Storage	30
3.8 Sequence Diagram Generate Answer.....	31
3.9 Sequence Diagram Get Citation File	32
3.10 Prompt การตอบคำถามจากโปรแกรมตัวอย่าง (azure-search-openai-demo).....	34
3.11 Prompt การกำหนด keyword จากโปรแกรมตัวอย่าง (azure-search-openai-demo)	34
3.12 Prompt การตอบคำถามของระบบแชทบอท	34
3.13 Prompt การสร้าง Keyword ของระบบแชทบอท	34
3.14 Cognitive search process search by embedding	39
3.15 Cognitive search process search by text.....	39
3.16 หน้าจอเริ่มต้นของระบบ	42
3.17 หน้าจอการสนทนาของระบบ	43
3.18 หน้าจอการแสดงผลเอกสารอ้างอิง	43
4.1 กราฟแสดงผลลัพธ์ของการเปรียบเทียบด้วย Euclidean Distance ทั้ง 100 คำถาม	46
4.2 กราฟแสดงผลลัพธ์ของการเปรียบเทียบด้วย Cosine Similarity ทั้ง 100 คำถาม	47
4.3 หน้าหลักของระบบแชทบอท	48
4.4 หน้าจอการสนทนากับระบบแชทบอท	48
4.5 หน้าจอการแสดงผลข้อมูลเอกสารอ้างอิง	49
4.6 แบบฟอร์มประเมินความพึงพอใจในการใช้งานระบบ	51
4.7 ผลการประเมินความพึงพอใจในการใช้งานระบบ.....	52
4.8 กราฟสรุปความพึงพอใจโดยรวมในการใช้งานระบบ	52
ก.1 การสร้าง index client.....	61
ก.2 การตรวจสอบรายการ index และการสร้าง index.....	62
ก.3 การสร้าง field ของ index	63
ก.4 รายชื่อ index ที่แสดงใน azure portal.....	64
ก.5 รายการ field ใน index ที่ถูกสร้างขึ้นจากโปรแกรม.....	64
ข.1 การเชื่อมต่อกับ container ใน Azure Blob Storage.....	65
ข.2 การสร้าง ContainerClient	65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
ข.3 การอัปโหลดเอกสารเข้า Blob Storage.....	66
ข.4 การสร้างชื่อของ blob สำหรับอัปโหลดเอกสาร	66
ข.5 รายชื่อ container ที่แสดงใน azure portal	67
ข.6 รายการไฟล์ใน container ที่ถูกอัปโหลดจากโปรแกรม.....	67
ค.1 แผนการดำเนินงาน Sprint 1-5.....	68
ค.2 แผนการดำเนินงาน Sprint 6-12.....	69



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเทคโนโลยีปัญญาประดิษฐ์หรือ Artificial Intelligence (AI) เข้ามามีบทบาทและช่วยอำนวยความสะดวกในชีวิตประจำวันให้กับมนุษย์มากขึ้น รวมทั้งมีการพัฒนาเป็นระบบต่างๆ ที่ใช้ภายในองค์กร ซึ่งการทำงานภายในองค์กรบางส่วนก็อาจเกิดปัญหาหรือความผิดพลาดจากการทำงานของบุคลากร (Human Error) รวมทั้งปัญหาในการเรียนรู้งาน ข้อจำกัดในการเข้าถึงข้อมูลความรู้ภายในองค์กร ความซับซ้อนของงาน การติดต่อสื่อสารระหว่างแผนก การจัดสรรเวลาในการทำงาน เป็นต้น ซึ่งเป็นปัญหาที่พบได้บ่อยเนื่องจากข้อมูลอาจไม่ได้ถูกรวบรวมไว้ที่เดียวกันหรือบุคลากรบางคนก็อาจไม่ทราบข้อมูลทั้งหมด บางครั้งจึงเป็นการทำให้การดำเนินงานภายในองค์กรเป็นไปด้วยความล่าช้าอย่างหลีกเลี่ยงไม่ได้ เช่น การจัดการเอกสารการรับสมัครงาน การฝึกอบรมและการให้ความรู้แก่บุคคลากรใหม่ ทั้งนี้ ข้อมูลภายในองค์กรที่ต้องเรียนรู้และงานที่ได้รับมอบหมายมีจำนวนมาก จึงอาจทำให้บุคลากรทำงานได้ไม่เต็มประสิทธิภาพเท่าที่ควร

การพัฒนาบุคลากรและองค์กรให้มีประสิทธิภาพสูงที่สุดไปพร้อมกันจึงต้องอาศัยเทคโนโลยีและปัญญาประดิษฐ์เข้ามาช่วยพัฒนาระบบภายในองค์กร เพื่อให้เกิดความสะดวกในการเก็บรักษาและการจัดการข้อมูล การนำข้อมูลมาใช้งาน รวมไปถึงการวิเคราะห์ผลลัพธ์ของกิจกรรมต่างๆ ภายในองค์กร ให้สามารถเข้าถึงได้อย่างสะดวก เกิดความผิดพลาดในการทำงานน้อยลงและมีคุณภาพมากยิ่งขึ้นจากการนำประโยชน์ในแต่ละด้านของเทคโนโลยีและปัญญาประดิษฐ์มาใช้งาน

การพัฒนาระบบแชทบอทเกี่ยวกับการจัดการองค์ความรู้ในองค์กรหรือ Knowledge Management (KM) เพื่อช่วยอำนวยความสะดวกให้กับบุคลากรและองค์กรในการรวบรวมข้อมูลจัดการข้อมูล และประยุกต์ใช้ความรู้ในองค์กรให้เกิดประโยชน์สูงสุด ทำให้บุคลากรสามารถเข้าถึงข้อมูล ติดต่อบุคลากรแผนกต่างๆ และเรียนรู้การทำงานได้อย่างสะดวก และมีประสิทธิภาพมากขึ้นสามารถใช้งานได้ง่ายเนื่องจากเป็นแชทบอท ซึ่งจะสามารถโต้ตอบกับผู้ใช้งานได้อย่างเป็นธรรมชาติเพื่อให้เกิดประโยชน์สูงสุดต่อบุคลากรและองค์กร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของงานสหกิจศึกษา

- 1) เพื่อพัฒนาระบบโต้ตอบอัตโนมัติ (chatbot) ที่อำนวยความสะดวกและมีความเหมาะสมกับผู้ใช้งาน ในการเข้าถึงข้อมูลภายในองค์กร เช่น เอกสาร ISO, IT Process และ Security Guideline เป็นต้น
- 2) เพื่อนำเทคโนโลยีและปัญญาประดิษฐ์มาใช้งานให้เกิดประโยชน์สูงสุดกับการดำเนินงานภายในองค์กร
- 3) เพื่อสร้างส่วนต่อประสานผู้ใช้งาน (UI) ให้ผู้ใช้งานสามารถโต้ตอบกับระบบได้อย่างสะดวก

1.3 ขอบเขตของงานสหกิจศึกษา

ในการศึกษาครั้งนี้มีการกำหนดขอบเขตของการศึกษาและพัฒนาระบบ คือ พัฒนาระบบโต้ตอบอัตโนมัติ (chatbot) ในรูปแบบของเว็บแอปพลิเคชัน (web application) ซึ่งรองรับการใช้งานสำหรับคอมพิวเตอร์ส่วนบุคคล หรือ PC บนระบบปฏิบัติการ Windows และ macOS โดยขอบเขตด้านเนื้อหาในการศึกษาจะมีการศึกษาการทำงานของบริการ Microsoft Azure 3 บริการหลัก ดังนี้

- 1) Azure OpenAI
- 2) Azure Cognitive Search
- 3) Azure Blob Storage

โดยจะนำทั้ง 3 บริการข้างต้น มาใช้งานร่วมกันในระบบแชทบอทสำหรับองค์กร โดยศึกษาจากปัญหาที่เกี่ยวกับการติดต่อสื่อสาร และการเข้าถึงข้อมูลภายในองค์กร โดยข้อมูลภายในระบบจะมุ่งเน้นไปที่การรวบรวมองค์ความรู้ภายในองค์กรหรือ Knowledge Management (KM) มาพัฒนาเข้ากับระบบแชทบอท โดยกลุ่มผู้ใช้งานของระบบจะเป็นบุคลากรภายในองค์กรเท่านั้น

1.4 ประโยชน์ที่ได้รับ

- 1) ผู้ใช้งานสามารถเข้าถึงข้อมูลได้สะดวกและรวดเร็วขึ้น เช่น ข้อมูลการติดต่อของแต่ละแผนก ข้อมูลการทำงาน กฎระเบียบข้อบังคับขององค์กร เป็นต้น
- 2) ข้อมูลมีความถูกต้องแม่นยำและมีความน่าเชื่อถือ เพราะข้อมูลที่นำมาโต้ตอบกับผู้ใช้งานจะเป็นข้อมูลที่มีอยู่ในฐานข้อมูลขององค์กร
- 3) การทำงานภายในองค์กรมีประสิทธิภาพ รวดเร็วและทันสมัยมากขึ้น ช่วยให้บุคลากรที่รับผิดชอบในการดูแลบุคลากรใหม่ มีเวลาทำงานส่วนอื่นของตนเองมากขึ้น
- 4) เพื่อแก้ไขการตอบคำถามในการสอบถามข้อมูลหรือข้อสงสัยต่างๆ ให้เป็นไปอย่างถูกต้องแม่นยำ รวมทั้งอาจมีเอกสารอ้างอิง ทำให้เกิดความน่าเชื่อถือ
- 5) เพื่อลดการทำงานของบุคลากรในการแนะนำการทำงานทุกครั้งที่มีการรับบุคลากรใหม่เข้ามาทำงานภายในองค์กร รวมทั้งลดการดำเนินงานในส่วนที่สามารถนำเทคโนโลยี

เอกสารนี้เป็นเอกสารเข้ามาช่วยอำนวยความสะดวกให้กับบุคลากรได้ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 เครื่องมือที่ใช้ในการพัฒนา

ในการพัฒนาระบบแชทบอทสำหรับองค์กร ได้มีการใช้เครื่องมือและเทคโนโลยีต่างๆ ดังนี้

1) เครื่องคอมพิวเตอร์ Dell Inc. (Inspiron 5515)

OS Name: Microsoft Windows 11 Home Single Language

Version: 10.0.22621 Build 22621

OS Manufacturer: Microsoft Corporation

Processor: AMD Ryzen 5 5500U with Radeon Graphics,
2100 MHz, 6 Core(s), 12 Logical Processor(s)

RAM: 8.00 GB

Hard Disk: Solid State Drive (SSD)

2) เครื่องคอมพิวเตอร์ LENOVO (82LN)

OS Name: Microsoft Windows 11 Home Single Language

Version: 10.0.22621 Build 22621

OS Manufacturer: Microsoft Corporation

Processor: AMD Ryzen 7 5700U with Radeon Graphics,
1801 MHz, 8 Core(s), 16 Logical Processor(s)

RAM: 8.00 GB

Hard Disk: Solid State Drive (SSD)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและเทคโนโลยีเกี่ยวข้อง

ในการศึกษาและดำเนินการพัฒนาระบบแซทบอทสำหรับองค์กร ซึ่งเป็นแซทบอทที่ช่วยอำนวยความสะดวกให้กับบุคลากรภายในองค์กร ในการตอบคำถามข้อมูลองค์กรทั่วไป มีการประยุกต์ใช้ทฤษฎีการจัดการองค์ความรู้ (KM) และปัญญาประดิษฐ์ (AI) เพื่อพัฒนาระบบ ทฤษฎีและเทคโนโลยีที่เกี่ยวข้องนี้เป็นส่วนสนับสนุนว่าการนำเทคโนโลยีปัญญาประดิษฐ์มาพัฒนาระบบแซทบอททำให้ระบบมีความสามารถและมีประสิทธิภาพมากขึ้นได้

2.1 การจัดการองค์ความรู้

การจัดการองค์ความรู้ (Knowledge Management: KM) เป็นการรวบรวมองค์ความรู้ภายในองค์กรทั้งจากเอกสารและบุคคล รวมถึงข้อมูลในการทำงาน ประสบการณ์ ทักษะ นำมารวบรวม วิเคราะห์ และพัฒนาขึ้นมาเป็นระบบ ทำให้การจัดการองค์ความรู้ภายในองค์กรเกิดประสิทธิภาพและประสิทธิผลมากขึ้น ซึ่งข้อมูลทั้งหมดนี้ล้วนมีความสำคัญต่อการพัฒนาบุคลากรและความก้าวหน้าขององค์กร แนวคิดและหลักการของการจัดการองค์ความรู้ มีดังนี้

- 1) องค์ความรู้เป็นสินทรัพย์ที่มีค่าขององค์กร
- 2) องค์ความรู้ควรได้รับการแบ่งปันและแลกเปลี่ยนอย่างมีประสิทธิภาพ
- 3) องค์ความรู้ควรได้รับการปรับปรุงและพัฒนาอยู่เสมอ

2.2 เทคโนโลยีที่ใช้ในการพัฒนา

2.2.1 Python

ภาษาโปรแกรม Python เป็นภาษาโปรแกรมระดับสูงที่ใช้การแปลความหมาย (Interpreted) เป็นภาษาสคริปต์และเป็นภาษาเชิงวัตถุ (Object-Oriented) มีการกำหนดประเภท dynamic typing และ dynamic binding ทำให้ไวยากรณ์ของภาษา Python มีความเรียบง่าย ทำงานได้เร็วขึ้น ไม่ซับซ้อน สามารถเข้าใจได้ง่าย ง่ายต่อการเรียนรู้ เหมาะสำหรับการพัฒนาซอฟต์แวร์ การเรียนรู้ของเครื่อง (Machine Learning) และการพัฒนาแอปพลิเคชันอย่างรวดเร็ว (Rapid Application Development) สามารถช่วยลดเวลาในการพัฒนาโปรแกรมได้ รองรับโมดูลและแพ็คเกจ มีไลบรารีให้ใช้จำนวนมาก รวมทั้งยังสามารถทำงานร่วมกับภาษาโปรแกรมอื่นได้ดี [10]



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
รูปที่ 2.1 สัญลักษณ์ Python

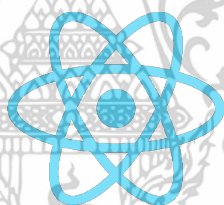
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 React TypeScript

React เป็นไลบรารี JavaScript ที่ช่วยในการสร้างส่วนต่อประสานผู้ใช้ (User Interface: UI) ในเว็บแอปพลิเคชัน โดยใช้แนวคิดของคอมโพเนนต์ (Component) ซึ่งช่วยแบ่ง UI ออกเป็นส่วนย่อย เพื่อจัดการกับส่วนต่างๆ ของ UI อย่างมีประสิทธิภาพและราบรื่นในการใช้งาน การใช้ React ช่วยให้สามารถสร้างแอปพลิเคชันที่มีประสิทธิภาพในการใช้งานได้มากยิ่งขึ้น [4]

TypeScript เป็นการพัฒนามาจาก JavaScript เป็นภาษาชนิด Static typed คือจะต้องกำหนดประเภทของตัวแปรก่อน และจะมีการตรวจสอบก่อนที่โปรแกรมเริ่มทำงาน ทำให้มีความสามารถในการตรวจสอบประเภทของข้อมูลได้ดีขึ้น จึงทำให้ TypeScript สามารถช่วยลดข้อผิดพลาดที่อาจเกิดขึ้นในระหว่างที่โปรแกรมทำงานได้ และการกำหนดประเภทของตัวแปรและฟังก์ชันยังช่วยในการจัดการโปรแกรมขนาดใหญ่หรือโปรแกรมที่มีโครงสร้างที่ซับซ้อนได้

การใช้ React TypeScript เป็นการใช้ภาษา TypeScript ร่วมกับไลบรารี React เพื่อสร้างเว็บแอปพลิเคชัน (Web Applications) โดยใช้ Typescript กำหนดประเภทให้กับตัวแปรภายในโปรแกรม โดย React ช่วยให้สามารถตรวจสอบประเภทข้อมูลของตัวแปรและฟังก์ชันได้ล่วงหน้า ซึ่งสามารถช่วยป้องกันข้อผิดพลาดที่อาจเกิดขึ้นได้ นอกจากนี้ Typescript ยังช่วยให้โค้ดที่เขียนสามารถอ่านง่ายและเข้าใจได้ง่ายขึ้นโดยการแสดงประเภทข้อมูลของตัวแปรและฟังก์ชันอย่างชัดเจน รวมทั้งยังทำให้โปรแกรมมีประสิทธิภาพมากขึ้นด้วย



รูปที่ 2.2 สัญลักษณ์ React



รูปที่ 2.3 สัญลักษณ์ TypeScript

2.2.3 CSS

CSS หมายถึง Cascading Style Sheets เป็นภาษาที่ใช้ในการตกแต่งหน้าเว็บที่สร้างขึ้นด้วย HTML หรือ XHTML เพื่อให้หน้าเว็บดูสวยงามและมีรูปแบบที่ถูกต้อง มีการกำหนดรูปแบบของข้อความ เส้นขอบ พื้นหลัง ระยะห่าง และอื่นๆ ตามต้องการ มีรูปแบบการเขียนโค้ดที่เฉพาะเจาะจงและถูกกำหนดมาตรฐานโดย W3C (World Wide Web Consortium) ซึ่งเป็นองค์กรเอกสารนี้มีมาตรฐานสำหรับเทคโนโลยีเว็บต่างๆ [8] ซึ่งประโยชน์ของ CSS ได้แก่ นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) ช่วยให้แก้ไขเอกสาร HTML ง่ายขึ้น การใช้ CSS ช่วยแยกแยะระหว่างเนื้อหาและรูปแบบการแสดงผล ทำให้สามารถแก้ไขเอกสาร HTML ได้อย่างง่ายและรวดเร็ว โดยไม่ต้องแก้ไขโค้ด HTML มาก เนื้อหาและรูปแบบจะแยกกันอย่างชัดเจน
- 2) ช่วยลดโค้ดในเอกสาร HTML ทำให้ขนาดไฟล์ลดลง ซึ่งช่วยให้เว็บไซต์โหลดได้เร็วขึ้น นักพัฒนาสามารถเลือกใช้ไฟล์ CSS ร่วมกันในหลายหน้าเว็บ เพื่อลดการโหลดข้อมูลซ้ำซ้อน
- 3) ช่วยให้สามารถกำหนดรูปแบบการแสดงผลที่เหมือนกันในหลายๆ ส่วนของเว็บไซต์ หรือหลายหน้าเว็บได้ ช่วยลดเวลาในการพัฒนาและปรับปรุงเว็บไซต์
- 4) ช่วยควบคุมการแสดงผล CSS ทำให้สามารถควบคุมการแสดงผลในรูปแบบที่เหมาะสมกับสื่อต่างๆ ได้ เช่น เว็บไซต์สามารถปรับขนาดและรูปแบบตามอุปกรณ์ที่เข้าชมได้อย่างถูกต้อง
- 5) ช่วยให้เว็บไซต์มีความเป็นมาตรฐานและสามารถรองรับการใช้งานในอนาคตได้ดี นอกจากนี้ ยังช่วยให้เว็บไซต์ดูทันสมัยมากขึ้น



รูปที่ 2.4 สัญลักษณ์ CSS

2.2.4 Azure OpenAI

Azure OpenAI เป็นการผสมผสานระหว่างบริการคลาวด์ Azure ของ Microsoft Azure และเทคโนโลยีปัญญาประดิษฐ์แบบ GPT-3 ของ OpenAI ช่วยให้นักพัฒนาสามารถใช้งานโมเดล OpenAI ร่วมกับแอปพลิเคชันของตนเองได้อย่างง่าย สามารถเข้าถึงโมเดลปัญญาประดิษฐ์ เครื่องมือ และทรัพยากรต่างๆ ได้อย่างหลากหลายและรวดเร็วในแอปพลิเคชันและบริการคลาวด์ของ Azure ได้อย่างสะดวกและปลอดภัย Azure OpenAI Service มี API ให้ใช้สำหรับงานที่ต่างกัน เช่น การสร้างข้อความ การแปลภาษา ซึ่งช่วยให้นักพัฒนาเริ่มต้นใช้งานได้อย่างง่ายดาย [24]

Azure OpenAI นำ OpenAI GPT-3 มาบนพื้นที่คลาวด์ Azure เพื่อให้ นักพัฒนาและองค์กรสามารถใช้งานความสามารถของ GPT-3 ในแอปพลิเคชันและบริการของตนเองได้ นักพัฒนาสามารถเข้าถึง Azure OpenAI API เพื่อนำ GPT-3 มาใช้ในโปรเจกต์และแอปพลิเคชันต่างๆ บน Azure โดยอัตโนมัติ และควบคุมการเรียกใช้งาน GPT-3 ได้อย่างยืดหยุ่นตามความต้องการของนักพัฒนาและองค์กร



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
รูปที่ 2.5 สัญลักษณ์ OpenAI
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.4.1 GPT-3.5-turbo-16k

GPT [14] ย่อมาจาก Generative Pre-trained Transformer ถูกพัฒนาโดย OpenAI ซึ่งเป็นเทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning: ML) แบบ Deep Learning และเป็นระบบปัญญาประดิษฐ์ (Artificial Intelligence: AI) ที่เกี่ยวข้องกับการประมวลผลภาษา สามารถสร้างข้อมูลข้อความเชิงธรรมชาติเหมือนที่มนุษย์เขียนได้ โดย GPT ใช้โครงสร้างของโครงข่ายประสาทเทียม (Neural Network) ชนิด Transformer ในการประมวลผลข้อมูลและสร้างข้อมูลใหม่

ในการทำระบบแชทบอทสำหรับองค์กรนี้ เลือกใช้โมเดล GPT-3.5-turbo-16k เนื่องจากเป็นโมเดลที่เข้าใจภาษาธรรมชาติ และสามารถสร้างข้อมูลข้อความภาษาธรรมชาติ หรือโค้ดได้ รองรับ token จำนวนมาก ทำให้ผู้ใช้งานสามารถสนทนาต่อเนื่องได้จำนวนมาก เป็นโมเดลที่มีประสิทธิภาพและมีความคุ้มค่าเมื่อเทียบกับโมเดล GPT-3.5 รุ่นอื่นๆ มีการปรับแต่งให้เหมาะสำหรับการสร้างคำตอบแบบสนทนาต่อเนื่อง (Chat completion) โดยใช้ API แต่ยังสามารถทำงานได้ดีในการสร้างคำตอบแบบการสนทนาครั้งเดียว (One-shot conversation)

ตารางที่ 2.1 ตารางแสดงโมเดลของ GPT 3.5 [26]

Model	Description	Max Token
gpt-3.5-turbo	GPT-3.5 ที่มีความสามารถมากที่สุด เหมาะสมสำหรับการแชท และมีราคาที่คุ้มค่ากว่าเวอร์ชันอื่นๆ ในรุ่น GPT-3.5 ถูกปรับแต่งให้เหมาะสำหรับการสนทนาแต่ยังทำงานได้ดีในงานการสร้างข้อมูลแบบดั้งเดิม	4,097
gpt-3.5-turbo-16k	ความสามารถเหมือน GPT-3.5-turbo แต่สามารถประมวลผล context ที่ความยาวมากกว่าถึง 4 เท่า	16,385
gpt-3.5-turbo-instruct	สามารถประมวลผลข้อความที่เกี่ยวข้องกับภาษาได้ดี สามารถสร้างคำตอบที่มีความยาวมากได้ ถูกออกแบบให้เข้ากับ Completions endpoint เวอร์ชันเก่าของ OpenAI API และไม่ได้ถูกปรับแต่งมาเป็น Chat Completions API	4,097

(ที่มา: <https://platform.openai.com/docs/models/gpt-3-5>)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานรูปที่ 2.6 สัญลักษณ์ GPT อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.4.2 OpenAI Text Embedding

Text Embedding เป็นเทคโนโลยีที่ใช้การเรียนรู้ของเครื่อง (Machine Learning) เพื่อแปลงข้อความให้เป็นเวกเตอร์ตัวเลขทศนิยม เพื่อใช้ในการวิเคราะห์และประมวลผลข้อความในรูปแบบที่คอมพิวเตอร์สามารถเข้าใจได้ โดยระยะห่างระหว่างเวกเตอร์ 2 ตัว จะใช้วัดความสัมพันธ์ของข้อความ หาก ค่า cosine similarity ของ 2 เวกเตอร์มีค่าใกล้เคียงกันมาก แสดงว่าข้อความนั้นมีความเกี่ยวข้องกันมากและมีความหมายใกล้เคียงกัน หากตัวเลขมีความใกล้เคียงกันน้อย แสดงว่าข้อความนั้นมีความเกี่ยวข้องกันน้อย เพราะสามารถใช้ค่า cosine ช่วยหาความหมายของประโยคได้ [27]

การเปลี่ยนข้อความเป็นเวกเตอร์ทำให้ง่ายต่อการดำเนินการต่างๆ เช่น การหาความคล้ายคลึงระหว่างข้อความ การจัดกลุ่มข้อมูลข้อความ การค้นหาข้อมูลที่เกี่ยวข้อง การแปลภาษา การวิเคราะห์รูปแบบของข้อมูล เป็นต้น โดยไม่จำเป็นต้องมีความเข้าใจเนื้อหาของข้อความจริงๆ เหมือนมนุษย์

OpenAI มีโมเดล embedding 2 รุ่น ซึ่งรุ่นแรกมีทั้งหมด 16 โมเดล และรุ่นที่ 2 มี 1 โมเดล คือ text-embedding-ada-002 ซึ่ง OpenAI แนะนำให้ใช้รุ่นนี้สำหรับทุกการใช้งาน เพราะมีการทำงานที่ดีกว่าในทุกด้าน มีราคาที่ถูกกว่า และง่ายต่อการใช้งาน Text-embedding-ada-002 สามารถประมวลผล request หรือ input ได้มากถึง 8,191 token และสามารถประมวลผลเวกเตอร์ หรือ output dimension ได้ถึง 1,536 ตัว [34]

2.2.4.3 Tiktoken

Tiktoken เป็นเครื่องมือที่ถูกพัฒนาโดย OpenAI ช่วยให้ผู้ใช้สามารถประมาณค่าใช้จ่ายในการใช้งาน OpenAI API ได้อย่างสะดวกและรวดเร็ว ใช้สำหรับนับจำนวน token ที่ใช้งานไปในข้อความโดยไม่ต้องทำการร้องขอ API จริง เป็นวิธีที่ช่วยให้ผู้ใช้สามารถประมาณค่าใช้จ่ายในการใช้งาน API ของ OpenAI ได้อย่างแม่นยำ การนับจำนวน token นี้ จะช่วยให้ผู้ใช้สามารถทราบว่าคุณสมบัติที่ต้องการประมวลผลจะมีค่าใช้จ่ายเท่าใดตามจำนวน token ที่ใช้ ซึ่งทำให้ง่ายต่อการวางแผนและการบริหารค่าใช้จ่าย

นอกจากนี้ Tiktoken ยังสามารถนำเสนอผลลัพธ์ในรูปแบบกราฟ เพื่อความเข้าใจที่ง่ายขึ้น เช่น การแสดงกราฟของจำนวน token ที่ใช้ในข้อความตามยอดรวมค่าใช้จ่าย เพื่อช่วยในการวางแผนและการตัดสินใจในการใช้งาน API ของ OpenAI [23]

2.2.5 Azure Cognitive Search

Azure Cognitive Search หรือ Azure Search คือ cloud search service มี API ให้นักพัฒนาสามารถใช้งานได้ เป็นปัญญาประดิษฐ์ที่มีความคล้ายคลึงกับมนุษย์ในด้านภาษา เป็นเครื่องมือสำหรับการสร้างการค้นหา สามารถแยกภาษาออกเป็นวลีแล้ววิเคราะห์ความหมายได้ รวมทั้งเพิ่มความเป็นธรรมชาติในการโต้ตอบกับมนุษย์บนอุปกรณ์ต่างๆ [11]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 สัญลักษณ์ Azure Cognitive Search

2.2.6 Azure Blob Storage

Azure Blob Storage เป็นบริการจัดเก็บข้อมูลแบบ Object Storage ในระบบคลาวด์ของ Microsoft Azure ที่ถูกออกแบบมาเพื่อรองรับการจัดเก็บข้อมูลที่มีปริมาณมากและเหมาะสมสำหรับข้อมูลแบบ Unstructured หรือข้อมูลที่ไม่มีโครงสร้างคงที่ เช่น ข้อความ ไฟล์ข้อมูล รูปภาพ วิดีโอ เสียง และอื่น ๆ อีกมากมาย บริการนี้มีความสามารถในการจัดเก็บข้อมูลในลักษณะของ Object เพื่อการจัดเก็บและการเข้าถึงข้อมูลที่มีปริมาณมากอย่างมีประสิทธิภาพและมีความเสถียรภายใน Azure Cloud [16]

การจัดเก็บข้อมูลของ Azure Blob Storage มี 2 ประเภทหลัก ได้แก่

- 1) Locally redundant storage (LRS) เป็นการจัดเก็บข้อมูลที่คัดลอกไปยังหลายๆ โชน (Zones) ในภูมิภาคหรือ Region เดียวกัน โดยคัดลอกข้อมูลอย่างน้อย 3 โชน และถูกจัดเก็บในโชนที่ไม่เกี่ยวข้องกัน ทำให้มีความคงทนต่อข้อขัดข้องในการให้บริการหากมีปัญหาเกิดขึ้นในโชนหนึ่ง
- 2) Zone-Redundant Storage (ZRS) เป็นการจัดเก็บข้อมูลที่คัดลอกไปยังหลายๆ โชน (Zones) ในภูมิภาคหรือ Region แต่การคัดลอกข้อมูลจะถูกจัดเก็บในโชนที่ไกลที่สุด ซึ่งเพิ่มความคงทนต่อความเสี่ยงในการให้บริการ การใช้ ZRS มีความสามารถในการสำรองข้อมูลในระดับ Zone ซึ่งเหมาะสำหรับการเก็บข้อมูลที่ต้องการความพร้อมใช้งานสูงและความเสถียรในระดับ Zone โดยเฉพาะ [19]



รูปที่ 2.8 สัญลักษณ์ Azure Blob Storage

2.2.7 โปรแกรมที่ใช้ในการพัฒนา

2.2.7.1 Visual Studio Code

Visual Studio Code เป็นโปรแกรมประเภท editor ใช้ในการแก้ไขโค้ด เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านธุรกิจ พัฒนาโดย Microsoft สามารถใช้งานได้กับ Windows, Linux และ macOS ใช้งานได้กับไม่ว่ากรณ์ใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาษาโปรแกรมหลากหลายภาษา มีส่วนของการ debugger การเชื่อมต่อและควบคุม Git ในตัว การแนะนำไวยากรณ์และการเติมโค้ดอัจฉริยะ ในการใช้งานโปรแกรมสามารถปรับแต่งได้หลายอย่าง เช่น ผู้ใช้งานสามารถเปลี่ยนธีม แป้นพิมพ์ลัด การตั้งค่า และติดตั้งส่วนขยายที่เพิ่มฟังก์ชันการทำงานเพิ่มเติมได้ สามารถใช้งานได้ฟรีสำหรับการใช้ส่วนตัวและการใช้งานเพื่อการค้า [2]



รูปที่ 2.9 สัญลักษณ์ Visual Studio Code

2.2.7.2 Azure Portal

Azure Portal เป็นเว็บแอปพลิเคชันที่ให้บริการโดย Microsoft Azure เป็นส่วนหนึ่งของ Microsoft Azure ซึ่งเป็นคลาวด์แพลตฟอร์ม (Cloud platform) ควบคุมการทำงานของ Azure ในรูปแบบ dashboard หรือการรวบรวมข้อมูลทั้งหมดให้อยู่ในหน้าจอเดียว มีความยืดหยุ่นในการใช้งานและพร้อมใช้งานเสมอ ให้บริการเกี่ยวกับ compute resources, networking services, database services, IoT services, data analytics and AI services และบริการอื่นๆ ในรูปแบบคลาวด์สำหรับองค์กรและนักพัฒนา สามารถใช้ Azure Portal เพื่อสร้างและจัดการได้หลายอย่าง ทั้งเว็บแอปพลิเคชันไปจนถึงการปรับใช้กับระบบคลาวด์ [6]



รูปที่ 2.10 สัญลักษณ์ Azure

2.2.7.3 Figma

Figma เป็นเครื่องมือออกแบบที่ใช้สำหรับการสร้าง แชนร์ และทดสอบ การแสดงผลหน้าจอสำหรับเว็บไซต์ แอปพลิเคชันมือถือ ผลิตภัณฑ์ดิจิทัล รวมถึงอุปกรณ์อื่นๆ ที่เกี่ยวข้อง เป็นเครื่องมือที่นิยมในการออกแบบและมีการใช้งานโดยนักออกแบบ ผู้จัดการผลิตภัณฑ์ นักเขียน และนักพัฒนา เพื่อช่วยให้ผู้ที่เกี่ยวข้องกับกระบวนการออกแบบสามารถมีส่วนร่วมให้ข้อเสนอแนะ และตัดสินใจที่ดีขึ้น โดยเร็วขึ้นได้ [18]



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
รูปที่ 2.11 สัญลักษณ์ Figma
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.7.4 GitLab

GitLab เป็นแพลตฟอร์ม open source สำหรับการพัฒนาซอฟต์แวร์และโปรเจกต์ขนาดใหญ่ ที่ให้ความสำคัญกับการทำงานร่วมกัน เหมือนมีที่เก็บโค้ดออนไลน์และเครื่องมือสำหรับติดตามปัญหา และการทดสอบโค้ด รองรับทั้งการพัฒนาแบบ public และ private มีเวอร์ชันฟรีสำหรับบุคคลที่ใช้งานรายบุคคล นักออกแบบ และนักพัฒนามากมาย เพื่อช่วยในกระบวนการออกแบบและพัฒนาซอฟต์แวร์ โดย GitLab มีความสามารถในการทำงานในทุกขั้นตอนของวงจรการพัฒนาซอฟต์แวร์ มีความปลอดภัยและมีความสามารถมากมาย ช่วยให้ทีมพัฒนามีประสิทธิภาพและมีความคุ้มค่ามากยิ่งขึ้น [33]



รูปที่ 2.12 สัญลักษณ์ GitLab

2.2.7.5 GitHub Desktop

GitHub Desktop เป็นแอปพลิเคชันที่ถูกพัฒนาขึ้นเพื่อช่วยให้ผู้ใช้งานสามารถสื่อสารและทำงานกับ GitHub ผ่านทาง GUI (Graphical User Interface) อย่างสะดวกและใช้งานได้ง่าย แอปพลิเคชันนี้จะช่วยในการจัดการโค้ด การตรวจสอบการเปลี่ยนแปลงของโค้ด และการทำงานร่วมกับทีมนักพัฒนา ด้วยการให้บริการทาง GUI ที่ใช้ง่ายและเพิ่มความสะดวกในการใช้ Git และ GitHub ทำให้สามารถทำงานกับพื้นที่เก็บข้อมูล ดูประวัติของการเปลี่ยนแปลง และทำการอัปเดตโค้ดอย่างรวดเร็ว ทั้งนี้ ทำให้การจัดการโปรเจกต์และการทำงานร่วมกับทีมนักพัฒนาเป็นเรื่องง่ายและมีประสิทธิภาพมากยิ่งขึ้น [21]



รูปที่ 2.13 สัญลักษณ์ GitHub Desktop

2.2.7.6 Jira Software

Jira โดย Atlassian เป็นเครื่องมือสำหรับการจัดการโปรเจกต์และงานที่เกี่ยวข้องกับการพัฒนาซอฟต์แวร์ ช่วยให้ทีมพัฒนาสามารถสร้างงาน ติดตามความคืบหน้า และจัดการกับงานอย่างมีระบบ ทำให้ทีมสามารถทำงานร่วมกันอย่างมีประสิทธิภาพและตรวจสอบการเปลี่ยนแปลงของโค้ดได้อย่างสะดวก [12]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Jira Software Board เป็นเครื่องมือที่ใช้ในการจัดการโปรเจกต์และงานที่เกี่ยวข้องใน Jira Software โดยเฉพาะ ช่วยให้ทีมนักพัฒนาสามารถแสดงงานและติดตามงาน ควบคุมความคืบหน้าของโปรเจกต์ และทำงานร่วมกันอย่างมีประสิทธิภาพ Jira Software Board มีความหลากหลายในการปรับแต่งและสร้างให้เหมาะกับความต้องการของแต่ละทีมและแต่ละงาน มีคุณสมบัติที่ช่วยให้ทีมมีการวางแผนและการทำงานที่เป็นระบบและเรียบง่าย [13]

Jira Software

รูปที่ 2.14 สัญลักษณ์ Jira Software

2.2.7.7 Postman

Postman เป็นแพลตฟอร์มสำหรับการทดสอบและจัดการกับ APIs (Application Programming Interfaces) ที่ถูกพัฒนาขึ้นเพื่อช่วยให้นักพัฒนาและทีมที่ทำงานกับ APIs สามารถทำงานได้อย่างรวดเร็วและมีประสิทธิภาพ Postman ช่วยในการออกแบบ ทดสอบ จัดการ และเปลี่ยนแปลง APIs ให้ง่ายและมีความมั่นคง นักพัฒนาสามารถใช้ Postman เพื่อสร้างคำขอ API ทดสอบฟังก์ชันการทำงานของ APIs จัดการรายการคำขอ และดูผลลัพธ์ของการเรียกใช้งาน API โดยละเอียด นอกจากนี้ Postman ยังมีคุณสมบัติที่ช่วยในการจัดการงานต่างๆ และการทำงานร่วมกับทีมนักพัฒนา [29]



รูปที่ 2.15 สัญลักษณ์ Postman

2.2.7.8 Draw.io

Draw.io เป็นเว็บแอปพลิเคชัน พัฒนาขึ้นโดยใช้เทคโนโลยี mxGraph เป็น Library ที่พัฒนาด้วยภาษา Javascript เป็นเครื่องมือสำหรับใช้ในการออกแบบและสร้าง diagram หรือแผนภาพผังงานต่างๆ แบบออนไลน์ โดยจะมีรูปแบบ diagram หลากหลายรูปแบบให้เลือกใช้ หรือสามารถออกแบบเองก็ได้เช่นกัน สามารถใช้งานได้ง่าย ไม่ต้องติดตั้งโปรแกรมเพิ่มเติมและสามารถใช้งานได้โดยไม่มีค่าใช้จ่ายใดๆ [5]



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.16 สัญลักษณ์ Draw.io

2.3 ทฤษฎีที่ใช้ในการเปรียบเทียบภาษาธรรมชาติ

ทฤษฎีที่นำมาใช้ในการเปรียบเทียบภาษาธรรมชาติ นำมาใช้ในการตรวจสอบว่าข้อความ ทั้ง 2 ชุดมีความคล้ายคลึงหรือมีความแตกต่างกันมากน้อยเพียงใด โดยจะศึกษาทั้งหมด 4 วิธี ได้แก่ ROUGE Score, BLEU Score, Cosine Similarity และ Euclidean Distance

2.3.1 ROUGE Score

ROUGE Score เป็นวิธีที่ใช้ในการประเมินประสิทธิภาพและคุณภาพของข้อความหรือคำอธิบาย ROUGE Score มีหลายรูปแบบ ซึ่งส่วนมากจะใช้ ROUGE-N เป็นวิธีการวัดความคล้ายคลึงของคำหรือวลีในข้อความต้นแบบกับข้อความเป้าหมายให้ออกมาในรูปแบบตัวเลขสัดส่วนของคำ หรือวลีทั้งหมดที่ถูพบในข้อความต้นแบบเทียบกับข้อความเป้าหมาย ผลลัพธ์ตัวเลขที่ได้นั้น ถ้ามีค่ามาก จะหมายถึงข้อความทั้ง 2 มีความคล้ายคลึงของคำกันมาก [22] ตัวอย่างวิธีการคำนวณ เช่น

$$\frac{\text{number of overlapping words}}{\text{total words in system summary}}$$

รูปที่ 2.17 ตัวอย่างการคำนวณความแม่นยำด้วย ROUGE Score

ตัวอย่างประโยค	ข้อความต้นแบบ	the cat was found under the bed
	ข้อความเป้าหมาย	the cat was under the bed

จะได้ค่าผลลัพธ์หรือ Precision = 0.86

การได้ค่าผลลัพธ์มีค่าเท่ากับ 0.86 เป็นการพิจารณาเฉพาะคำแต่ละคำในประโยค ซึ่งจำนวนคำที่คู่กันหรือเหมือนกันระหว่างข้อความต้นแบบกับข้อความเป้าหมายคือ 6 นำมาเทียบกับจำนวนคำของข้อความต้นแบบ ซึ่งมีค่าเป็น 7 เมื่อนำมาเทียบกันผลลัพธ์จึงมีค่าเท่ากับ 0.86 หมายความว่า 6 ใน 7 คำของข้อความเป้าหมายเมื่อเทียบกับข้อความต้นแบบแล้วมีความเกี่ยวข้องกันค่อนข้างมาก

2.3.2 BLEU Score

BLEU Score หรือ Bilingual Evaluation Understudy Score ใช้วัดความคล้ายคลึงระหว่างข้อความจากการแปลของเครื่องและข้อความอ้างอิง โดยคำนวณคะแนนที่ตรงกับคำและวลีที่มีอยู่ในข้อความจากการแปลของเครื่องเทียบกับข้อความอ้างอิง ซึ่งมีไวยากรณ์ที่ถูกต้องและถูกแปลโดยมนุษย์ BLEU Score คำนวณคะแนนโดยดูที่วลีที่ถูกต้องและที่ถูกแปล ทั้งในเรื่องความคล้ายและความยาวของข้อความที่ถูกแปล [32]

ผลลัพธ์ของ BLEU Score จะอยู่ในช่วงระหว่าง 0 ถึง 1 โดยคะแนนสูงหมายถึงความคล้ายสูงระหว่างข้อความแปลของระบบและข้อความอ้างอิง BLEU Score

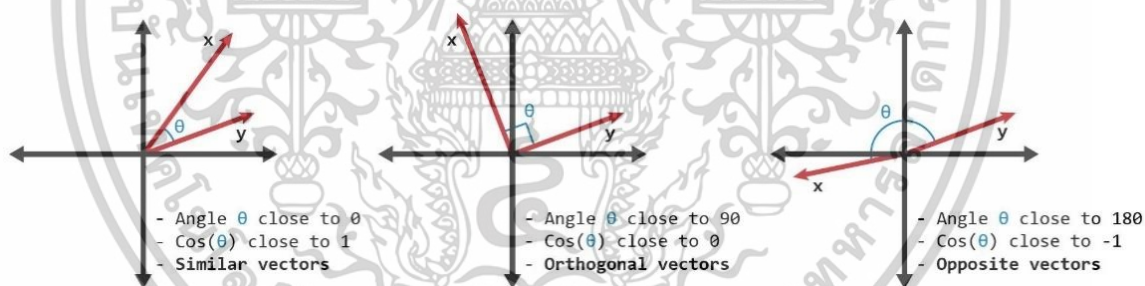
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระบวนการคำนวณ BLEU Score มีขั้นตอนที่คำนวณความคล้ายคลึงทั้งในระดับคำ (n-gram) และระดับประโยค การใช้ n-gram ที่มากขึ้น (เช่น unigram, bigram, trigram) ช่วยให้ BLEU Score สามารถคำนวณความคล้ายโดยพิจารณาลำดับคำหรือวลีที่ติดต่อกันในข้อความ แต่ยังมีข้อจำกัดในการประเมินคุณภาพการแปลเนื่องจากมองความคล้ายของข้อความ และไม่สนใจความหมายของข้อความ [31]

2.3.3 Cosine Similarity

Cosine Similarity เป็นวิธีที่มีประสิทธิภาพในการวัดความคล้ายคลึงทางความหมายของข้อมูล เป็นเครื่องมือสำคัญในการประเมินความคล้ายคลึงทางความหมายของคำและใช้ในการวัดประสิทธิภาพของระบบแชทบอทได้ โดยการเปรียบเทียบเวกเตอร์ของคำตอบที่ระบบสร้างขึ้นกับเวกเตอร์ของคำถามอ้างอิงที่ถูกตั้งเป็นมาตรฐาน ซึ่งทำให้สามารถประเมินการทำงานของระบบและวิเคราะห์ประสิทธิภาพของการตอบคำถามของระบบได้ [17]

Cosine Similarity คำนวณได้จาก Cosine Similarity $(A, B) = (A \cdot B) / (||A|| * ||B||)$ โดยที่ A และ B คือเวกเตอร์ของคำตอบที่จะเปรียบเทียบ และ $||A||$ และ $||B||$ แทนความยาว (magnitude) ของเวกเตอร์ทั้งสอง ค่า Cosine Similarity จะอยู่ในช่วงระหว่าง -1 ถึง 1 โดยที่ค่า 1 หมายถึงความคล้ายคลึงสูงสุดระหว่างคำตอบ และค่า -1 หมายถึงความคล้ายคลึงต่ำที่สุด ในขณะที่ค่า 0 หมายถึงไม่มีความคล้ายคลึงใดๆ ระหว่างชุดข้อมูลเวกเตอร์ทั้ง 2



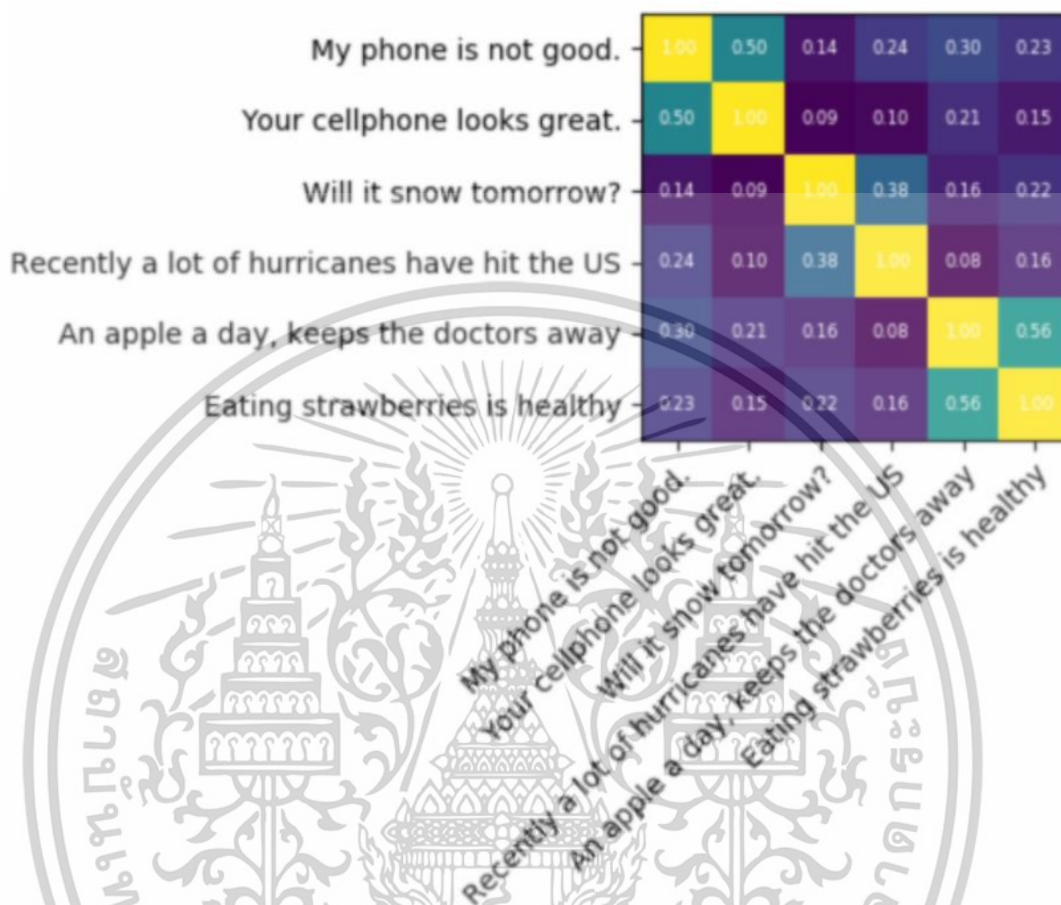
จากภาพแสดงรูปร่างกราฟของกราฟ Cosine Similarity เป็นกราฟที่ใช้เพื่อแสดงความคล้ายคลึงระหว่างความสัมพันธ์ของสองเวกเตอร์โดยใช้ค่า Cosine Similarity ตั้งแต่ -1 ถึง 1 บนแกน y และแสดงมุมระหว่างสองเวกเตอร์บนแกน x ของกราฟ

ค่า Cosine Similarity เท่ากับ 1 หมายถึงทั้งสองเวกเตอร์เหมือนกันอย่างสมบูรณ์ กราฟจะแสดงค่า Cosine Similarity เท่ากับ 1 ที่มุม 0 องศาบนกราฟ ซึ่งหมายถึงเวกเตอร์ทั้งสองมีความคล้ายคลึงและชี้ไปในทิศทางเดียวกัน หากเวกเตอร์มีค่าเข้าใกล้ 1 มาก จะหมายความว่าเวกเตอร์ทั้งสองมีความใกล้เคียงกันมาก

ค่า Cosine Similarity เท่ากับ 0 หมายถึงทั้งสองเวกเตอร์ไม่มีความสัมพันธ์ต่อกัน กราฟจะแสดงค่า Cosine Similarity เท่ากับ 0 ที่มุม 90 องศา หรือตั้งฉากกันบนกราฟ ซึ่งหมายถึงเวกเตอร์ทั้งสองไม่มีความคล้ายคลึงกันเลย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต หากมีข้อสงสัยหรือต้องการข้อมูลเพิ่มเติม กรุณาติดต่อฝ่ายบริการลูกค้า

ค่า Cosine Similarity เท่ากับ -1 ถ้าสองเวกเตอร์มีตรงข้ามกัน กราฟจะแสดงค่า Cosine Similarity เท่ากับ -1 ที่มุม 180 องศาบนกราฟ ซึ่งหมายถึงเวกเตอร์ทั้งสองมีความหมายในทิศทางตรงข้ามกัน



รูปที่ 2.19 ตัวอย่างการเปรียบเทียบประโยคด้วย Cosine Similarity

จากภาพที่ 2.19 เป็นตัวอย่างของการเปรียบเทียบประโยคด้วยการใช้วิธี Cosine Similarity เพื่อหาความคล้ายคลึงและความเกี่ยวข้องกัน เช่น ประโยค My phone is not good. กับ My phone is not good. เป็นประโยคเดียวกันมีความหมายเหมือนกัน ผลลัพธ์จึงได้ออกมาเป็น 1.00 แต่ถ้าเปรียบเทียบกันด้วยประโยค My phone is not good. กับ Eating strawberries is healthy ประโยคจะมีความหมายต่างกันและไม่มีความเกี่ยวข้องกัน จึงทำให้ได้ผลลัพธ์ออกมาที่มีค่าน้อยมากหรือประมาณ 0.23

2.3.4 Euclidean Distance

Euclidean Distance มีที่มาจากทฤษฎีบทพีทาโกรัสทางคณิตศาสตร์ เป็นการวัดระยะห่างระหว่างจุดทั้ง 2 จุดแนวเส้นตรงในมิติ 2 มิติหรือมิติสูงกว่านี้ การใช้ Euclidean Distance ง่ายต่อการเข้าใจและการคำนวณ รวมทั้งมีความสามารถในการวัดระยะห่างระหว่างจุดในมิติหลายๆ มิติ ซึ่งทำให้มีความหลากหลายในการใช้งาน แต่ก็เหมาะกับงานที่มีจำนวนมิติ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่มากจนเกินไป และการใช้งาน Euclidean Distance จะมีข้อจำกัด เช่น ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) เมื่อข้อมูลมีจำนวนมิติมากขึ้น การใช้ Euclidean Distance มักมีข้อจำกัดในด้านความสามารถในการทำงาน เนื่องจากในการวัดระยะห่างที่มีจำนวนหลายมิติจะทำให้ระยะห่างเพิ่มขึ้นอย่างไม่จำเป็น
- 2) Euclidean Distance ไม่คำนึงถึงความสัมพันธ์ระหว่างข้อมูล ซึ่งอาจทำให้ไม่เหมาะสมสำหรับข้อมูลที่มีความสัมพันธ์กัน

$$d_1(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

รูปที่ 2.20 การคำนวณ Euclidean Distance

จากสูตรการคำนวณระยะทางจากจุด p ไปยังจุด q โดยวัดในแบบ Euclidean ค่าของตัวแปร p และ q คือจุดใดๆ และ n คือจำนวนมิติของข้อมูล ผลลัพธ์จะได้เป็นระยะห่างระหว่างระยะทางของตัวแปรทั้ง 2 ในแนวเส้นตรง โดยถ้าผลลัพธ์ที่ได้จากการคำนวณมีค่าน้อยหมายความว่าข้อมูลนั้นมีความต่างกันเล็กน้อย แต่ถ้าข้อมูลมีความแตกต่างกันมากค่าผลลัพธ์ของ Euclidean Distance จะมีค่ามาก

2.4 เทคโนโลยีที่เกี่ยวข้อง

2.4.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

ปัญญาประดิษฐ์หมายถึงระบบหรือเทคโนโลยีการประมวลผลของคอมพิวเตอร์เป็นศาสตร์หนึ่งทางคอมพิวเตอร์ มุ่งเน้นการพัฒนาาระบบที่สามารถคิดและเรียนรู้ได้เหมือนมนุษย์ สามารถปรับปรุงให้มีความรู้มากขึ้นได้ โดยนำมาประยุกต์ต่อยอดใช้ร่วมกับระบบแชทบอท เช่น ใช้การประมวลผลภาษาธรรมชาติ (Natural Language Processing) ใช้การเรียนรู้ของเครื่อง (Machine Learning) และใช้การสร้างโมเดลภาษาขนาดใหญ่ (Large Language Model) เป็นการนำมาใช้ร่วมกันเพื่อเพิ่มประสิทธิภาพและเพิ่มความสามารถให้กับแชทบอทได้

การประมวลผลภาษาธรรมชาติ คือเทคโนโลยีที่ช่วยให้คอมพิวเตอร์สามารถตีความและเข้าใจภาษาของมนุษย์ได้ เป็นส่วนสำคัญในการวิเคราะห์ข้อมูลข้อความได้อย่างมีประสิทธิภาพ สามารถเข้าใจภาษาท้องถิ่นหรือประโยคที่ผิดไวยากรณ์ได้ มักนำมาใช้ช่วยในการประมวลผลหรือวิเคราะห์เอกสาร วิเคราะห์ข้อความ แบ่งประเภทข้อความ แปลภาษา นำมาช่วยในการสื่อสารและช่วยอำนวยความสะดวกให้กับมนุษย์ [15]

เทคโนโลยีการประมวลผลภาษาธรรมชาติสามารถนำมาประยุกต์ใช้กับระบบแชทบอทได้หลายวิธี เช่น ใช้ในการวิเคราะห์ข้อความหรือคำถามของผู้ใช้งาน การสร้างการตอบสนองที่สอดคล้องกับบริบทของบทสนทนา และเข้าใจความซับซ้อนของการสนทนาได้ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.2 โมเดลภาษาขนาดใหญ่ (Large Language Model)

โมเดลภาษาขนาดใหญ่เป็นโมเดลการเรียนรู้ของเครื่องขนาดใหญ่ โดยผ่านการฝึกฝนและวิเคราะห์ข้อมูลจำนวนมาก มีการทดสอบความสามารถโดยการวิเคราะห์ข้อมูลขนาดใหญ่ ใช้ในการประมวลผลภาษาธรรมชาติ สามารถสร้างรูปแบบการสนทนาและข้อมูลได้คล้ายคลึงกับมนุษย์ สามารถเรียนรู้การประมวลผล มีความสามารถในการตีความและประมวลผลข้อความ โดยอาศัยการคาดเดาและการวิเคราะห์ [7]

โมเดลภาษาขนาดใหญ่สามารถนำมาประยุกต์ใช้กับระบบแชทบอทได้หลายด้าน เช่น ใช้ในการตอบคำถามหรือข้อสงสัยของผู้ใช้งานได้อย่างรวดเร็วและถูกต้อง คิดค้นคำตอบและเชื่อมโยง ใช้เวลาในการประมวลผลน้อย รวมทั้งใช้ในการแปลภาษาและใช้ในการสร้างเนื้อหาใหม่

2.4.3 เทคโนโลยี Cloud Computing

เทคโนโลยี Cloud Computing เป็นบริการที่ครอบคลุมถึงการประมวลผล การจัดเก็บข้อมูล และระบบออนไลน์ต่างๆ จากผู้ให้บริการ ซึ่งสามารถช่วยลดต้นทุนได้ เนื่องจากสามารถเลือกจำนวนทรัพยากรได้ตามความต้องการ ลดความรับผิดชอบ และประหยัดเวลาในการดูแลระบบได้ สามารถเข้าถึงข้อมูลได้จากที่ต่างๆ [1]

การใช้ Cloud Computing ช่วยให้ระบบแชทบอทสามารถทำงานได้อย่างมีประสิทธิภาพมากขึ้น มีความยืดหยุ่นในการทำงาน สามารถทำงานร่วมกันได้ง่าย และสามารถเข้าถึงทรัพยากรได้จากทุกที่ผ่านอินเทอร์เน็ต

2.5 Prompt

Prompt หรือข้อความที่ป้อนให้กับ model AI เปรียบเสมือนคำสั่ง แนวทางในการสื่อสาร หรือแนวทางในการทำงานที่ส่งให้กับ AI เพื่อให้ระบบหรือโปรแกรมทำงานตามที่ได้กำหนดไว้ เป็นเครื่องมือที่มีประสิทธิภาพในการควบคุมการทำงานของระบบ AI ให้มีความเหมาะสมตามความต้องการที่ได้ระบุไว้ ซึ่งเป็นข้อความส่วนที่สามารถบอกความต้องการให้กับระบบได้ว่าต้องการให้ระบบทำงานอย่างไรบ้างและสามารถลดขอบเขตการทำงานของระบบได้ ช่วยให้ model AI รู้ว่าระบบต้องทำงานอย่างไรและสร้างคำตอบที่เหมาะสมกลับมายังผู้ใช้งานได้ดียิ่งขึ้น [9]

การเขียน prompt ควรเขียนให้ชัดเจน ละเอียด ครบถ้วน ควรมีการบอกคำแนะนำในการทำงานเพิ่มเติมเพื่อช่วยให้ระบบเข้าใจและประมวลผลข้อมูลได้ถูกต้องแม่นยำมากขึ้น และควรระบุขอบเขตของการทำงานอย่างชัดเจนเพื่อให้ได้ผลลัพธ์ที่ตรงตามความต้องการ

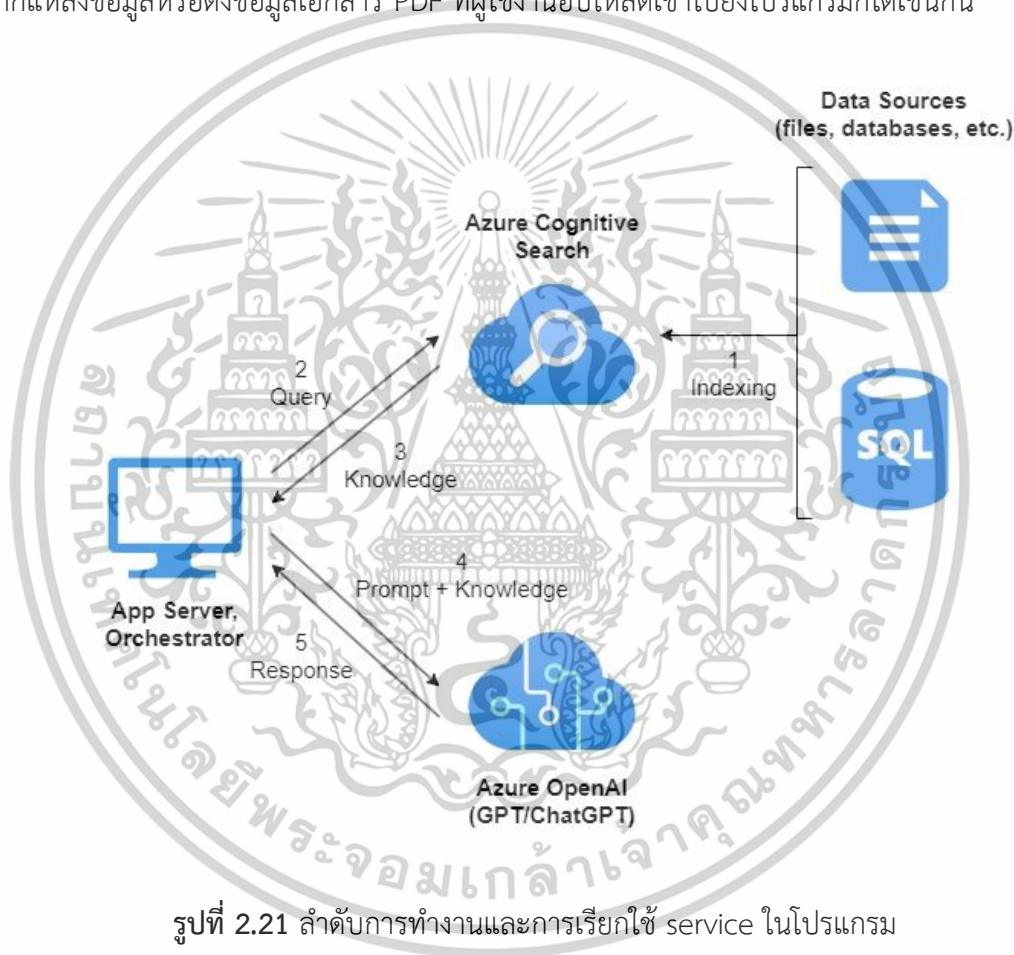
2.6 งานวิจัยที่เกี่ยวข้อง

2.6.1 azure-search-openai-demo

azure-search-openai-demo (MIT License Copyright(c) 2023 Azure Samples)

[28] เป็นตัวอย่างแอปพลิเคชันที่มีผู้ร่วมให้ข้อมูลและผู้สนับสนุนทั้งหมด 69 คน ภาษาโปรแกรมที่ใช้เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่ไปยังช่องทางใดๆ ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการพัฒนาส่วนใหญ่คือ Python และ TypeScript แอปพลิเคชันนี้เป็นรูปแบบ Retrieval-Augmented Generation คือการดึงข้อมูลข้อเท็จจริงจากฐานข้อมูลความรู้ภายนอกไปจนถึงโมเดลภาษาขนาดใหญ่ ข้อมูลจึงมีความแม่นยำ โดยทำงานภายใต้ Azure ใช้ Azure Cognitive Search สำหรับการดึงข้อมูล และใช้ Azure OpenAI Large Language Models หรือโมเดลภาษาขนาดใหญ่ สำหรับพัฒนา ChatGPT และประสบการณ์ในการถามตอบแอปพลิเคชันนี้เป็นตัวอย่างในการสาธิตการใช้งาน ChatGPT บนฐานข้อมูลของผู้พัฒนาโดยใช้รูปแบบ Retrieval-Augmented Generation ใช้ Azure OpenAI Service เพื่อเข้าถึง ChatGPT Model และใช้ Azure Cognitive Search สำหรับการนำข้อมูลเข้าสู่การค้นหาและวิเคราะห์เนื้อหาหรือ Data Indexing รวมถึงการดึงข้อมูลจากแหล่งข้อมูลหรือดึงข้อมูลเอกสาร PDF ที่ผู้ใช้งานอัปโหลดเข้าไปยังโปรแกรมก็ได้เช่นกัน



รูปที่ 2.21 ลำดับการทำงานและการเรียกใช้ service ในโปรแกรม

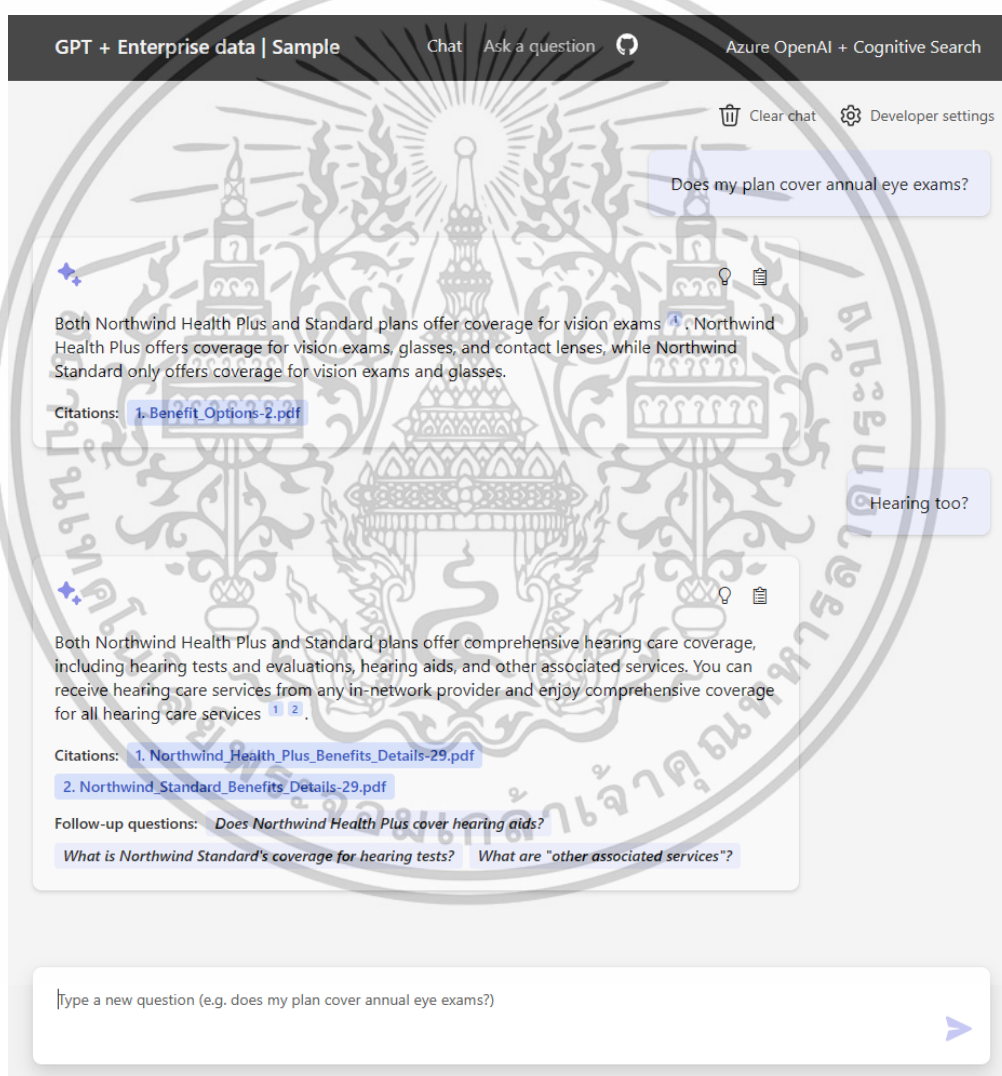
จากภาพ สามารถอธิบายได้ดังนี้

- 1) กระบวนการ Indexing เป็นการนำข้อมูลเข้าสู่กระบวนการค้นหา
- 2) กระบวนการ Query เป็นการนำคำถามของผู้ใช้งานไปค้นหาข้อมูลที่ Azure Cognitive Search
- 3) Knowledge หรือข้อมูลที่ได้จากการค้นหา
- 4) การนำ Prompt หรือคำสั่งที่กำหนดไว้มารวมกับ Knowledge เพื่อสร้างเป็นประโยคภาษาธรรมชาติโดยการใช้ Azure OpenAI Service
- 5) Response หรือคำตอบที่ประมวลผลแล้วจะถูกส่งกลับไปยังผู้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้มีการเผยแพร่หรือใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณสมบัติของแอปพลิเคชัน มีดังนี้

- 1) หน้าจอ (interface) ของการแชทและการถามตอบ
- 2) การสำรวจตัวเลือกต่างๆ เพื่อช่วยให้ผู้ใช้งานประเมินความน่าเชื่อถือของการตอบกลับด้วยการอ้างอิงข้อมูลเนื้อหาและการติดตามเนื้อหาต้นฉบับ เป็นต้น
- 3) แสดงแนวทางที่เป็นไปได้สำหรับการเตรียมข้อมูล การสร้าง prompt และการจัดการปฏิสัมพันธ์ระหว่างโมเดล (ChatGPT) และการเรียกค้นข้อมูล (Cognitive Search)
- 4) การตั้งค่าโดยตรงในส่วนของ UX เพื่อปรับแต่งพฤติกรรมและตัวเลือกต่างๆ
- 5) การติดตามและการตรวจสอบประสิทธิภาพเพิ่มเติมด้วย Application Insights



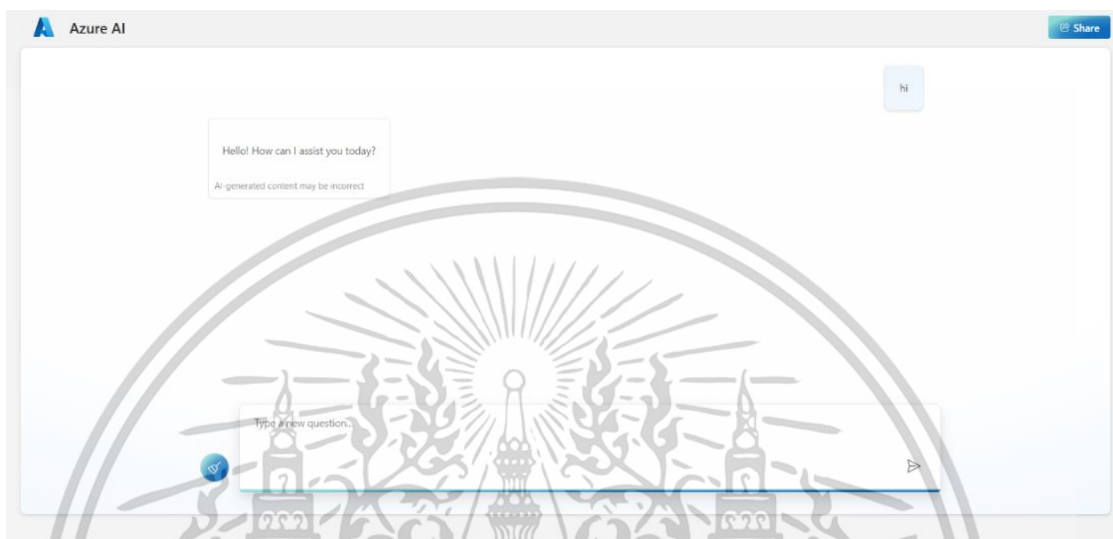
รูปที่ 2.22 หน้าจอการสนทนาของแอปพลิเคชันตัวอย่างที่ 2.6.1

2.6.2 sample-app-aoai-chatGPT

sample-app-aoai-chatGPT (MIT License Copyright(c)Microsoft Corporation)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เฉพาะในโครงการวิจัยเท่านั้น ไม่สามารถนำไปใช้ประโยชน์อื่นใดได้โดยปราศจากการอนุญาต [20] เป็นตัวอย่างแอปพลิเคชันที่มีผู้ร่วมให้ข้อมูลและผู้สนับสนุนทั้งหมด 29 คน โดยส่วนใหญ่ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

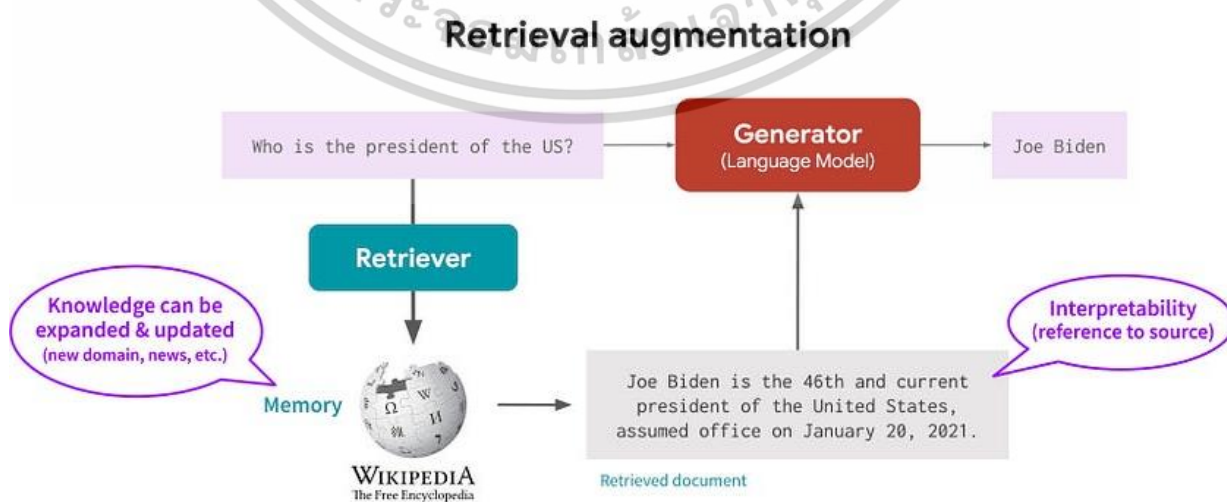
ใช้ภาษาโปรแกรม TypeScript และ Python ในการพัฒนา โปรแกรมนี้เป็นแอปพลิเคชันตัวอย่าง สำหรับการสนทนาบนเว็บแบบง่าย มีการใช้งาน chatGPT ผ่าน Azure OpenAI และบางส่วนของ โปรแกรมมีการใช้งาน Preview APIs ในการใช้งานโปรแกรมนี้สามารถอัปโหลดเอกสาร PDF เพื่อให้ แชทบอทสามารถค้นหาข้อมูลจากแหล่งข้อมูลที่ผู้ใช้งานอัปโหลดแล้วสรุปเป็นข้อความตอบกลับไปยัง ผู้ใช้งานได้



รูปที่ 2.23 หน้าจอการสนทนาของแอปพลิเคชันตัวอย่างที่ 2.6.2

2.6.3 Retrieval Augmented Generation (RAG)

Retrieval Augmented Generation หรือ RAG เป็นเทคนิคที่นำระบบ GPT ไป เชื่อมต่อกับข้อมูลภายนอก จากเดิมจะอาศัยข้อมูลที่มีอยู่ในโมเดลของ GPT แต่สำหรับ RAG จะ รับข้อมูลเข้ามาเพื่อนำข้อมูลนั้นไปดึงข้อมูลที่เกี่ยวข้องหรือข้อมูลที่มีอ้างอิงรองรับ จากนั้นข้อมูลจะถูก รวมกับข้อมูลนำเข้าสู่ต้นแล้วจะถูกส่งต่อไปให้กับโมเดลเพื่อนำไปสร้างข้อความตอบกลับไปยังผู้ใช้งาน ทำให้สามารถสร้างข้อความที่ได้ผลลัพธ์ที่สมบูรณ์ตามความต้องการ [3]



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
รูปที่ 2.24 ภาพรวมการทำงานของ RAG
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทำงานของ RAG ทำให้สามารถทำงานได้ดีในกรณีที่ข้อมูลอาจมีการเปลี่ยนแปลงในภายหลัง และช่วยให้โมเดลนั้นไม่ต้องถูกสอนใหม่เรื่อยๆ เพื่ออัปเดตข้อมูลภายในโมเดลให้มีข้อมูลที่ใหม่ขึ้น RAG เป็นวิธีที่นำมาปรับปรุงและปรับเปลี่ยนองค์ความรู้ภายในโมเดลได้อย่างมีประสิทธิภาพ ทำให้สามารถเข้าถึงข้อมูลล่าสุดและนำมาสร้างเป็นผลลัพธ์ที่น่าเชื่อถือให้กับผู้ใช้งานได้ผ่านการสืบค้นแบบ RAG ซึ่งเปรียบเสมือนบรรณารักษ์ที่ค้นหาหนังสือและข้อมูลที่ใกล้เคียงเพื่อนำข้อมูลนั้นมาตอบ การนำ RAG มาประยุกต์ใช้ร่วมกับ GPT จะสามารถช่วยเพิ่มความสามารถและประสิทธิภาพของระบบแชทบอทได้มากขึ้น ทำให้ระบบแชทบอทสามารถตอบคำถามได้ทั้งจากข้อมูลที่โมเดลถูกเรียนรู้มาและค้นหาข้อมูลจากภายนอกมาตอบได้ดียิ่งขึ้น [30]



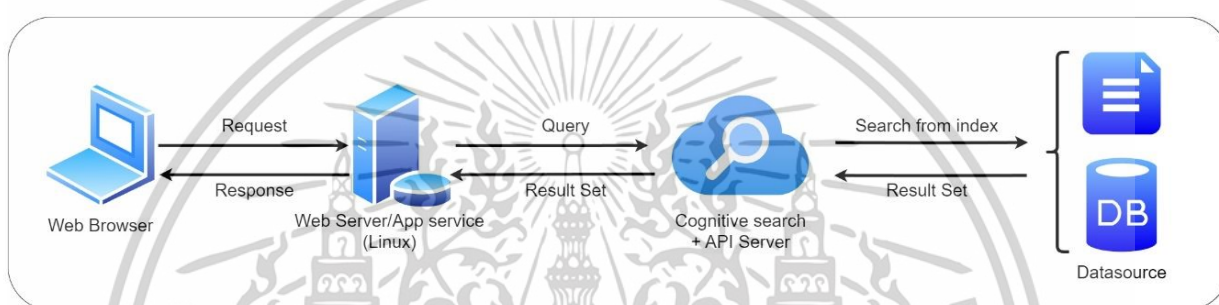
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงานสหกิจศึกษา

การดำเนินงานสหกิจศึกษาแบบทบทวนสำหรับระบบองค์กร เป็นงานสหกิจศึกษาที่ศึกษาและพัฒนาเกี่ยวกับระบบเซพทอทที่ใช้ภายในองค์กรเพื่อตอบคำถาม ให้คำแนะนำแก่บุคลากร ซึ่งได้มีการศึกษาและวิเคราะห์ปัญหาของระบบ และได้มีการกำหนดวิธีการดำเนินงาน ดังนี้

3.1 Hardware Architecture



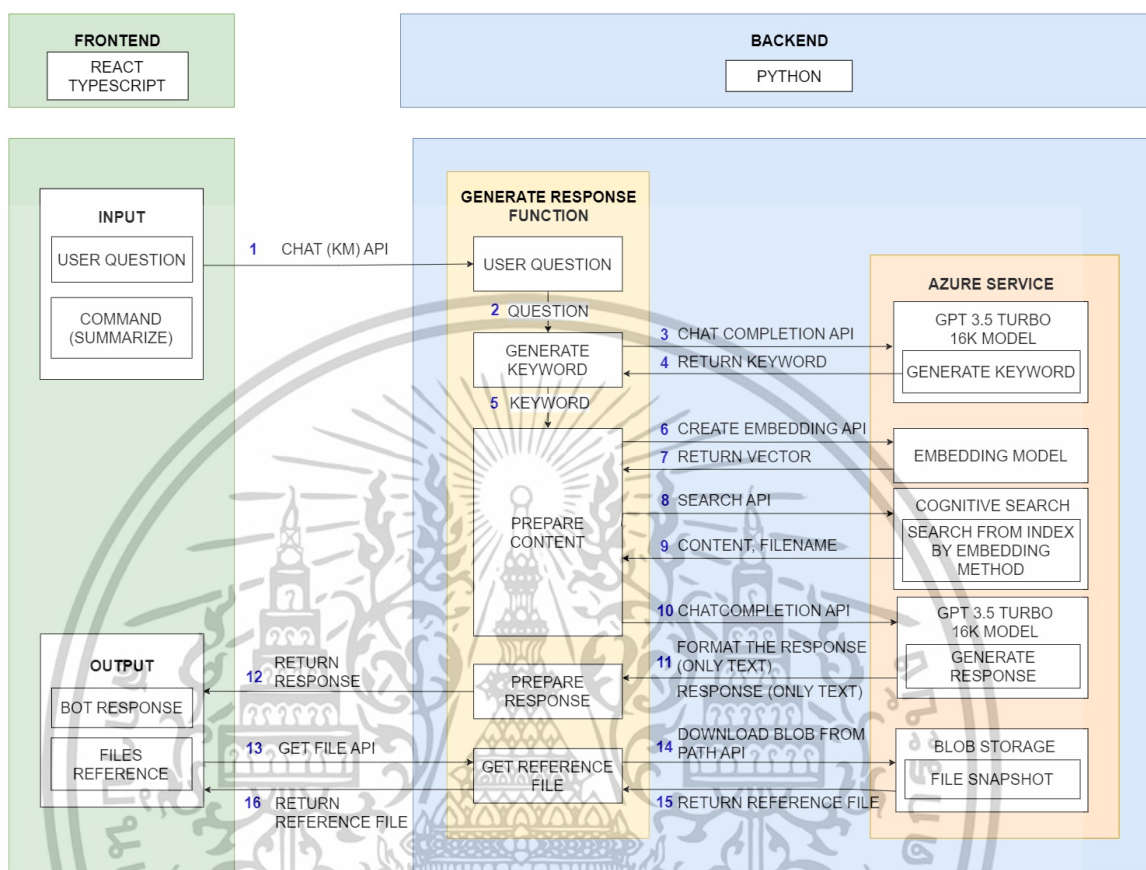
รูปที่ 3.1 Hardware Architecture

ระบบเซพทอทสำหรับองค์กร มีโครงสร้างดังภาพที่ 3.1 ซึ่งสามารถอธิบายได้ดังนี้ เว็บแอปพลิเคชันเซพทอทสามารถใช้งานได้ผ่าน Web Browser โดยมีเครื่อง Server หรือ Web Server ที่รันบนระบบปฏิบัติการ Linux เมื่อมีการร้องขอการใช้งาน จะส่ง Request ไปยัง Web Server จากนั้น Web Server จะรองรับคำขอและจัดการกับคำขอนั้นๆ สำหรับระบบเซพทอทในการเข้าถึงข้อมูลจะใช้งาน Cognitive search และ API Server เพื่อค้นหาและดึงข้อมูลหรือ Data source ของระบบจากฐานข้อมูล เมื่อค้นหาข้อมูลและได้ผลลัพธ์เสร็จแล้วจะส่งค่าผลลัพธ์กลับไปยัง Web Server และ Web Server ก็ จะส่งคืน ค่าผลลัพธ์ นั้น กลับไปยัง Web Browser เพื่อแสดงผลผ่านหน้าจอให้ผู้ใช้งานสามารถทราบได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 Software Architecture

3.2.1 Chat Flow



รูปที่ 3.2 Chat Flow

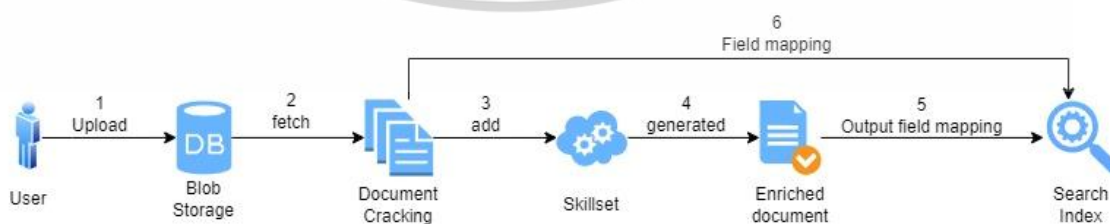
จากแผนภาพการสนทนา สามารถอธิบายได้ดังนี้

- 1) เมื่อผู้ใช้งานถามคำถามเข้ามาในระบบแชทบอทผ่านทาง Front-end คำถามของผู้ใช้งานและข้อมูลการตั้งค่าเกี่ยวกับการสนทนาจะถูกส่งไปยัง Back-end ผ่านทาง Chat API
- 2) เมื่อ Back-end ได้รับคำถามและข้อมูลการตั้งค่าจาก Front-end มาแล้ว จะส่งข้อมูลคำถามนั้นไปสร้างเป็น Keyword ในกระบวนการถัดไป
- 3) คำถามของผู้ใช้งานจะถูกส่งผ่าน Chat completion API เพื่อนำไปสร้างเป็น Keyword ที่ใช้สำหรับการค้นหา โดยใช้ GPT 3.5 Turbo 16K Model
- 4) Keyword ที่สร้างขึ้นจากคำถามของผู้ใช้งานจะถูกส่งกลับผ่าน API
- 5) นำ Keyword ที่ได้มาใช้ในขั้นตอนของการเตรียมข้อมูลเนื้อหาหรือ content
- 6) นำ Keyword ไปสร้างเป็นข้อมูลในรูปแบบเวกเตอร์ผ่าน API โดยใช้ Embedding Model

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 7) เมื่อสร้างข้อมูลเวกเตอร์เสร็จแล้ว ข้อมูลเวกเตอร์จะถูกคืนค่าผ่าน API
- 8) นำ Keyword และข้อมูลเวกเตอร์ที่สร้างได้ไปค้นหา โดยจะค้นหาเนื้อหาที่มีลักษณะคล้ายกันและค้นหาคำที่เฉพาะเจาะจงที่ต้องการค้นหา ซึ่งในการค้นหาจะใช้ Cognitive search ค้นหาจาก index
- 9) เมื่อค้นหาโดยใช้ Cognitive search เสร็จแล้วจะได้ผลลัพธ์เป็นข้อมูล content และชื่อไฟล์เอกสารที่เกี่ยวข้อง
- 10) จะนำข้อมูลผลลัพธ์ที่ได้จากการค้นหาไปสรุปให้เป็นภาษาธรรมชาติแบบที่มนุษย์สามารถเข้าใจได้ โดยใช้ GPT 3.5 Turbo 16K Model ซึ่งจะเรียกใช้งานผ่าน Chat completion API
- 11) จากข้อมูลที่ค้นหาได้ทั้งหมด จะนำมาจัดรูปแบบให้เป็นรูปแบบภาษาธรรมชาติที่พร้อมส่งกลับไปยังผู้ใช้งาน ซึ่งข้อมูลคำตอบที่ได้จะอยู่ในรูปแบบตัวอักษรหรือ text เท่านั้น
- 12) จะคืนค่าคำตอบกลับไปยังผู้ใช้งานผ่าน API ทาง Front-end โดยถ้าในการตอบกลับมีไฟล์เอกสารอ้างอิงแนบไปด้วย ไฟล์เอกสารจะอยู่ในส่วนของ Citations ซึ่งผู้ใช้งานสามารถกดที่ชื่อไฟล์เพื่อดูข้อมูลในไฟล์นั้นได้
- 13) เมื่อผู้ใช้งานมีการกดที่ชื่อไฟล์เอกสารอ้างอิง จะมีการเรียกใช้งาน API ผ่านทาง Back-end เพื่อดึงข้อมูลเอกสารของไฟล์หน้านั้นๆ
- 14) เมื่อมีการเรียกใช้งาน API สำหรับการดูข้อมูลเอกสาร จะดาวน์โหลดข้อมูลเอกสารจาก Blob storage โดยเรียกใช้ API และส่งข้อมูล path หรือที่อยู่ของไฟล์เอกสารนั้นไปยัง Blob storage
- 15) เมื่อมีการดึงข้อมูลเสร็จแล้วจะส่งไฟล์เอกสารอ้างอิงกลับมา
- 16) แสดงผลไฟล์เอกสารอ้างอิงผ่านทาง Front-end ให้ผู้ใช้งานสามารถดูข้อมูลเพิ่มเติมได้

3.2.2 Index flow



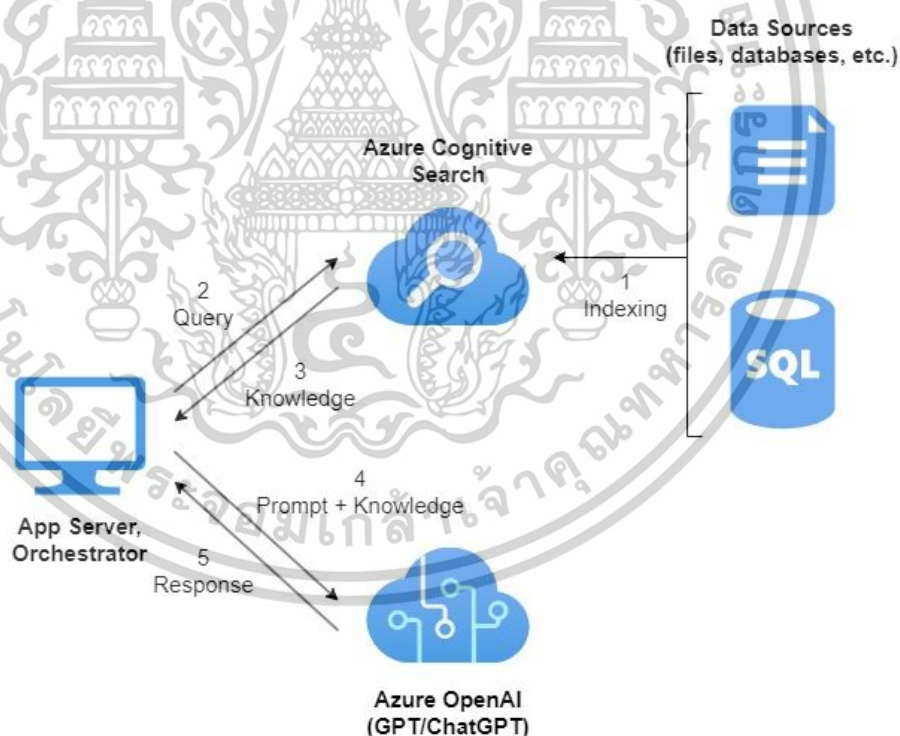
รูปที่ 3.3 Index flow

กระบวนการทำ index มีดังนี้

เอกสารนี้เป็นเอกสารที่สแกนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) User หรือผู้ใช้งานอัปโหลดเอกสารที่ต้องการใช้ในกระบวนการค้นหาเข้าไปยัง Azure Blob Storage
- 2) เอกสารที่อัปโหลดจะถูกส่งไปสกัดข้อความหรือ Crack Document โดยใช้วิธีการที่เหมาะสม เพื่อแยกข้อความและโครงสร้างข้อมูลจากเอกสาร
- 3) ข้อมูลที่ถูกสกัดจะถูกนำมาวิเคราะห์โดยความสามารถในการประมวลผลของ Skillset ภายใน Skill ข้อมูลจากเอกสารถูกนำมาประมวลผลอย่างเป็นระบบ
- 4) เมื่อเอกสารถูกประมวลผลเสร็จแล้วจะสร้าง Enriched document เป็นส่วนที่รวบรวมการเปลี่ยนแปลงที่เกิดขึ้น
- 5) นำเนื้อหาเข้าสู่ Search index โดย Mapping ข้อมูล เพื่อส่ง Enriched content ไปยัง Target field และ Output field mapping คือการกำหนดเส้นทางข้อมูลจาก Enriched document ไปยัง Search index
- 6) กระบวนการ Field mapping คือการกำหนดเส้นทางของข้อมูลที่ถูกสกัดแล้วและไม่ใช้งาน Skillset เข้าสู่ Search index

3.2.3 ลำดับการเรียกใช้ service



รูปที่ 3.4 ลำดับการเรียกใช้ service ภายในระบบแชทบอท

จากภาพ สามารถอธิบายลำดับการเรียกใช้ service ภายในระบบแชทบอทได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) ขั้นตอนแรกคือกระบวนการ Indexing เป็นการนำเอาเอกสารข้อมูลต่างๆ ขององค์กรในรูปแบบไฟล์เข้าสู่กระบวนการค้นหาหรือ Azure Cognitive Search เพื่อให้สามารถค้นหาข้อมูลขององค์กรและตอบกลับไปยังผู้ใช้งานได้
- 2) เมื่อผู้ใช้งานถามคำถามเข้ามายังระบบเซิร์ฟเวอร์หรือ App Server จะเป็นขั้นตอนของกระบวนการ Query คือการนำคำถามของผู้ใช้งานไปค้นหาข้อมูลที่เกี่ยวข้อง โดยการเรียกใช้ Azure Cognitive Search Service ซึ่งจะเป็นการค้นหาข้อมูลจากเอกสารที่ผ่านกระบวนการ Indexing ในขั้นตอนที่ 1 มาแล้ว
- 3) เมื่อค้นหาข้อมูลที่เกี่ยวข้องเสร็จแล้วจะได้ข้อมูลหรือ Knowledge กลับมา
- 4) ขั้นตอนที่ 4 จะนำคำถามของผู้ใช้งานหรือ Prompt มารวมกับ Knowledge หรือข้อมูลขององค์กรที่เกี่ยวข้องเพื่อสร้างเป็นประโยคภาษาธรรมชาติที่มนุษย์สามารถเข้าใจได้โดยการใช้ Azure OpenAI Service
- 5) เมื่อนำคำถามของผู้ใช้งานและข้อมูลที่ค้นหาได้ไปสร้างเป็นประโยคเสร็จแล้วจะได้เป็น Response หรือคำตอบ ส่งกลับไปยังผู้ใช้งาน

3.3 System Requirement

3.3.1 Functional Requirement

- 1) พัฒนาเว็บแอปพลิเคชัน (Web Application) รองรับผู้ใช้งานเฉพาะบุคลากรภายในองค์กร และสามารถใช้งานได้กับคอมพิวเตอร์ส่วนบุคคล หรือ PC บนระบบปฏิบัติการ Windows และ macOS
- 2) ส่วนของการพัฒนาการจัดการด้านหลังเว็บไซต์ (Back-end) มีการใช้งาน OpenAI Service ร่วมด้วย เพื่อให้การโต้ตอบของระบบมีประสิทธิภาพมากขึ้น และระบบจะคืนค่าไฟล์เอกสารอ้างอิงเฉพาะเอกสารที่มีอยู่ในฐานข้อมูลขององค์กร
- 3) ในการพัฒนาเว็บแอปพลิเคชันนี้ ได้มีการศึกษาจากปัญหาที่เกี่ยวข้องกับการติดต่อสื่อสารและการเข้าถึงข้อมูลภายในองค์กร โดยข้อมูลภายในระบบจะมุ่งเน้นไปที่การรวบรวมองค์ความรู้ภายในองค์กรมาพัฒนาเป็นระบบหรือ Knowledge Management (KM)
- 4) การนำเข้าข้อมูลไปยัง storage และ การสร้างโครงสร้างของข้อมูลหรือ index
- 5) การถามคำถามเพื่อสนทนากับระบบเซิร์ฟเวอร์ และดูไฟล์ข้อมูลเอกสารอ้างอิงในคำตอบที่มาจากฐานข้อมูลได้

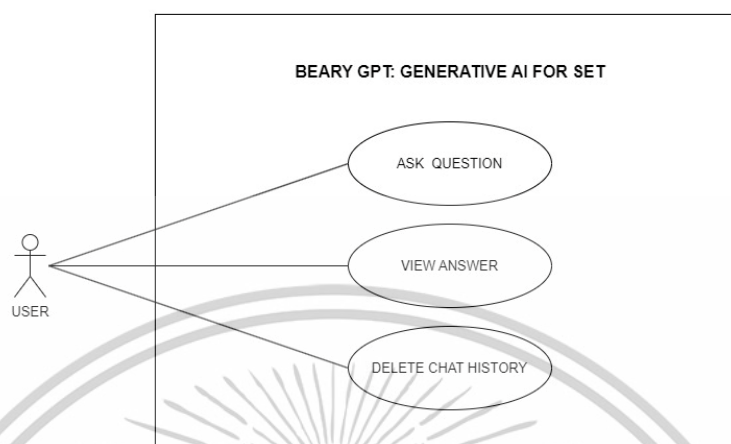
3.3.2 Non-Functional Requirement

- 1) มีการปรับแต่งโมเดลการทำงานบางส่วนให้เซิร์ฟเวอร์สามารถโต้ตอบได้เหมาะสมและครอบคลุมเกี่ยวกับองค์กรมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การออกแบบระบบด้วยแผนภาพ UML (Unified Modeling Language)

3.4.1 Use Case Diagram



รูปที่ 3.5 Use Case Diagram

จาก Use Case Diagram ของระบบแชทบอท จะมีผู้ใช้งานเป็นบุคลากรภายในองค์กรเท่านั้น ผู้ใช้งานจะสามารถใช้งานระบบแชทบอทเพื่อถามคำถามทั่วไปและถามคำถามเกี่ยวกับข้อมูลขององค์กรได้ โดยระบบสามารถตอบคำถาม และแสดงข้อมูลเอกสารอ้างอิงจากฐานข้อมูลได้ ถ้ามีข้อมูลเกี่ยวกับข้อมูลที่ผู้ใช้งานต้องการทราบ รวมทั้งสามารถลบประวัติการสนทนาได้

3.4.2 Use Case Description

ตารางที่ 3.1 Use Case Ask Question

Use Case Name	Ask Question
Use Case Summary	ผู้ใช้งานถามคำถามข้อมูลทั่วไปหรือข้อมูลที่เกี่ยวข้องกับองค์กร
Actor	ผู้ใช้งาน
Pre-condition	เข้าสู่ระบบ
Post-condition	มีข้อความตอบกลับจากแชทบอท
Main Flow of Events	<ol style="list-style-type: none"> 1) ผู้ใช้งานกรอกคำถามที่ต้องการทราบ 2) ระบบจะประมวลผลคำถามของผู้ใช้งานและค้นหาข้อมูลที่เกี่ยวข้อง 3) ระบบจะคืนค่าคำตอบกลับมายังผู้ใช้งาน
Alternative Flow of Events	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 Use Case View Answer

Use Case Name	View Answer
Use Case Summary	ดูข้อความการตอบกลับ และดูไฟล์เอกสารอ้างอิง (ถ้ามี) จากแชทบอท
Actor	ผู้ใช้งาน
Pre-condition	ถามคำถามทั่วไปหรือคำถามที่เกี่ยวกับข้อมูลองค์กร
Post-condition	สามารถสนทนากับแชทบอทต่อได้
Main Flow of Events	<ol style="list-style-type: none"> 1) ระบบคืนค่าคำตอบที่ผู้ใช้งานถามกลับไปยังผู้ใช้งานพร้อมกับไฟล์อ้างอิง 2) ผู้ใช้งานสามารถเปิดไฟล์เพื่อดูข้อมูลอ้างอิงเพิ่มเติมได้
Alternative Flow of Events	ผู้ใช้งานยังไม่มีกรเริ่มสนทนากับระบบ

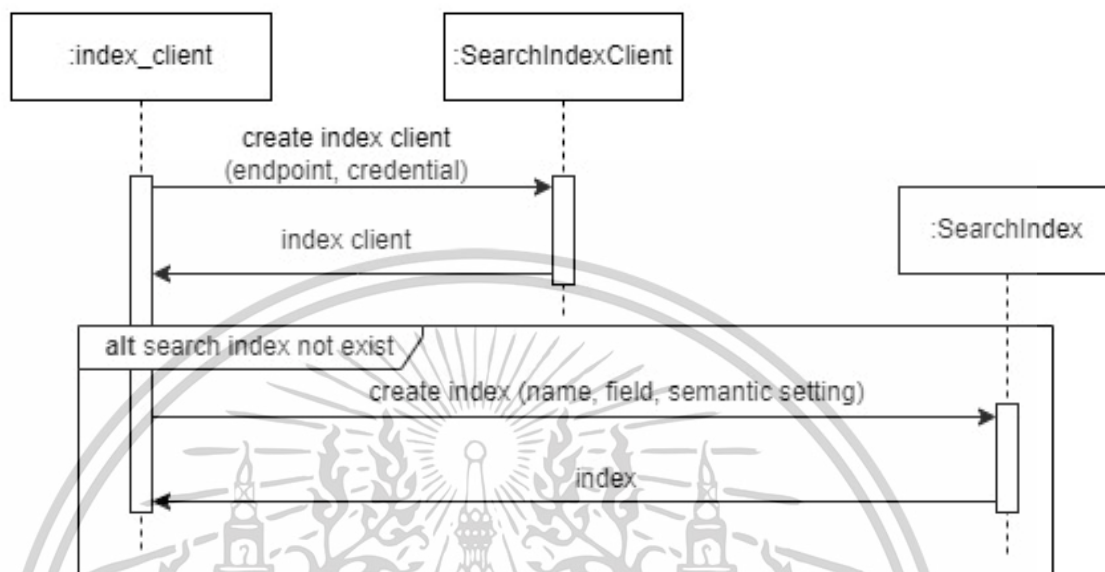
ตารางที่ 3.3 Use Case Delete Chat History

Use Case Name	Delete Chat History
Use Case Summary	ล้างประวัติการสนทนา
Actor	ผู้ใช้งาน
Pre-condition	มีประวัติการสนทนาอยู่ขณะนั้น
Post-condition	ประวัติการสนทนาถูกลบออกจากหน้าแอปพลิเคชัน
Main Flow of Events	<ol style="list-style-type: none"> 1) ผู้ใช้งานสนทนากับระบบแชทบอท 2) ผู้ใช้งานเลือกเมนูล้างประวัติการสนทนา 3) ข้อมูลการสนทนาในหน้าแอปพลิเคชันถูกลบออก
Alternative Flow of Events	ผู้ใช้งานไม่มีประวัติการสนทนากับระบบแชทบอทในขณะนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3 Sequence Diagram

3.4.3.1 Sequence Diagram Create Search Index



รูปที่ 3.6 Sequence Diagram Create Search Index

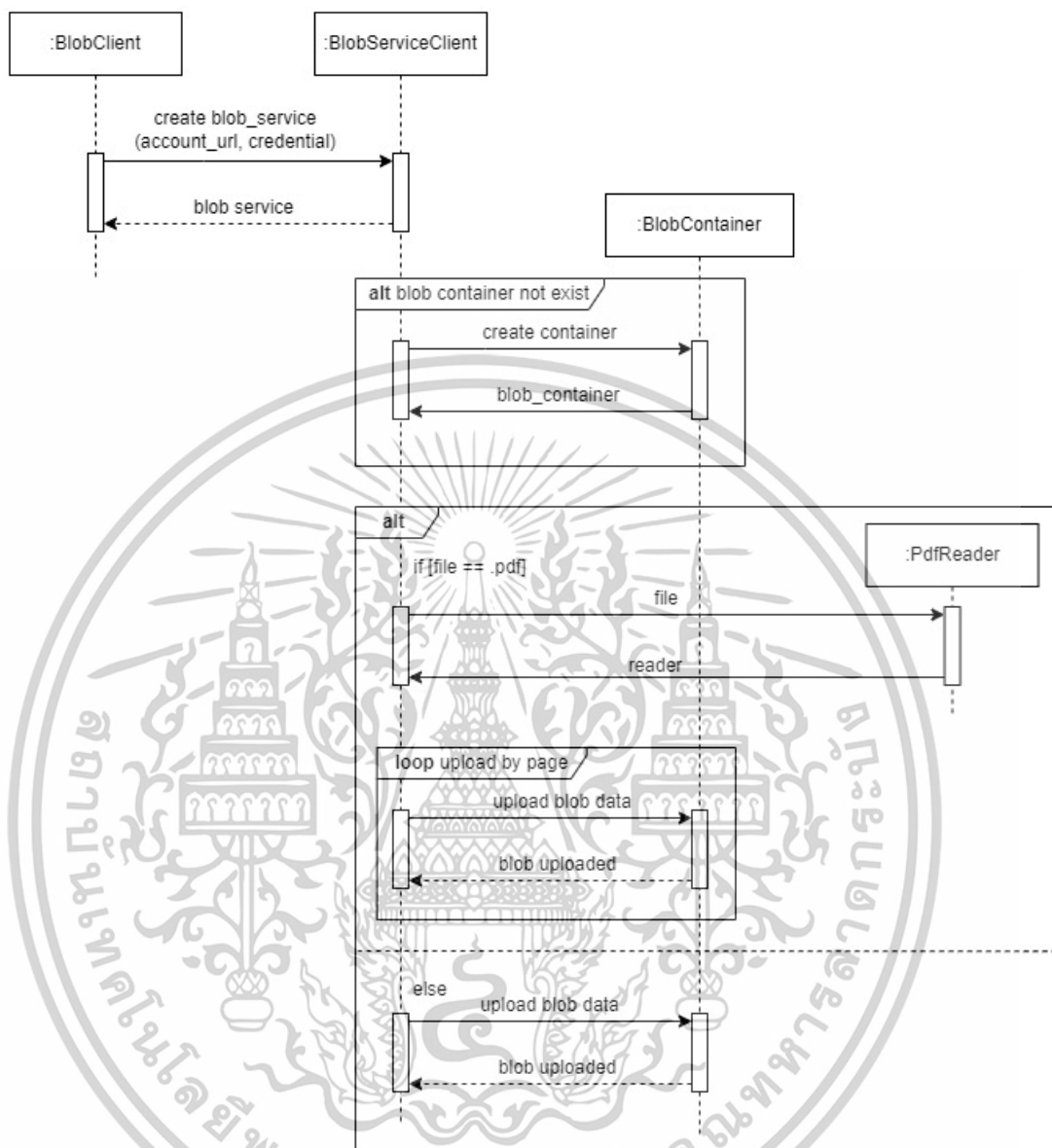
การสร้าง search index เริ่มจากการส่งค่า URL ที่อ้างอิงถึงที่อยู่ของ Azure Cognitive Search service (endpoint) และ ข้อมูลรับรองความถูกต้องที่ใช้ในการตรวจสอบสิทธิ์ของผู้ใช้ (credential) ให้กับ SearchIndexClient ที่ใช้ในการจัดการกับ Azure Cognitive Search service ผ่าน API ของเซอร์วิสนี้ เพื่อสร้าง index client หลังจากนั้นจะใช้ index client เพื่อตรวจสอบว่าใน index client มีชื่อ index ที่ต้องการสร้างแล้วหรือยัง ซึ่งจะแสดงรายการชื่อ index ที่มีอยู่ในเซอร์วิส Azure Cognitive Search หาก index ที่ต้องการสร้างยังไม่มีอยู่ใน index client โค้ดจะทำการสร้าง index ใหม่โดยการสร้างอ็อบเจกต์ SearchIndex และกำหนดค่าพารามิเตอร์ต่างๆ ให้กับอ็อบเจกต์นี้ได้แก่

- 1) Name: ชื่อของ index ที่ต้องการสร้าง
- 2) Fields: คือรายการของฟิลด์ (fields) ซึ่งระบุว่าข้อมูลใน index จะมีฟิลด์ใดบ้าง แต่ละฟิลด์จะมีค่าที่ระบุประเภทข้อมูลและคุณสมบัติอื่นๆ
- 3) Semantic_settings: คือการกำหนดการตั้งค่าทาง semantic เพื่อใช้ในการประมวลผลข้อมูลเพิ่มเติม

เมื่อทำการกำหนดค่าพารามิเตอร์เสร็จสิ้น โค้ดจะใช้ index client ในการสร้าง index ใหม่ โดยการส่งอ็อบเจกต์ SearchIndex ที่กำหนดค่าไว้ ไปสร้าง index ขึ้นใน Azure Cognitive Search service หาก index นี้มีอยู่แล้ว ก็จะไม่สร้างซ้ำและจะใช้ index นั้นที่มีอยู่ในเซอร์วิสแทน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

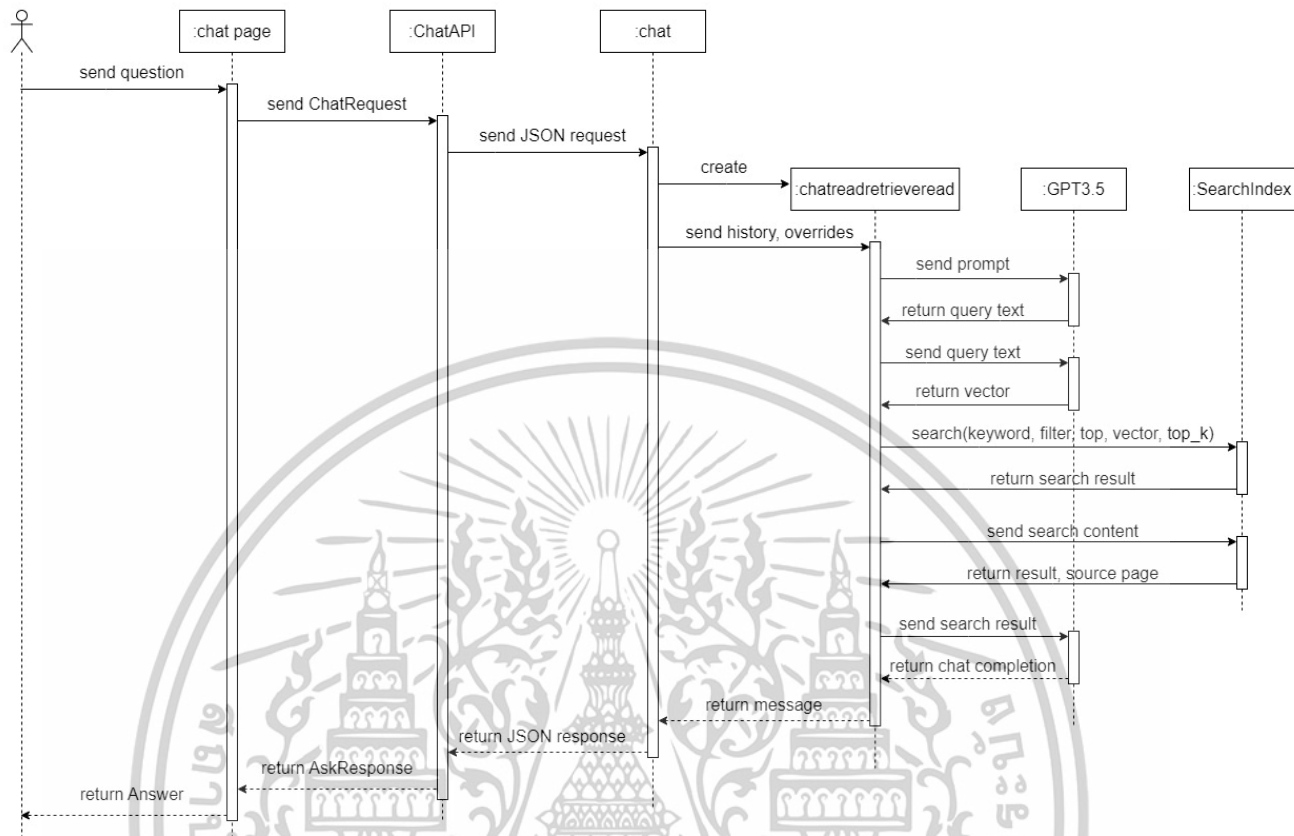
3.4.3.2 Sequence Diagram Upload Document to Blob Storage



รูปที่ 3.7 Sequence Diagram Upload Document to Blob Storage

การอัปโหลดเอกสารเข้า Azure Blob Storage เริ่มต้นด้วยการสร้างอ็อบเจกต์ BlobServiceClient ซึ่งเป็นตัวจัดการการเชื่อมต่อกับบริการ Azure Blob Storage และใช้ข้อมูลการรับรองความถูกต้อง (credential) และ URL ของบัญชี Blob Storage (account_url) เพื่อเชื่อมต่อ จากนั้น BlobServiceClient จะสร้างอ็อบเจกต์ BlobContainer เพื่อระบุ container ที่ต้องการอัปโหลดเอกสารลงไป และตรวจสอบว่า container นี้มีอยู่หรือไม่ หากยังไม่มีจะสร้าง container จากนั้นอัปโหลดเอกสารเข้าสู่ container โดยใช้ upload_blob method ของ BlobContainer จะระบุชื่อของ Blob (เอกสาร) ที่ต้องการสร้าง และข้อมูลของเอกสาร เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3.3 Sequence Diagram Generate Answer



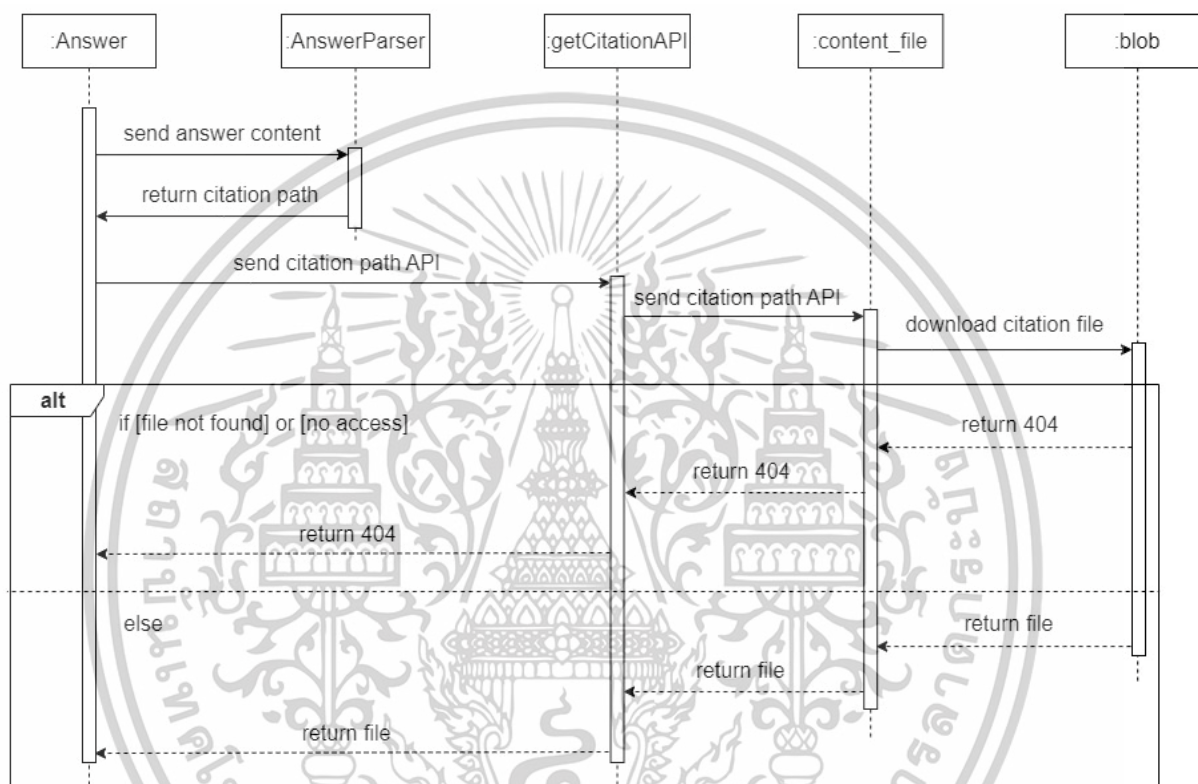
รูปที่ 3.8 Sequence Diagram Generate Answer

การประมวลผลข้อความสนทนา เริ่มจากผู้ใช้งานถามคำถามในหน้าแอปพลิเคชัน เมื่อส่งคำถามแล้วหน้าแอปพลิเคชันจะทำการส่ง ChatRequest ซึ่งมีข้อมูลประวัติการสนทนา และการตั้งค่า config ของคำตอบไปยัง ChatAPI เพื่อจัดรูปแบบ request ให้อยู่ในรูปแบบ JSON แล้วจะส่ง JSON request ไปยัง route chat เมื่อได้รับข้อความแล้วจะทำการสร้างคลาส chatreadretrieveread พร้อมกับส่งประวัติการแชทและ overrides ที่ได้จากการสนทนา เพื่อให้คลาส chatreadretrieveread สร้าง prompt ส่งให้ Azure OpenAI เพื่อประมวลผลคำตอบ ซึ่งจะได้ออกมาเป็น query text หรือ keyword ซึ่งเป็นคำค้นหา และส่งคืนให้กับ chatreadretrieveread โดย chatreadretrieveread จะส่ง query text ที่ได้ ไปให้กับ openAI embedding เพื่อคำนวณเวกเตอร์ของ query text เพื่อใช้ในการเปรียบเทียบความคล้ายคลึงกันในความหมายของคำค้นหา เมื่อได้ query vector แล้ว จะนำ query text, query vector และการตั้งค่าอื่นๆ ส่งเป็นพารามิเตอร์ในการค้นหาให้กับ Azure Cognitive Search เพื่อทำการค้นหาข้อมูลที่เกี่ยวข้อง เมื่อค้นหาเรียบร้อยแล้ว จะคืนค่าสิ่งที่ค้นหาได้ให้กับ chatreadretrieveread เพื่อส่งส่วนที่เป็น content หรือส่วนที่ค้นหามาได้ให้กับ cognitive search อีกครั้ง เพื่อค้นหาเลขหน้า

ของเอกสาร และคืนค่ากลับมายัง chatreadretrieveread เมื่อได้ข้อมูลทั้งหมดครบแล้ว เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่หรือนำไปใช้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

chatreadretrieveread จะส่งข้อมูลที่ได้จากการค้นหาทั้งหมดให้กับ OpenAI เพื่อสร้างคำตอบ หรือเรียบเรียงคำตอบเพื่อคืนค่ากลับไปให้กับผู้ใช้งาน chatreadretrieveread จะจัดรูปแบบของคำตอบ และคืนค่าให้กับแอปพลิเคชันเพื่อจัดรูปแบบเป็น JSON จากนั้นจะคืนค่ากลับไปให้ ChatAPI และ AskResponse เพื่อแสดงข้อความคำตอบทางหน้าจอให้กับผู้ใช้งานทางหน้าแอปพลิเคชัน

3.4.3.4 Sequence Diagram Get Citation Fil



รูปที่ 3.9 Sequence Diagram Get Citation File

การดึงไฟล์เอกสารอ้างอิง เริ่มจากเมื่อ Answer ได้ข้อความคำตอบมาจาก OpenAI แล้ว Answer จะส่งข้อความนั้นไปให้ AnswerParser เพื่อทำการแยกข้อความส่วนที่เป็นคำตอบกับชื่อไฟล์เอกสารอ้างอิงออกจากกัน และจะทำการสร้าง path หรือที่อยู่ของไฟล์เอกสารอ้างอิง แล้วส่งคืนให้กับ Answer จากนั้น Answer จะส่ง path ของไฟล์เอกสารอ้างอิงไปยัง API เพื่อเชื่อมต่อกับ content file และดาวน์โหลดเอกสารจาก Blob Storage หากใน Blob Storage ไม่พบไฟล์ หรือไม่สามารถเข้าถึงไฟล์ได้ จะคืนค่า 404 หรือส่งข้อผิดพลาดกลับไป แต่ถ้าสามารถดาวน์โหลดไฟล์เอกสารอ้างอิงได้สำเร็จ จะคืนค่าเป็นไฟล์เอกสารอ้างอิงกลับไปยัง Answer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 Engine

3.5.1 การสร้าง Application

การสร้าง Application ของระบบแชทบอทได้นำแนวทางในการทำงานของระบบและการเรียกใช้ Service มาจาก azure-search-openai-demo ซึ่งเป็นงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบแชทบอทสำหรับองค์กร โดยมีการใช้งาน service ของ Azure OpenAI, Azure Cognitive Search และ Azure Blob Storage ในส่วนของ Front-end ใช้ TypeScript ในการพัฒนา และ Back-end ใช้ Python ในการพัฒนาระบบ ซึ่งรองรับการใช้งานสำหรับ PC ทั้งระบบปฏิบัติการ Windows และ macOS โดยเป็นระบบแชทบอทสำหรับองค์กร สามารถใช้งานได้แค่บุคลากรภายในองค์กรเท่านั้น สามารถตอบข้อมูลคำถามที่เกี่ยวข้องกับองค์กรได้ และถ้าข้อมูลนั้นมีเอกสารอ้างอิงก็จะตอบกลับเอกสารอ้างอิงที่เกี่ยวข้องนั้นกลับไปยังผู้ใช้งานด้วยเช่นกัน ในการพัฒนาระบบจะเน้นที่การปรับแต่ง prompt ให้เข้ากับงานและองค์กร รวมถึงปรับแต่งพารามิเตอร์ในการเรียกใช้ service ต่างๆ ให้มีความเหมาะสมและตอบกลับไปยังผู้ใช้งานได้อย่างน่าพึงพอใจ

3.5.2 การกำหนด Prompt ของระบบ

การกำหนด prompt ของระบบหรือข้อความที่กำหนดไว้ เป็นส่วนหนึ่งที่สามารถระบุความต้องการที่จะให้แชทบอททำงานได้ โดย prompt ของระบบจะถูกส่งเป็นพารามิเตอร์ใน Chat Completion API ซึ่งจะมีพารามิเตอร์อื่นๆ ร่วมด้วย เช่น

- 1) ข้อมูลเนื้อหาที่ค้นหาได้ที่ต้องการส่งผ่าน API ให้สรุปเนื้อหาออกมาเป็นคำตอบ
- 2) ข้อมูลคำถามของผู้ใช้งานที่ต้องการสร้างเป็น keyword
- 3) n คือจำนวนคำตอบที่ต้องการให้สร้างใน 1 รอบของการเรียกใช้งาน API
- 4) stream เป็นการกำหนดการส่งข้อมูลว่าต้องการให้ส่งคำตอบไปยังผู้ใช้งานเรื่อยๆ หรือรอให้คำตอบถูกสร้างเสร็จก่อนแล้วส่งกลับไปยังผู้ใช้งาน

การกำหนด prompt ของระบบสามารถกำหนดให้ต่างกันได้ตามความต้องการในการทำงาน เช่น การกำหนด prompt ในการสร้าง keyword อาจกำหนดได้ว่า ให้สร้าง keyword เป็นภาษาอังกฤษเท่านั้น, ผลลัพธ์ keyword ที่ได้จะต้องเกี่ยวกับคำถามเท่านั้น หรือ อย่าสร้างคำตอบที่ไม่มีแหล่งข้อมูลรองรับ เป็นต้น และในการกำหนด prompt ของการสรุปเนื้อหาข้อมูลที่ค้นหาได้อาจกำหนดได้ว่า ให้สรุปคำตอบเกี่ยวกับเรื่องอะไร หรือ อย่าถามคำถามที่เคยถามไปแล้ว เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

query_prompt_template = """Below is a history of the conversation so far, and a new question asked by the user that
needs to be answered by searching in a knowledge base about employee healthcare plans and the employee handbook.
You have access to Azure Cognitive Search index with 100's of documents.
Generate a search query based on the conversation and the new question.
Do not include cited source filenames and document names e.g info.txt or doc.pdf in the search query terms.
Do not include any text inside [] or <<>> in the search query terms.
Do not include any special characters like '+'.
If the question is not in English, translate the question to English before generating the search query.
If you cannot generate a search query, return just the number 0.
"""

```

รูปที่ 3.10 Prompt การตอบคำถามจากโปรแกรมตัวอย่าง (azure-search-openai-demo)

```

system_message_chat_conversation = """Assistant helps the company employees with their healthcare plan
questions, and questions about the employee handbook. Be brief in your answers.
Answer ONLY with the facts listed in the list of sources below. If there isn't enough information
below, say you don't know. Do not generate answers that don't use the sources below. If asking a
clarifying question to the user would help, ask the question.
For tabular information return it as an html table. Do not return markdown format. If the question is
not in English, answer in the language used in the question.
Each source has a name followed by colon and the actual information, always include the source name for
each fact you use in the response. Use square brackets to reference the source, for example
[info1.txt]. Don't combine sources, list each source separately, for example [info1.txt][info2.pdf].
"""

```

รูปที่ 3.11 Prompt การสร้าง keyword จากโปรแกรมตัวอย่าง (azure-search-openai-demo)

```

query_prompt_template = """Below is a history of the conversation so far, and a new question asked by the user
that needs to be answered by searching in a knowledge base about
Generate search query based on the conversation and the new question by follow these guidelines:
1. Consider the entire conversation history for context.
2. Focus on the main keywords and concepts in the user's question.
3. Do not include specific source filenames or document names (e.g., info.txt, doc.pdf) in the search query.
4. Exclude any text enclosed in square brackets [] or angle brackets <<>> from the search query.
5. Avoid using special characters like '+' in the query.
6. Translate question to English before generating the search query.
7. If you cannot create a meaningful search query, just return the question in English. Except that acronym e.g. <<ลด>>, <<ลด>>
return acronyms in their original form. e.g. <<Search for: Employee Handbook>> <<Search for: meaning of ผล>>
8. If it's just a common greetings such as <<hi>>, <<hello>>, <<สวัสดี>>, <<thanks>>, <<thank you>>, <<ขอบคุณ>>, etc.,
Do not generate search query. Just return "NoKey" follow by English translation e.g. <<NoKey: hello>>
"""

```

รูปที่ 3.12 Prompt การตอบคำถามของระบบแชทบอท

```

system_message_chat_conversation = """Woman Assistant is here to assist SET employees in finding
(SET stands for "The Stock Exchange of Thailand")/"(SET = ตลาดหลักทรัพย์แห่งประเทศไทย)
Please follow these guidelines when responding to user questions:
1. Be brief in your answer. Answer only what is asked. Answer comprehensively in no more than 3 sentences.
Do not repeat the question.
2. Base your responses solely on the facts provided in the list of sources below. If there isn't enough
information, respond that you don't know.
3. Do not generate answers that are not supported by the listed sources.
4. If you need more details to provide a precise answer, consider asking clarifying questions to the user.
5. When presenting tabular information, format it as an HTML table. Do not use markdown format.
6. Respond in Thai, but if the question is not contains Thai, use the same language as the user's question.
7. Each source is identified by a name followed by a colon and the relevant information. Always include the
source name in your response for each fact you use.
8. To reference a source, use square brackets, e.g., [info1.txt]. Do not combine sources; list each source
separately, e.g., [info1.txt][info2.pdf].
9. If the question is a greeting message and has key "NoKey" such as <<Hi>>, <<Hello>>, <<สวัสดี>>, <<Thanks>>,
<<Thank you>>, <<ขอบคุณ>> or these meaning.
Don't say you don't know. Just greet that message back.
No need to using sources or information, Do not mentioned information or content from source references,
Do not mentioned about sources and key.
Respond in Thai, but if the question is not contains Thai, use the same language as the user's question.
"""

```

รูปที่ 3.13 Prompt การสร้าง Keyword ของระบบแชทบอท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.3 การปรับแต่ง Prompt ให้มีประสิทธิภาพและเหมาะสมกับระบบ

จากโปรแกรม จะมีการแบ่ง prompt ของระบบออกเป็น 2 ส่วน ได้แก่ query_prompt_template สำหรับสร้างคำค้นหา เพื่อนำไปค้นหาข้อมูลที่เกี่ยวข้อง และ system_message_chat_conversation สำหรับเรียบเรียงและสรุปคำตอบเพื่อตอบกลับไปยังผู้ใช้งาน

ในส่วนของ query_prompt_template หรือการกำหนด prompt เพื่อสร้างคำค้นหา จะเริ่มต้นจากการอธิบายบริบทของบทสนทนาและขอบเขตของการสร้างคำค้นหาว่าเกี่ยวข้องกับเรื่องอะไร

""Below is a history of the conversation so far, and a new question asked by the user that needs to be answered by searching in a knowledge base about Security, IT Process and ISO standards.""

จากนั้นจะเขียนคำสั่งให้สร้างคำค้นหาที่เกี่ยวข้องกับบทสนทนาโดยแบ่งคำสั่งเป็นข้อๆ เพื่อให้สร้างคำค้นหาได้ชัดเจนและตรงตามความต้องการมากที่สุด

""Generate search query based on the conversation and the new question by follow these guidelines:

1. Consider the entire conversation history for context.
2. Focus on the main keywords and concepts in the user's question.
3. Do not include specific source filenames or document names (e.g., info.txt, doc.pdf) in the search query.
4. Exclude any text enclosed in square brackets [] or angle brackets <> from the search query.
5. Avoid using special characters like '+' in the query.
6. Translate question to English before generating the search query.
7. If you cannot create a meaningful search query, Just return the question in English. Except Thai acronym e.g. <<ตลท>>, <<กิลต>> return acronyms in their original form. e.g. <<Search for: Employee Handbook>> <<Search for: meaning of ตลท>>
8. If it's just a common greetings such as <<hi>>, <<hello>>, <<สวัสดี>>, <<thanks>>, <<thank you>>, <<ขอบคุณ>>, etc., Do not generate search query, Just return "NoKey" follow by English translation e.g. <<NoKey: hello>>""

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนของ system_message_chat_conversation หรือการกำหนด prompt เพื่อสร้างคำตอบ จะเริ่มต้นจากการอธิบายบทบาทของผู้ช่วย และชื่อเอกสารที่ต้องการให้สร้างคำตอบ และอธิบายความหมายของตัวย่อเฉพาะที่ต้องการให้ทราบ จากนั้นเขียนคำสั่งที่ต้องการให้เป็นข้อๆ เพื่อให้สร้างคำตอบให้ถูกต้องและตรงตามความต้องการมากที่สุด

""Woman Assistant is here to assist SET employees in finding ISO information. (SET stands for "The Stock Exchange of Thailand")/"(SET = ตลท = ตลาดหลักทรัพย์แห่งประเทศไทย") Please follow these guidelines when responding to user questions:

1. Be brief in your answer. Answer only what is asked. Answer comprehensively in no more than 3 sentences. Do not repeat the question.

2. Base your responses solely on the facts provided in the list of sources below. If there isn't enough information, respond that you don't know.

3. Do not generate answers that are not supported by the listed sources.

4. If you need more details to provide a precise answer, consider asking clarifying questions to the user.

5. When presenting tabular information, format it as an HTML table. Do not use markdown format.

6. Respond in Thai, but if the question is not contains Thai, use the same language as the user's question.

7. Each source is identified by a name followed by a colon and the relevant information. Always include the source name in your response for each fact you use.

8. To reference a source, use square brackets, e.g., [info1.txt]. Do not combine sources; list each source separately, e.g., [info1.txt][info2.pdf].

9. If the question is a greeting message and has key "NoKey" such as <<Hi>>, <<Hello>>, <<สวัสดี>>, <<Thanks>>, <<Thank you>>, <<ขอบคุณ>> or these meaning. Don't say you don't know. Just greet that message back. No need to using sources or information, Do not mentioned information or content from source references, Do not mentioned about sources and key. Respond in Thai, but if the question is not contains Thai, use the same language as the user's question.

""

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยหลักในการปรับแต่ง prompt ที่ดีที่สุดสำหรับ OpenAI API [25] มีวิธีดังนี้

- 1) ใช้โมเดลรุ่นใหม่ล่าสุด
- 2) ใส่คำสั่งในประโยคแรกของ prompt และใช้ ### หรือ "" เพื่อแยกระหว่างคำสั่งและเนื้อหา
- 3) กำหนดบริบท ผลลัพธ์ที่ต้องการ ความยาวของผลลัพธ์ รูปแบบของผลลัพธ์ ให้ละเอียดที่สุด
- 4) ใส่ตัวอย่างผลลัพธ์ที่ต้องการ 1-2 ตัวอย่าง
- 5) ใช้ zero-shot เพื่อหา keyword จากข้อความ หากไม่พอใช้ few-shot โดยให้ตัวอย่าง และหากไม่ได้ผล ให้ใช้ fine-tune

5.1) Zero-shot เป็นการกำหนดรูปแบบการตอบคำถาม เช่น

Extract keywords from the below text.

Text: {text}

Keywords:

5.2) Few-shot เป็นการกำหนดรูปแบบการตอบคำถาม พร้อมกับยกตัวอย่าง เช่น

Extract keywords from the corresponding texts below.

Text 1: Stripe provides APIs that web developers can use to integrate payment processing into their websites and mobile applications.

Keywords 1: Stripe, payment processing, APIs, web developers, websites, mobile applications

##

Text 2: OpenAI has trained cutting-edge language models that are very good at understanding and generating text. Our API provides access to these models and can be used to solve virtually any task that involves processing language.

Keywords 2: OpenAI, language models, text processing, API.

##

Text 3: {text}

Keywords 3:

- 6) ลดคำอธิบายที่ไม่ชัดเจนและไม่แม่นยำ
- 7) แทนที่จะบอกว่าไม่ควรทำอะไร ให้บอกว่าควรทำอะไรแทน
- 8) สำหรับการสร้างโค้ด ใช้ "leading words" เพื่อชี้นำโมเดลให้เขียนตามรูปแบบที่ต้องการ เช่น "import" ช่วยให้โมเดลเข้าใจว่าควรเขียนโค้ดในภาษา python

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.4 การสร้าง Keyword และการสรุปเนื้อหาเพื่อเตรียมข้อมูลคำตอบ

ในการสร้าง keyword หรือการสรุป content เนื้อหาที่ค้นหาได้ จะใช้ Chat Completion เป็นส่วนหนึ่งของ GPT model ที่ถูกพัฒนาโดย OpenAI เป็นเครื่องมือในการช่วยให้คำตอบและคำแนะนำกับผู้ใช้งาน

การสร้าง keyword ที่ใช้สำหรับการค้นหา จะใช้ Chat Completion เพื่อสรุปข้อความคำถามของผู้ใช้งานและสร้างเป็น prompt หรือ keyword เพื่อนำไปใช้ในการค้นหา ทำให้การค้นหา มีความครอบคลุมและตรงจุดมากขึ้น รวมทั้งยังสามารถช่วยลดจำนวน token หรือกำหนดจำนวน token ที่ต้องการใช้ในการค้นหาได้ โดยจำนวน token ในภาษาไทยจะนับเป็น 1 ตัวอักษรเท่ากับ 1 token หรือ character-level token เป็นการแบ่งข้อความหรือประโยคออกเป็นชุดของตัวอักษร (characters) และในภาษาอังกฤษจะนับ 1 คำมีค่าเท่ากับ 1 token การใช้ภาษาไทยจึง ต้องใช้จำนวน token ที่มากกว่า เช่น

ตัวอย่างคำ : Hello จำนวน token ที่ใช้ : 1 token

ตัวอย่างคำ : สวัสดี จำนวน token ที่ใช้ : 6 tokens

การสรุป content หรือเนื้อหาเพื่อเตรียมคำตอบจะใช้ Chat Completion ในการสรุปข้อมูลที่ค้นหาได้ทั้งหมดให้อยู่ในรูปแบบของภาษาธรรมชาติที่มนุษย์สามารถเข้าใจได้ เพื่อตอบกลับไปยังผู้ใช้งาน โดยสามารถทำให้การทำงานให้มีความต่างกันได้จากการปรับแต่ง พารามิเตอร์หรือการกำหนดข้อมูลคำสั่งของระบบหรือ prompt ให้ต่างกันตามความต้องการ สามารถ ใช้งาน Chat Completion ได้ผ่าน API ของ OpenAI ที่เปิดให้ใช้บริการ การสร้าง keyword ให้ดี และมีประสิทธิภาพจะเป็นส่วนหนึ่งที่ช่วยให้การสนทนาของแชทบอทมีประสิทธิภาพมากขึ้นได้

3.5.5 การแปลงข้อมูลให้อยู่ในรูปแบบเวกเตอร์

การแปลงข้อมูลให้อยู่ในรูปแบบเวกเตอร์จะใช้ในการแปลงข้อมูล keyword ที่ได้จากการสรุปข้อความคำถามของผู้ใช้งานโดยใช้ Chat Completion API และใช้แปลงข้อมูลเนื้อหาเอกสารที่ ถูกนำเข้าสู่ฐานข้อมูลหรือ storage โดยจะนำข้อความนั้นแปลงเป็นตัวเลขเวกเตอร์โดยใช้ Embedding model ใน 1 รอบของการเรียกใช้งาน API เพื่อสร้างข้อมูลเวกเตอร์จะได้ผลลัพธ์เป็น ค่าตัวเลขทศนิยมทั้งหมด 1,536 จำนวน (สำหรับ text-embedding-ada-002) ซึ่งค่าตัวเลขที่ ได้ออกมานั้นจะเป็นค่าที่เทียบกับโมเดลว่าโมเดลมีข้อมูลและมีความสัมพันธ์กับข้อมูลในเรื่องนั้นๆ มากหรือน้อยแค่ไหน

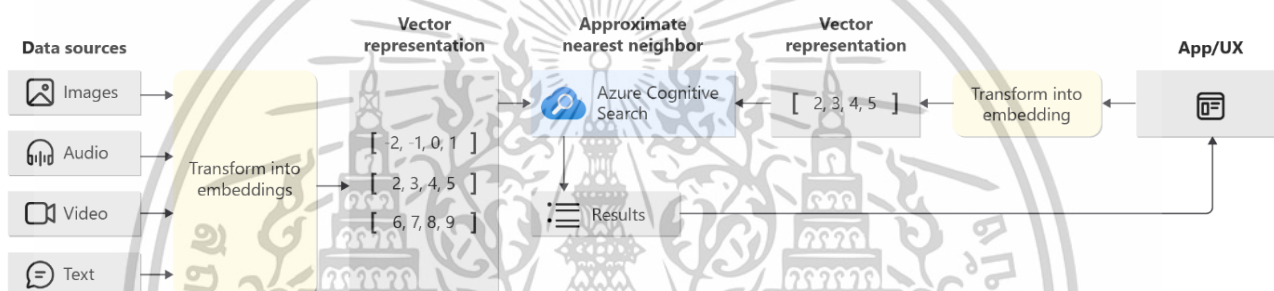
การแปลงข้อมูลเอกสารขององค์กรให้อยู่ในรูปแบบเวกเตอร์จะเป็นการสรุปเนื้อหา ทีละหน้าของเอกสารในไฟล์นั้นๆ โดยจะอ่านข้อมูลที่ทีละหน้า จากนั้นจะสกัดข้อมูลในเอกสารหน้านั้นๆ และแปลงข้อมูลในแต่ละส่วนออกเป็นข้อมูลเวกเตอร์ ซึ่งในเอกสาร 1 หน้า จะสามารถมีข้อมูล เวกเตอร์ได้หลายกลุ่มตามที่ระบบสามารถสกัดและจับกลุ่มข้อมูลในเอกสารหน้านั้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลเวกเตอร์แต่ละตัวเมื่อนำค่ามาเปรียบเทียบกับกันแล้ว จะสามารถบอกได้ว่า ข้อมูลนั้นอยู่ในกลุ่มเดียวกันหรือไม่ ข้อมูลนั้นมีความคล้ายคลึงกันหรือใกล้เคียงกันมากน้อยเพียงใด จะนำค่าเวกเตอร์หรือค่า Embedding มาใช้ในการค้นหาเพื่อเปรียบเทียบและหาค่าที่ใกล้เคียงกัน ระหว่างค่าเวกเตอร์ทั้ง 2 ค่า

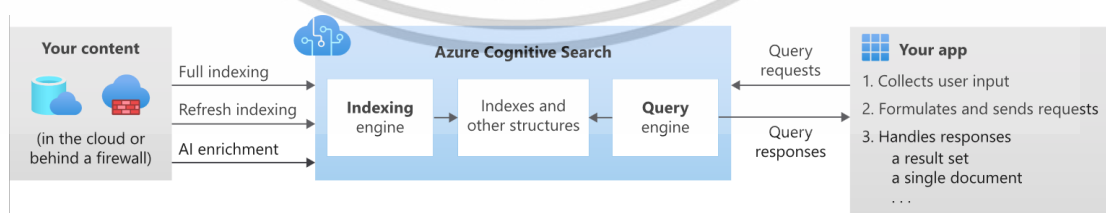
3.5.6 การค้นหาข้อมูลและเอกสารที่เกี่ยวข้อง

การค้นหาข้อมูลและเอกสารที่เกี่ยวข้องเป็นกระบวนการที่ทำก่อนการสรุป Content หรือเนื้อหา เพื่อเตรียมข้อมูลคำตอบสำหรับส่งกลับไปยังผู้ใช้งาน ในการค้นหาจะใช้ Search API ของ Cognitive search เป็นเครื่องมือในการค้นหาข้อมูลและเอกสารที่เกี่ยวข้องจากแหล่งข้อมูลหรือ storage ที่มีการรวบรวมเอกสารขององค์กรไว้



รูปที่ 3.14 Cognitive search process search by embedding

การค้นหาของระบบแชทบอทสำหรับองค์กรจะใช้การค้นหาจากเวกเตอร์หรือค่า embedding ที่คำนวณได้และค้นหาจาก text หรือ keyword ที่ได้จากการสรุปข้อความคำถามของผู้ใช้งาน โดยการค้นหาจากเวกเตอร์หรือการค้นหาจากค่า embedding ที่คำนวณได้จะเป็นการเปรียบเทียบและหาค่าของเวกเตอร์ที่ใกล้เคียงกันระหว่างค่าเวกเตอร์ของ keyword หรือคำถามของผู้ใช้งาน และค่าเวกเตอร์ข้อมูลเอกสารขององค์กร เพื่อหากลุ่มของข้อมูลและความใกล้เคียงของคำถามกับข้อมูลในฐานข้อมูลที่เกี่ยวข้อง



รูปที่ 3.15 Cognitive search process search by text

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การค้นหาด้วย text หรือ keyword คำถามของผู้ใช้งาน ในกระบวนการค้นหาจะใช้ข้อมูล keyword ค้นหาโดยตรงข้อมูลที่เกี่ยวข้องโดยตรง ซึ่งผลลัพธ์ที่ได้จะเป็นผลลัพธ์ที่เกี่ยวข้องและมีความหมายที่เกี่ยวข้องกับ keyword ที่ถูกค้นหา

ระบบจะใช้ทั้งคำเวกเตอร์และ keyword ใน 1 รอบของการค้นหา การค้นหาด้วย keyword จะได้ผลลัพธ์ที่เกี่ยวข้องกับคำที่ค้นหา และการใช้เวกเตอร์จะทำให้ได้ผลลัพธ์ที่มีความคล้ายคลึงกันมากขึ้น การค้นหาแบบนี้จะทำให้ได้ข้อมูลที่เกี่ยวข้อง อยู่ในกลุ่มเดียวกัน และมีความคล้ายคลึงกันมากที่สุด เพื่อให้ผลลัพธ์ที่ได้จากการค้นหาและคำตอบมีประสิทธิภาพมากกว่าการค้นหาแบบใช้ text หรือข้อความ keyword เพียงอย่างเดียว เพื่อให้ผู้ใช้งานได้รับผลลัพธ์ที่มีความครอบคลุมทั้งในความหมายและความคล้ายคลึงของข้อความ

3.5.7 การดูข้อมูลเอกสารอ้างอิง

การดูข้อมูลเอกสารอ้างอิงจากคำตอบที่ระบบแชทบอทตอบกลับไปยังผู้ใช้งานจะเป็นการใช้งาน API โดยส่ง request หรือส่งคำขอไปยัง Azure Blob storage ที่มีการรวบรวมข้อมูลขององค์กรไว้ โดยใน API นั้นจะมีรายละเอียดของ path หรือที่อยู่และชื่อของเอกสารนั้นในฐานข้อมูล เมื่อผู้ใช้งานกดชื่อไฟล์ที่แชทบอทตอบกลับในส่วนของ Citations แล้วจะมีการเรียกใช้งาน API นี้ทุกครั้ง เพื่อดึงข้อมูลเอกสารอ้างอิงจากฐานข้อมูลมาแสดงให้ผู้ใช้งานสามารถดูรายละเอียดเพิ่มเติมได้จาก Front-end

3.5.8 การตรวจสอบความถูกต้องของคำตอบ

การตรวจสอบความถูกต้องของคำตอบจากแชทบอทจะตรวจสอบโดยการเปรียบเทียบภาษาธรรมชาติ โดยเปรียบเทียบระหว่างชุดคำถาม-คำตอบจากผู้เชี่ยวชาญทางด้านข้อมูลขององค์กรกับคำตอบที่ได้จากระบบแชทบอท จะมีการเก็บข้อมูลชุดคำตอบที่ได้จากผู้เชี่ยวชาญและชุดคำตอบที่ได้จากแชทบอทไว้ และจะนำชุดคำตอบนั้นไปตรวจสอบโดยมีการเขียนโปรแกรมเพิ่มเติมสำหรับใช้ในการตรวจสอบความถูกต้องของคำตอบโดยเฉพาะ ซึ่งจากการศึกษาทฤษฎีที่ใช้ในการเปรียบเทียบภาษาธรรมชาติทั้ง 4 วิธี พบว่าวิธี Cosine similarity และ Euclidean distance เป็นวิธีที่เหมาะสมในการใช้เปรียบเทียบ เนื่องจากสามารถใช้ร่วมกับ Embedding model เพื่อแปลงข้อมูลข้อความให้อยู่ในรูปแบบเวกเตอร์แล้วนำค่าเวกเตอร์นั้นไปหาความใกล้เคียงและความต่างได้ แต่การใช้วิธี ROUGE Score และ BLEU Score อาจไม่เหมาะสมในการนำมาใช้ตรวจสอบความถูกต้องของคำตอบที่ได้จากแชทบอทและคำตอบจากผู้เชี่ยวชาญ เนื่องจากพบว่ามีข้อจำกัดทางด้านภาษาคือภาษาไทย 1 คำสามารถเขียนได้หลายแบบ และข้อจำกัดของ model คือการวัดแบบ ROUGE Score และ BLEU Score เป็นการวัดที่ดูจากคำที่เหมือนกันเท่านั้น แต่ไม่ได้ดูที่ความหมายว่าคำนั้นมีความหมายเหมือนกันหรือไม่ ทำให้เมื่อเป็นคำที่มีความหมายเหมือนกันแต่เขียนได้หลายแบบ model ก็จะนับเป็นคนละคำและทำให้ได้ผลลัพธ์ที่มีความผิดพลาดค่อนข้างมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการตรวจสอบความถูกต้องแบบ Cosine similarity และ Euclidean distance จะใช้วิธีแปลงข้อมูลคำตอบจากผู้เชี่ยวชาญและคำตอบจากแชทบอทให้อยู่ในรูปแบบของข้อมูลเวกเตอร์ด้วย Embedding model เมื่อได้ข้อมูลเวกเตอร์ของชุดคำตอบทั้งหมดแล้วจะนำมาตรวจสอบความถูกต้องของคำตอบโดยการใช้ Cosine similarity และ Euclidean distance การใช้วิธี Cosine similarity จะเป็นการเปรียบเทียบหาความคล้ายคลึงกันของค่าเวกเตอร์ทั้ง 2 ที่นำมาเปรียบเทียบกับกัน ทำให้สามารถทราบได้ว่าข้อมูลทั้ง 2 นั้นมีความใกล้เคียงกันมากเพียงใด และการใช้วิธี Euclidean distance จะเป็นการนำข้อมูลเวกเตอร์ทั้ง 2 มาเปรียบเทียบหาความต่างของระยะระหว่างเวกเตอร์ทั้ง 2 ผลลัพธ์ที่ได้จาก Cosine similarity ถ้ามีค่ามากหมายถึงข้อมูลมีความคล้ายคลึงกันมาก แต่ผลลัพธ์ที่ได้จาก Euclidean distance ถ้าผลลัพธ์มีค่าน้อยหมายถึงข้อมูลทั้ง 2 มีความแตกต่างกันเพียงเล็กน้อย

3.6 ค่าใช้จ่ายสำหรับการใช้งานบริการของ Azure Service

ในการใช้งานบริการของ Azure Service ประกอบไปด้วย Azure OpenAI, Azure Cognitive Search และ Azure Blob Storage โดยแต่ละบริการมีค่าใช้จ่ายที่แตกต่างกัน ดังนี้

- 1) Azure OpenAI Service
 - 1.1) Language models

ตารางที่ 3.4 ค่าจ่ายของ Azure OpenAI Language Model

Model	Prompt (ต่อ 1,000 tokens)	Completion (ต่อ 1,000 tokens)
GPT-3.5-Turbo 16K	\$0.003	\$0.004

- 1.2) Embedding models

ตารางที่ 3.5 ค่าจ่ายของ Azure OpenAI Embedding Model

Model	ต่อ 1,000 tokens
Ada	\$0.0001

- 2) Azure Cognitive Search Service

การใช้งาน Azure Cognitive Search Service มีค่าใช้จ่ายในการใช้งานบริการ ดังนี้

ตารางที่ 3.6 ค่าใช้จ่ายของ Azure Cognitive Search Service

	Standard S1 (ต่อชั่วโมง)
Price per SU (Scale Unit)	\$0.34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) Azure Blob Storage

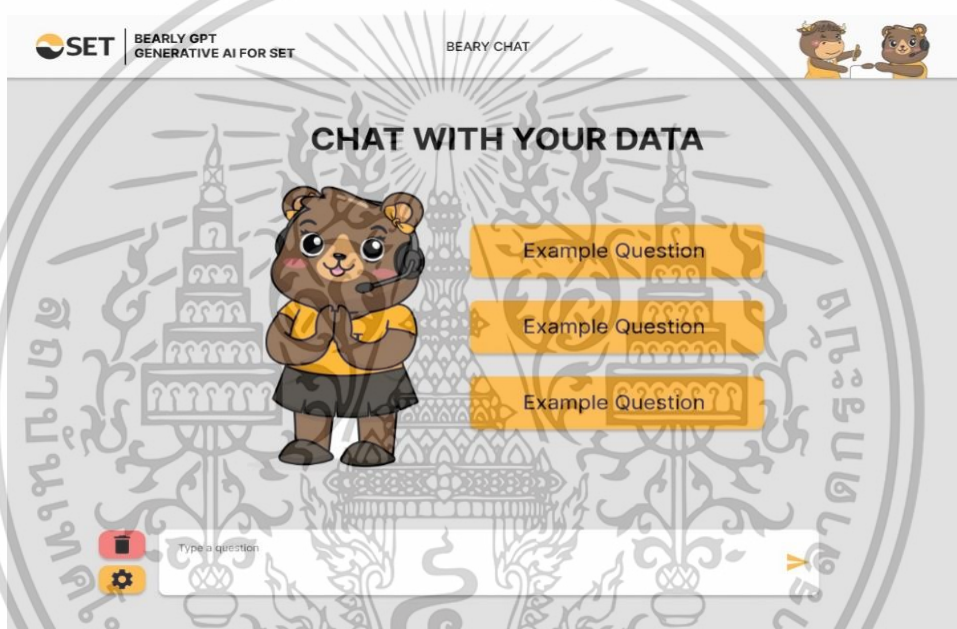
การใช้งาน Azure Blob Storage มีค่าใช้จ่ายในการใช้งานบริการ ดังนี้

ตารางที่ 3.7 ค่าใช้จ่ายของ Azure Blob Storage

Data storage prices pay-as-you-go	Hot tier (ต่อ GB)
First 50 terabyte (TB) / month	\$0.018

3.7 การออกแบบหน้าจอปฏิสัมพันธ์ (User Interface Design)

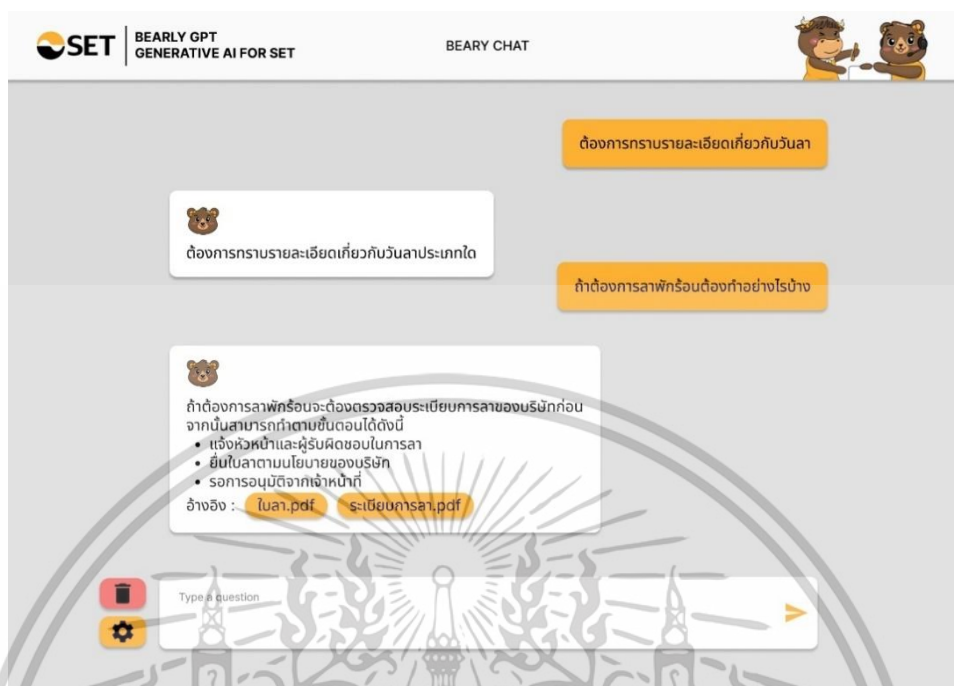
3.7.1 หน้าจอเริ่มต้นของแชทบอท



รูปที่ 3.16 หน้าจอเริ่มต้นของระบบ

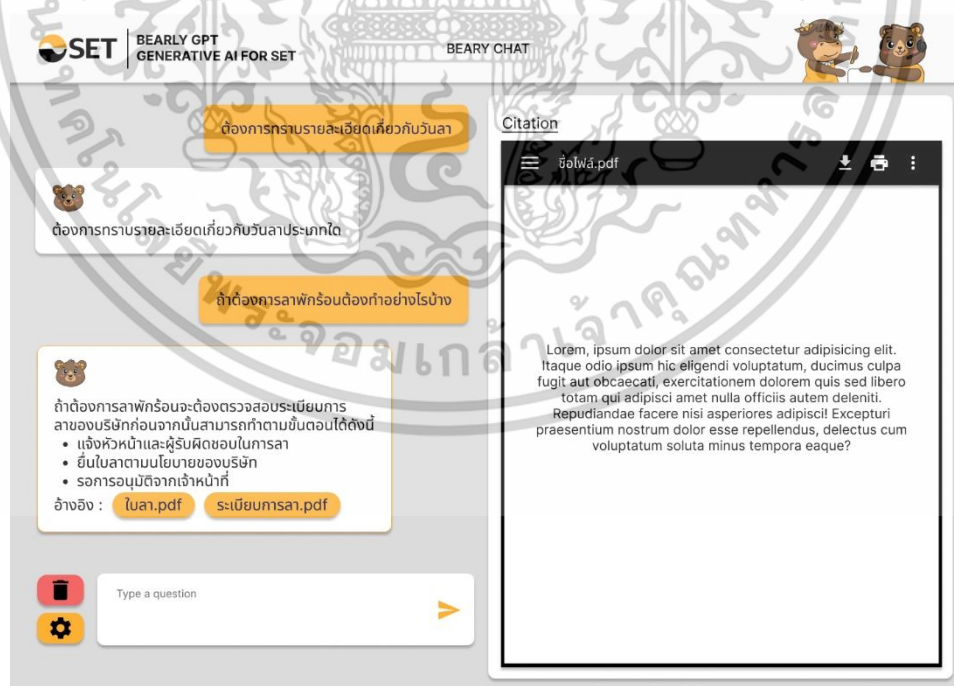
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.2 หน้าจอการสนทนาของระบบ



รูปที่ 3.17 หน้าจอการสนทนาของระบบ

3.7.3 หน้าจอการแสดงผลไฟล์เอกสารอ้างอิง



รูปที่ 3.18 หน้าจอการแสดงผลเอกสารอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลสทกจศึกษาและการอภิปรายผล

ผลของการดำเนินงานสทกจศึกษาระบบแชทบอทสำหรับองค์กร ชุดข้อมูลต่างๆ ที่ใช้ในการทดสอบและประเมินผลการทำงานของระบบแชทบอท รวมถึงหน้าจอบปฏิสัมพันธ์ของระบบ และการประเมินความพึงพอใจในการใช้งานระบบ สามารถอธิบายได้ดังนี้

4.1 ชุดข้อมูลที่ใช้ในการทดสอบระบบ

ชุดข้อมูลที่นำมาใช้ในการทดสอบระบบเป็นข้อมูลที่ได้มาจากบุคลากรภายในองค์กรที่มีความรู้ความเชี่ยวชาญทางด้านข้อมูลและการจัดการเอกสารขององค์กรทั้งหมด 2 ท่าน โดยข้อมูลทั้งหมดที่นำมาใช้ในการทดสอบระบบจะมีทั้งหมด 38 ไฟล์ ซึ่งเป็นข้อมูลภายในองค์กรที่ไม่สามารถเผยแพร่ได้ทั้งหมด จะสร้างชุดข้อมูลคำถาม-คำตอบที่ใช้สำหรับการทดสอบระบบทั้งหมด 100 ข้อ และจะนำไปใช้ในการประเมินประสิทธิภาพการโต้ตอบของระบบแชทบอทด้วยโปรแกรมการตรวจสอบความคล้ายคลึงและความถูกต้องต่อไป

4.2 การประเมินประสิทธิภาพของระบบ

การประเมินประสิทธิภาพของระบบมีวัตถุประสงค์เพื่อวัดความคล้ายคลึงทางความหมายของคำตอบระหว่างค่าเวกเตอร์ของคำตอบอ้างอิงที่ได้จากผู้เชี่ยวชาญด้านข้อมูลและเวกเตอร์คำตอบของระบบแชทบอท โดยใช้การวัดแบบ Cosine Similarity และ Euclidean Distance ซึ่งเป็นเครื่องมือทางคณิตศาสตร์ที่นำมาใช้ในการวัดความคล้ายคลึงและความแตกต่างระหว่างเวกเตอร์ที่แทนคำตอบทั้งสองที่ได้จาก Embedding model ของ OpenAI

การประเมินประสิทธิภาพของระบบได้มีการกำหนดเกณฑ์ที่ใช้เป็นเกณฑ์ในการตัดสินว่าคำตอบที่ระบบตอบกลับมาเป็นคำตอบที่น่าพึงพอใจ จากการวิเคราะห์และทดลองกับชุดข้อมูลตัวอย่างคู่คำถาม-คำตอบซึ่งเป็นข้อมูลทั่วไป จำนวน 100 ตัวอย่าง พบว่าค่า Cosine Similarity หรือค่าเฉลี่ยความคล้ายคลึงของชุดข้อมูลเวกเตอร์อยู่ที่ประมาณ 0.89 (89%) และได้ค่า Euclidean Distance หรือค่าความต่างของข้อมูลประมาณ 0.57 (57%) ดังตารางที่ 4.1 จากผลการทดสอบนี้ พบว่าค่า Cosine Similarity ที่ 0.8 และค่า Euclidean Distance ที่ 0.6 เป็นค่าที่เหมาะสมสำหรับงานนี้ จึงได้มีการกำหนดเกณฑ์ว่าค่า Cosine Similarity หรือค่าความคล้ายคลึงของเวกเตอร์ควรมีค่าเฉลี่ยไม่ต่ำกว่า 0.8 และค่า Euclidean Distance หรือค่าความแตกต่างของเวกเตอร์คำตอบต้นแบบและเวกเตอร์คำตอบของแชทบอทควรมีค่าเฉลี่ยไม่เกิน 0.6 จึงจะถือว่าระบบผ่านการประเมิน ซึ่งหมายความว่าค่า Cosine Similarity ควรมีความคล้ายคลึงในระดับที่สูงกว่า 80% และค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Euclidean Distance ควรมีความแตกต่างกันต่ำกว่า 60% จึงจะสรุปได้ว่าระบบมีประสิทธิภาพ ในการตอบคำถาม และประมวลผลข้อมูลได้เหมาะสมและถูกต้องตามชุดข้อมูลที่ใช้ในการฝึกฝนระบบ

ตารางที่ 4.1 สรุปผลการทดสอบการตอบคำถามของระบบที่ใช้ในการกำหนดเกณฑ์

ครั้งที่ทำการทดสอบ	Euclidean Distance	Cosine Similarity
1	0.5372	0.8977
2	0.6293	0.8868
3	0.5521	0.8965
สรุปผลโดยเฉลี่ย	≈ 0.57	≈ 0.89

โดยมีตัวอย่างชุดข้อมูลคำถาม-คำตอบ ดังนี้

ตารางที่ 4.2 ตัวอย่างชุดคำถาม-คำตอบเกี่ยวกับข้อมูลทั่วไปที่ใช้ในการทดสอบระบบ

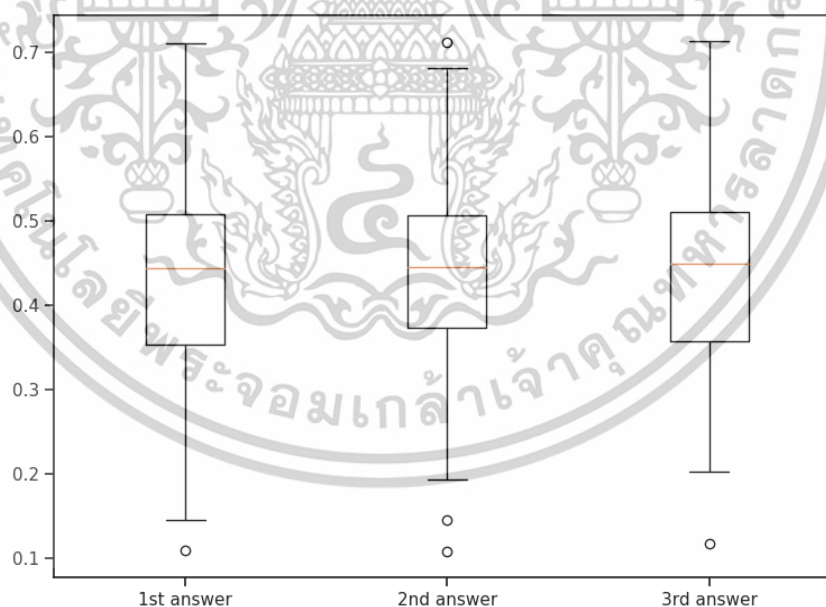
คำถาม	คำตอบอ้างอิง	คำตอบครั้งที่ 1	คำตอบครั้งที่ 2	คำตอบครั้งที่ 3
สายรุ้งมีกี่สี สีอะไรบ้าง	รุ้งกินน้ำ ประกอบด้วยแถบสี ทั้งหมด 7 แถบ คือ สีม่วง สีคราม สีน้ำเงิน สีเขียว สีเหลือง สีส้ม และสีแดง	สายรุ้งมีสีทั้งหมด 7 สีได้แก่ สีแดง, สีส้ม, สีเหลือง, สีเขียว, สีฟ้า, สีน้ำเงิน, และสีม่วง	สายรุ้งมี 7 สี ได้แก่ สีแดง, สีส้ม, สีเหลือง, สีเขียว, สีฟ้า, สีน้ำเงิน, และสีม่วง	สายรุ้งมี 7 สี คือ สีแดง, สีส้ม, สีเหลือง, สีเขียว, สีฟ้า, สีน้ำเงิน, และสีม่วง
เมืองหลวงของประเทศอินเดียคืออะไร	นิวเดลี	เมืองหลวงของประเทศอินเดียคือ เมืองนิวเดลี	เมืองหลวงของประเทศอินเดียคือ เมืองนิวเดลี	เมืองหลวงของประเทศอินเดียคือ เมืองนิวเดลี
ฝนเกิดจากอะไร	ฝนเกิดจาก อนุภาคของไอน้ำขนาดต่างๆ ในก้อนเมฆเมื่อมีขนาดใหญ่ขึ้นจนไม่สามารถลอยตัวอยู่ในก้อนเมฆได้ ก็จะตกลงมาเป็นฝน	ฝนเกิดจากการระเหยของน้ำที่อยู่ในบรรยากาศ เมื่อน้ำที่อยู่ในรูปของละอองน้ำหรือเม็ดฝนเข้าสู่บรรยากาศแล้ว เจอกับอากาศที่มีความเย็น น้ำจะเปลี่ยนเป็นรูปของ	ฝนเกิดจากการระเหยของน้ำที่อยู่ในบรรยากาศ เมื่อน้ำที่อยู่ในรูปของละอองน้ำหรือเม็ดฝนเข้าสู่บรรยากาศแล้ว เจอกับอากาศที่มีความเย็น น้ำจะเริ่มเปลี่ยนเป็นรูปของละอองน้ำหรือ	ฝนเกิดจากการระเหยของน้ำที่อยู่ในบรรยากาศ เมื่อน้ำที่ระเหยขึ้นสูงกว่าระดับความหนาแน่นของอากาศ จะเกิดการเปลี่ยนแปลงของสภาพอากาศที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

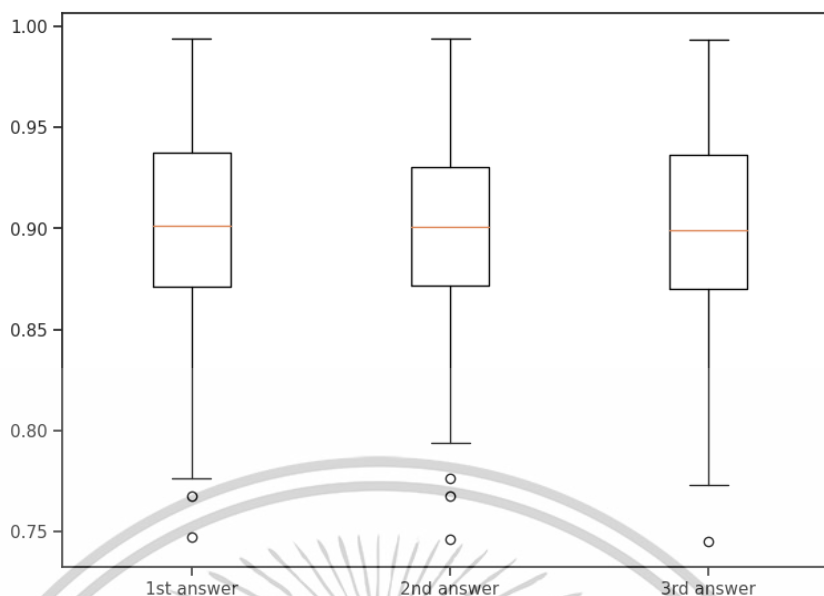
		ละอองน้ำหรือเม็ดฝน และตกลงมาบน พื้นผิวโลก	เม็ดฝนและตกลงมา บนพื้นผิวโลก	เกี่ยวข้องกับกาเกิด ฝน
ภารกิจใดของ Apollo moon ที่ บรรทุกยานสำรวจ ดวงจันทร์เป็น คนแรก	ภารกิจพอลลโล 15	ภารกิจของ Apollo moon ที่บรรทุกยาน สำรวจดวงจันทร์เป็น คนแรกคือภารกิจ Apollo 11	ภารกิจของ Apollo moon ที่บรรทุกยาน สำรวจดวงจันทร์เป็น คนแรกคือภารกิจ Apollo 11	ภารกิจของ Apollo moon ที่บรรทุกยาน สำรวจดวงจันทร์เป็น คนแรกคือภารกิจ Apollo 11
ชื่อทางวิทยาศาสตร์ ของผงชูรสคืออะไร	monosodium L- glutamate	ชื่อทางวิทยาศาสตร์ ของผงชูรสคือ "Monosodium glutamate"	ชื่อทางวิทยาศาสตร์ ของผงชูรสคือ "Monosodium glutamate"	ชื่อทางวิทยาศาสตร์ ของผงชูรสคือ "Monosodium glutamate"

ในการทดสอบเพื่อประเมินประสิทธิภาพของระบบแชทบอทจะทำการทดสอบการตอบคำถามของระบบโดยใช้ชุดคำถามเดียวกัน จะถามทั้งหมด 3 ครั้งและจะนำชุดคำถาม-คำตอบทั้ง 3 ชุดมาเปรียบเทียบกับชุดคำถาม-คำตอบจากผู้เชี่ยวชาญ และหาค่าเฉลี่ยเพื่อสรุปผลจากการทดสอบโดยใช้ทั้ง 2 วิธี สามารถสรุปผลได้ดังนี้



รูปที่ 4.1 กราฟแสดงผลลัพธ์ของการเปรียบเทียบด้วย Euclidean Distance ทั้ง 100 คำถาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 กราฟแสดงผลลัพธ์ของการเปรียบเทียบด้วย Cosine Similarity ทั้ง 100 คำถาม

ตารางที่ 4.3 สรุปผลการทดสอบการตอบคำถามของระบบ

ครั้งที่ทำการทดสอบ	Euclidean Distance	Cosine Similarity
1	0.5310	0.9010
2	0.5422	0.8981
3	0.5444	0.8956
สรุปผลโดยเฉลี่ย	≈ 0.54	≈ 0.90

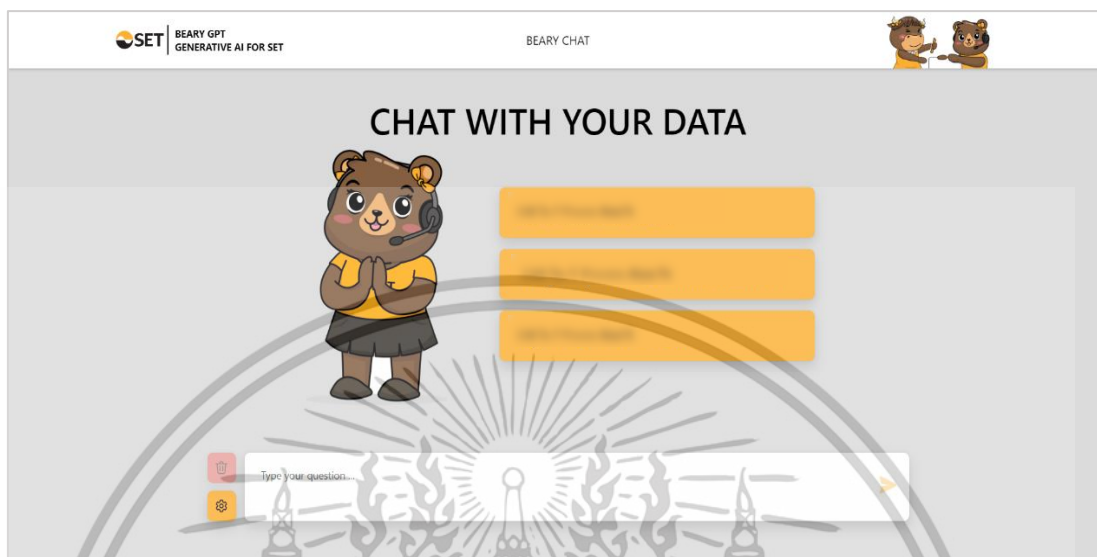
- 1) การทดสอบด้วยวิธี Cosine Similarity ได้ผลลัพธ์ประมาณ 0.90 หรือมีค่าความใกล้เคียงอยู่ที่ประมาณ 90% ซึ่งมีค่ามากกว่า 80% จึงสรุปได้ว่าผ่านเกณฑ์ที่กำหนดไว้
 - 2) การทดสอบด้วยวิธี Euclidean Distance ได้ผลลัพธ์ประมาณ 0.54 หรือมีค่าความต่างของชุดคำตอบประมาณ 54% ซึ่งมีค่าน้อยกว่า 60% จึงสรุปได้ว่าผ่านเกณฑ์ที่กำหนดไว้
- เมื่อนำชุดคำถาม-คำตอบทั้ง 100 ข้อของการทดสอบทั้ง 3 ครั้ง มาเปรียบเทียบความถูกต้องโดยมนุษย์ พบว่าคำตอบที่ได้จากระบบแชทบอทมีทั้งหมด 14 ข้อ ที่แชทบอทตอบแตกต่างไปจากชุดคำถาม-คำตอบที่ได้จากผู้เชี่ยวชาญ รวมทั้งมีบางข้อที่แชทบอทสามารถตอบได้ใกล้เคียง ซึ่งข้อที่มีความถูกต้องและใกล้เคียงคิดเป็น 86% และแชทบอทสามารถตอบได้ถูกต้องเป็นส่วนใหญ่ เมื่อเทียบกับชุดคำถาม-คำตอบที่ได้จากผู้เชี่ยวชาญ โดยจากผลการทดสอบด้วยวิธี Cosine Similarity และ Euclidean Distance พบว่าการใช้วิธี Cosine Similarity และ Euclidean Distance ในการทดสอบและประเมินประสิทธิภาพการตอบคำถามของระบบแชทบอท ทำให้ได้ข้อสรุปว่าผลลัพธ์จาก

ทั้ง 2 วิธีนี้มีความใกล้เคียงกับผลลัพธ์ที่ตรวจสอบโดยมนุษย์ และผลลัพธ์นี้ผ่านเกณฑ์ที่กำหนดไว้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 หน้าจอปฏิสัมพันธ์ของระบบ (User Interface)

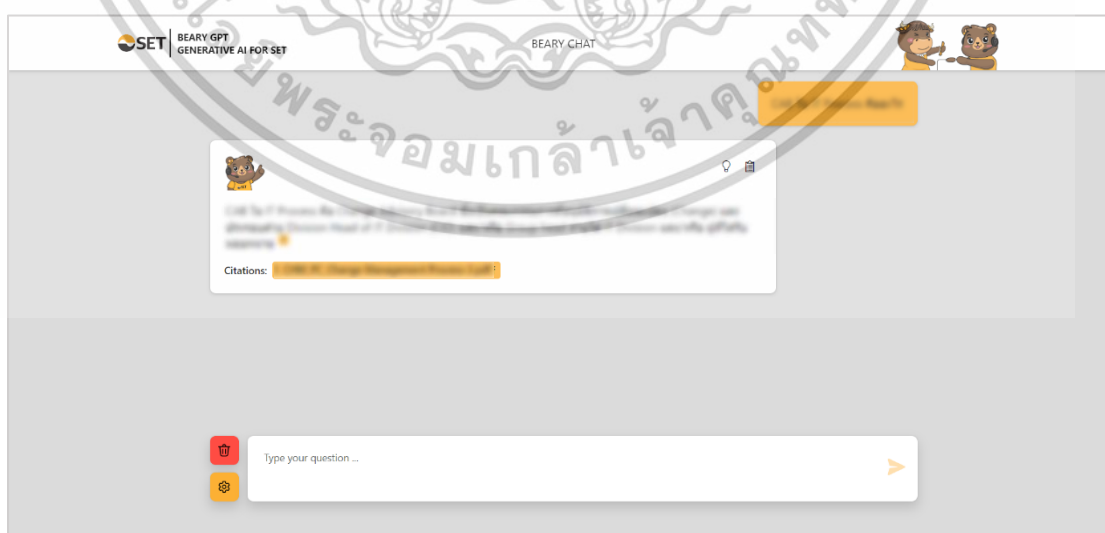
4.3.1 หน้าหลักของระบบแชทบอท



รูปที่ 4.3 หน้าหลักของระบบแชทบอท

หน้าหลักของระบบแชทบอทหรือหน้าเริ่มต้นของเว็บแอปพลิเคชัน ผู้ใช้งานสามารถเริ่มต้นสนทนากับแชทบอทได้ที่หน้านี้ โดยจะมีตัวอย่างคำถาม 3 คำถาม และมีปุ่มลบบประวัติการสนทนา

4.3.2 หน้าจอการสนทนากับระบบแชทบอท

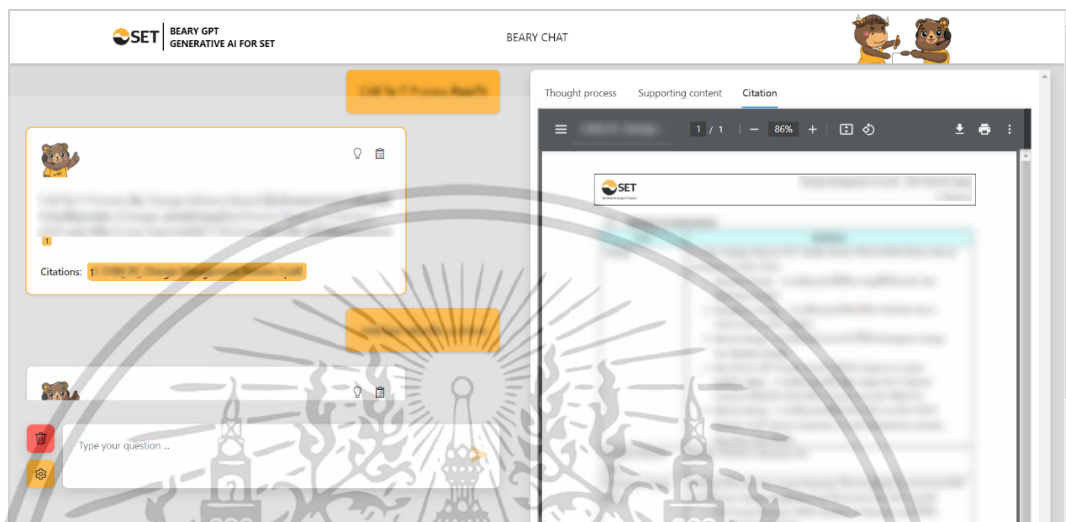


รูปที่ 4.4 หน้าจอการสนทนากับระบบแชทบอท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อผู้ใช้งานสนทนากับระบบแชทบอทแล้ว ในการตอบกลับของระบบแชทบอท ถ้ามีไฟล์เอกสารอ้างอิงแนบมาในการตอบด้วย ไฟล์เอกสารอ้างอิงจะอยู่ในส่วนของ Citations ซึ่งสามารถกดเพื่อดูข้อมูลของเอกสารในหน้านั้นๆ ได้

4.3.3 หน้าจอการแสดงผลข้อมูลเอกสารอ้างอิง



รูปที่ 4.5 หน้าจอการแสดงผลข้อมูลเอกสารอ้างอิง

เมื่อผู้ใช้งานกดที่ชื่อไฟล์ที่แชทบอทมีการอ้างอิงมาพร้อมกับคำตอบ ข้อมูลของไฟล์เอกสารอ้างอิงนั้นจะแสดงขึ้นที่ด้านขวาของหน้าจอการสนทนา

4.4 การทดสอบการทำงานของระบบ

ตารางที่ 4.4 รายละเอียดการทดสอบการส่ง API

ลำดับ	การทดสอบ	ขั้นตอนการทดสอบ	ผลลัพธ์ที่คาดหวัง	ผลการทดสอบ	
				ผ่าน	ไม่ผ่าน
01	ตรวจสอบการส่ง API ไปยัง Azure OpenAI เพื่อประมวลผลข้อความ	1. ใส่ API Key 2. ใส่ข้อมูลที่ต้องการประมวลผล	สามารถส่งข้อมูลได้ และมีข้อมูลตอบกลับ	√	
02	ตรวจสอบการส่ง API ไปยัง Azure Cognitive Search เพื่อค้นหาข้อมูล	1. ใส่ API Key 2. ใส่ข้อมูลที่ต้องการค้นหา	สามารถส่งข้อมูลได้ และมีข้อมูลที่ค้นหาตอบกลับ	√	
03	ตรวจสอบการส่ง API ไปยัง Azure Blob Storage เพื่อเรียกไฟล์อ้างอิง	1. ใส่ API Key 2. ใส่ชื่อไฟล์ที่ต้องการเรียก	สามารถส่งข้อมูลได้ และมีข้อมูลไฟล์ตอบกลับ	√	

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาก็เป็นได้ ไม่นับเป็นประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 รายละเอียดการทดสอบหน้าจอการใช้งาน

ลำดับ	การทดสอบ	ขั้นตอนการทดสอบ	ผลลัพธ์ที่คาดหวัง	ผลการทดสอบ	
				ผ่าน	ไม่ผ่าน
01	ทดสอบการเข้าสู่ระบบโดยผู้ใช้งานที่มีสิทธิ์เข้าถึงเท่านั้น (บุคลากรภายในองค์กร)	1. เข้าสู่ระบบ IT Portal 2. เลือกเมนู Beary GPT	บุคลากรภายในองค์กรสามารถเข้าใช้งานระบบได้	√	
02	ทดสอบการถามคำถามกับระบบ	1. เข้าสู่ระบบ IT Portal 2. เลือกเมนู Beary GPT 3. ถามคำถามกับระบบ	ผู้ใช้งานสามารถส่งคำถามไปยังระบบได้	√	
03	ทดสอบการตอบกลับจากระบบ	1. เข้าสู่ระบบ IT Portal 2. เลือกเมนู Beary GPT 3. ถามคำถามกับระบบ	มีคำตอบจากระบบตอบกลับมายังผู้ใช้งาน	√	
04	การกดปุ่มลบบประวัติการสนทนาในหน้าการสนทนา	1. เข้าสู่ระบบ IT Portal 2. เลือกเมนู Beary GPT 3. ถามคำถามกับระบบ 4. กดลบบประวัติสนทนา	สามารถลบบประวัติการสนทนาได้ และกลับไปยังหน้าแรก	√	

4.5 การประเมินความพึงพอใจในการใช้งานระบบ

การทำระบบแชทบอทสำหรับองค์กรมีการทำแบบประเมินความพึงพอใจในการใช้งานระบบแชทบอทเพื่อช่วยให้ทราบถึงข้อดีและข้อเสียของระบบจากผู้ที่ใช้งานโดยตรง ซึ่งทำให้สามารถเข้าใจในมุมมองของผู้ใช้งานได้ รวมทั้งสามารถนำมาพัฒนาระบบ และปรับปรุงส่วนต่างๆ ของระบบให้ดีขึ้นได้

ในการประเมินความพึงพอใจในการใช้งานระบบแชทบอท ผู้ที่ทำการประเมินจะเป็นกลุ่มตัวอย่างของบุคลากรภายในองค์กรเท่านั้น จำนวน 10 คน โดยแบบประเมินความพึงพอใจในการใช้งานระบบจะเป็นการให้ผู้ใช้งานประเมินระดับความพึงพอใจ 5 ระดับ ได้แก่ พึงพอใจมากที่สุด พึงพอใจมาก พึงพอใจปานกลาง พึงพอใจน้อย และพึงพอใจน้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความพึงพอใจในการใช้งานระบบ *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
ระบบแชทบอทสามารถใช้งานได้ง่ายในระดับใด	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ระบบแชทบอทสามารถเข้าใจคำถามและโต้ตอบกับผู้ใช้งานได้อย่างถูกต้อง	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ระบบสามารถเสนอคำแนะนำและช่วยแก้ปัญหาได้อย่างตรงจุด	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ระบบแชทบอทสามารถเข้าถึงข้อมูลตามที่ผู้ใช้งานต้องการได้ในระดับใด	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ความพึงพอใจในความพร้อมของคำตอบและเอกสารอ้างอิง	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ระบบมีความเสถียรในระดับใด	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
การใช้งานระบบแชทบอทที่มีประโยชน์มาคน้อยเพียงใด	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ความพึงพอใจโดยรวมในการใช้งานระบบแชทบอท	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
จากการใช้งานในครั้งนี้ คุณคิดว่าจะใช้งานระบบแชทบอทนี้อีกครั้งหรือไม่	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

ข้อเสนอแนะเพิ่มเติม

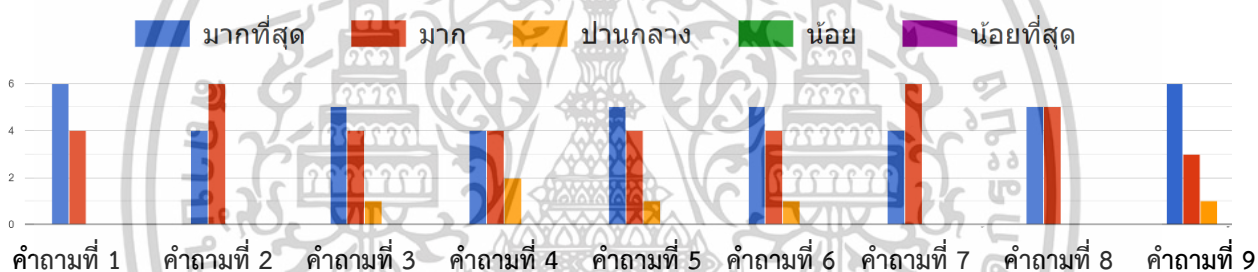
Your answer _____

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 รูปที่ 4.6 แบบฟอร์มประเมินความพึงพอใจในการใช้งานระบบ
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

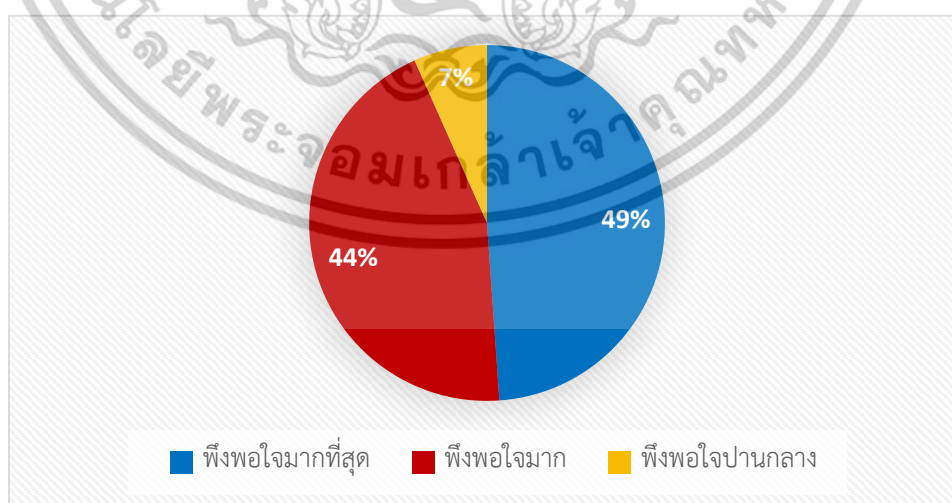
จากภาพ มีคำถามที่ใช้ในการประเมินความพึงพอใจในการใช้งานระบบทั้งหมด 9 คำถาม ดังนี้

- 1) ระบบเซทบอทสามารถใช้งานได้ง่ายในระดับใด
- 2) ระบบเซทบอทสามารถเข้าใจคำถามและโต้ตอบกับผู้ใช้งานได้อย่างถูกต้อง
- 3) ระบบสามารถเสนอคำแนะนำและช่วยแก้ปัญหาได้อย่างตรงจุด
- 4) ระบบเซทบอทสามารถเข้าถึงข้อมูลตามที่ต้องการได้ในระดับใด
- 5) ความพึงพอใจในความถูกต้องของคำตอบและเอกสารอ้างอิง
- 6) ระบบมีความเสถียรในระดับใด
- 7) การใช้งานระบบเซทบอทนี้มีประโยชน์มากน้อยเพียงใด
- 8) ความพึงพอใจโดยรวมในการใช้งานระบบเซทบอท
- 9) จากการใช้งานในครั้งนี้ คุณคิดว่าจะใช้งานระบบเซทบอทนี้อีกครั้งหรือไม่

เมื่อกลุ่มตัวอย่างของผู้ใช้งานซึ่งเป็นบุคลากรภายในองค์กรทั้ง 10 คน ในฝ่าย Enterprise Systems ได้ทำการประเมินแล้ว สามารถสรุปผลการประเมินความพึงพอใจในการใช้งานระบบได้ดังนี้



รูปที่ 4.7 ผลการประเมินความพึงพอใจในการใช้งานระบบ



รูปที่ 4.8 กราฟสรุปความพึงพอใจโดยรวมในการใช้งานระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 ผลการประเมินความพึงพอใจในการใช้งานระบบ

ข้อ	คำถามประเมิน	จำนวนคนที่ประเมิน					คิดเป็น %
		มากที่สุด (100%)	มาก (75%)	ปานกลาง (50%)	น้อย (25%)	น้อยที่สุด (0%)	
1	ระบบแชทบอทสามารถใช้งานได้ง่ายในระดับใด	6	4	-	-	-	90%
2	ระบบแชทบอทสามารถเข้าใจคำถามและโต้ตอบกับผู้ใช้งานได้อย่างถูกต้อง	4	6	-	-	-	85%
3	ระบบสามารถเสนอคำแนะนำและช่วยแก้ปัญหาได้อย่างตรงจุด	5	4	1	-	-	85%
4	ระบบแชทบอทสามารถเข้าถึงข้อมูลตามที่ผู้ใช้งานต้องการได้ในระดับใด	4	4	2	-	-	80%
5	ความพึงพอใจในความถูกต้องของคำตอบและเอกสารอ้างอิง	5	4	1	-	-	85%
6	ระบบมีความเสถียรในระดับใด	5	4	1	-	-	85%
7	การใช้งานระบบแชทบอทนี้มีประโยชน์มากน้อยเพียงใด	4	6	-	-	-	85%
8	ความพึงพอใจโดยรวมในการใช้งานระบบแชทบอท	5	5	-	-	-	87.5%
9	จากการใช้งานในครั้งนี้ คุณคิดว่า จะใช้งานระบบแชทบอทนี้ อีกครั้งหรือไม่	6	3	1	-	-	87.5%
ความพึงพอใจในการใช้งานระบบแชทบอทโดยเฉลี่ย							85.56%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการประเมิน ผู้ใช้งานมีความพึงพอใจสูงต่อระบบแชทบอท โดยมีคะแนนสูงที่สุดในการให้คะแนนในส่วนของ "ระบบแชทบอทสามารถใช้งานได้ง่าย" และ "การใช้งานระบบแชทบอทนี้อีกครั้ง" ซึ่งมีผลความพึงพอใจ "มากที่สุด" มากที่สุดที่ถูกประเมินจากผู้ใช้งาน โดยคิดเป็นร้อยละ 60% สำหรับทั้งสองคำถามนี้ แต่ยังมีบางประการที่สามารถพัฒนาและปรับปรุงได้ เช่น ในด้านการเข้าถึงข้อมูลตามที่ต้องการ ซึ่งมีผู้ใช้งานบางคนให้คะแนนที่ระดับปานกลาง และมีความพึงพอใจที่ 80% ซึ่งอาจเป็นโอกาสในการพัฒนาระบบให้มีประสิทธิภาพมากยิ่งขึ้นตามความคิดเห็นของผู้ใช้งาน

จากผลการประเมินความพึงพอใจในการใช้งานระบบทั้ง 9 ข้อ สามารถสรุปได้ว่าผู้ใช้งานมีความพึงพอใจสูงในการใช้งานระบบแชทบอทขององค์กร โดยมีคะแนนเฉลี่ยความพึงพอใจทั้งหมดที่ 85.56% ซึ่งเป็นที่น่าพึงพอใจ โดยสามารถสรุปผลความพึงพอใจในแต่ละด้านได้ดังนี้

- 1) ระดับความพึงพอใจ 90% เป็นส่วนที่ได้รับการประเมินความพึงพอใจจากผู้ใช้งานสูงที่สุด ได้แก่ "ระบบแชทบอทสามารถใช้งานได้ง่าย" (คำถามที่ 1) แสดงถึงความสะดวกสบายในการใช้งานของผู้ใช้งาน
- 2) ระดับความพึงพอใจ 87.5% ได้แก่ "ความพึงพอใจในการใช้งานระบบโดยรวม" (คำถามที่ 8) และ "ความประสงค์ที่ผู้ใช้งานจะใช้งานระบบอีกครั้ง" (คำถามที่ 9)
- 3) ระดับความพึงพอใจ 85% ได้แก่ "การเข้าใจและการโต้ตอบของระบบ" (คำถามที่ 2), "การเสนอคำแนะนำที่ช่วยแก้ปัญหาได้อย่างตรงจุด" (คำถามที่ 3), "ความถูกต้องของคำตอบและเอกสารอ้างอิง" (คำถามที่ 5), "ความเสถียรของระบบ" (คำถามที่ 6) และ "ประโยชน์ที่ได้รับจากการใช้งานระบบแชทบอท" (คำถามที่ 7)
- 4) ระดับความพึงพอใจ 80% ได้แก่ "ระบบแชทบอทสามารถเข้าถึงข้อมูลได้ตามที่ผู้ใช้งานต้องการ" (คำถามที่ 4)

จากผลการประเมินความพึงพอใจในการใช้งานระบบข้างต้น สามารถสรุปได้ว่าระบบแชทบอทนี้มีความพึงพอใจในการใช้งานสูง ประสบความสำเร็จในการให้บริการ ได้รับความไว้วางใจและได้รับการยอมรับจากผู้ใช้งานในองค์กร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลสหกิจศึกษาและข้อเสนอแนะ

5.1 สรุปผลสหกิจศึกษา

จากการดำเนินงานสหกิจศึกษาระบบแซทบอทสำหรับองค์กร (Beary GPT: Generative AI for SET) ซึ่งเป็นเว็บแอปพลิเคชันแซทบอทที่ช่วยอำนวยความสะดวกให้กับบุคลากรภายในองค์กร ในการตอบคำถามข้อมูลองค์กรทั่วไปและมีการประยุกต์ใช้ทฤษฎีการจัดการองค์ความรู้ (KM) รวบรวมข้อมูลความรู้ต่างๆ ภายในองค์กรไว้เพื่อช่วยให้ข้อมูลความรู้ แนะนำ หรือตอบข้อสงสัยให้แก่บุคลากร โดยในการพัฒนาระบบแซทบอทสำหรับระบบองค์กรจะใช้ Azure Service และ OpenAI Service มาใช้ร่วมกันให้ระบบแซทบอทสามารถเข้าใจ ตีความ และตอบกลับเป็นภาษาธรรมชาติในรูปแบบที่คล้ายคลึงกับภาษาของมนุษย์ และมนุษย์สามารถเข้าใจได้ โดยมีการพัฒนาระบบด้วย React TypeScript ในส่วนของ Front-end และ Python ในส่วนของ Back-end การเก็บข้อมูลจะเก็บข้อมูลในรูปแบบ Object Storage เป็นข้อมูลที่ไม่มีโครงสร้าง หรือ Unstructured Data ซึ่งส่วนใหญ่ข้อมูลขององค์กรจะเป็นข้อมูลประเภทไฟล์ PDF รวมทั้งมีการทดสอบประสิทธิภาพการสร้างคำตอบของระบบแซทบอทโดยการใช้ Embedding Model ในการเทียบค่าความใกล้เคียงระหว่างคำตอบจากแซทบอทและคำตอบจากผู้ที่มีความรู้ความเชี่ยวชาญทางด้านข้อมูลและการจัดการเอกสารขององค์กร

เมื่อพัฒนาระบบแซทบอทเสร็จสมบูรณ์และทำการทดสอบประสิทธิภาพในการสนทนา โดยเปรียบเทียบค่าความใกล้เคียงและความแตกต่างของคำตอบทั้งหมด 100 คำถาม เมื่อเปรียบเทียบโดยใช้ Embedding Model แปลงข้อมูลให้อยู่ในรูปแบบเวกเตอร์แล้วนำเวกเตอร์นั้นไปเปรียบเทียบต่อ สามารถสรุปผลได้ว่า ระบบแซทบอทสามารถตอบคำถามได้ใกล้เคียงกับมนุษย์ ประมาณ 89% เมื่อใช้วิธี Cosine Similarity ในการเปรียบเทียบ ซึ่งเป็นการเปรียบเทียบโดยการหาค่าความคล้ายคลึงกันระหว่างชุดข้อมูลเวกเตอร์ที่ได้จาก Embedding Model นำมาใช้เพื่อประเมินประสิทธิภาพการตอบของระบบแซทบอทโดยเทียบกับเวกเตอร์คำตอบของผู้เชี่ยวชาญ และมีความแตกต่างของค่าเวกเตอร์ที่แทนคำตอบประมาณ 54% เมื่อใช้ Euclidean Distance ในการวัดความแตกต่างของคำตอบ ซึ่งเป็นการวัดระยะห่างระหว่างจุดทั้ง 2 จุดแนวเส้นตรง จึงสรุปได้ว่าการทดสอบผ่านเกณฑ์ Cosine Similarity ที่มากกว่า 80% และ Euclidean Distance ที่ไม่เกิน 60% ตามที่กำหนดไว้

จากการให้ผู้ใช้งานได้ใช้งานระบบแซทบอทและได้มีการประเมินความพึงพอใจในการใช้งานระบบแล้ว จากผลการประเมินความพึงพอใจในการใช้งานระบบ ผู้ใช้งานมีความพึงพอใจสูงต่อระบบแซทบอท ซึ่งมีความพึงพอใจโดยรวมที่ 85.56% แสดงให้เห็นถึงความมีประโยชน์และความสะดวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการใช้งานของผู้ใช้งาน สามารถสรุปได้ว่าระบบแชทบอทสำหรับองค์กรนี้ประสบความสำเร็จในการให้บริการและได้รับความไว้วางใจจากผู้ใช้งานซึ่งเป็นบุคคลากรภายในองค์กร

5.2 ปัญหาที่พบ

ในการจัดทำระบบแชทบอทสำหรับองค์กร ปัญหาที่พบคือการกำหนด prompt ของระบบ และการสร้าง keyword ที่ใช้ในการค้นหา เนื่องจาก prompt ของระบบเป็นส่วนหนึ่งของการสร้าง keyword ซึ่งถ้ากำหนด prompt ได้ไม่ดีจะส่งผลให้ keyword ที่ถูกสร้างขึ้นมีความหมายไม่ครอบคลุมมากพอและเมื่อนำ keyword นั้นไปค้นหาจะทำให้ได้ผลลัพธ์ที่ไม่ถูกต้องหรือไม่ตรงตามความต้องการของผู้ใช้งาน ในบางครั้งถ้าสร้าง keyword ออกมาได้ไม่ถูกต้องและไม่มีคำที่คล้ายคลึงกับข้อมูลในเอกสารก็จะทำให้กระบวนการค้นหาข้อมูลและเอกสารที่เกี่ยวข้องจะหาข้อมูลที่มีความเกี่ยวข้องกันไม่เจอ ทำให้ผู้ใช้งานไม่ได้รับคำตอบที่ต้องการทราบหรือได้รับคำตอบที่ผิดพลาดได้ จากปัญหานี้ทำให้จำเป็นต้องใช้เวลามากในการศึกษาค้นคว้าหาข้อมูลในการสร้างและการกำหนด prompt ของระบบ รวมถึงต้องทดลองกำหนด prompt หลากๆ รูปแบบแล้วนำผลลัพธ์มาเปรียบเทียบกัน เพื่อให้ได้ keyword หรือคำค้นหาที่ดีที่สุด

5.3 ข้อเสนอแนะ

ข้อเสนอแนะของระบบแชทบอทสำหรับองค์กรคือ ในอนาคตระบบแชทบอทสามารถพัฒนาเพื่อนำไปใช้งานร่วมกับแผนกอื่นๆ หรืองานอื่นๆ อย่างเฉพาะเจาะจงและมีความเหมาะสมกับงานในแต่ละด้านได้ เช่น พัฒนาระบบแชทบอทเพื่อช่วยอ่านและวิเคราะห์เอกสารสมัครงาน หรือ CV (Curriculum Vitae) เพื่อช่วยอำนวยความสะดวกและช่วยลดระยะเวลาในการทำงานสำหรับบุคลากรในแผนกการบริหารจัดการทรัพยากรมนุษย์ ซึ่งจะช่วยให้ช่วยสามารถคัดกรองบุคลากรเข้าทำงานในองค์กรได้ในระดับหนึ่ง เป็นต้น เนื่องจากแชทบอทสามารถนำไปประยุกต์และใช้งานร่วมกับงานด้านอื่นๆ ได้หลายทาง รวมถึงอาจมีการพัฒนาการแสดงผลในส่วนที่ติดต่อกับผู้ใช้งานให้สามารถแสดงผลคำตอบในรูปแบบของตารางได้อย่างชัดเจน ซึ่งอาจทำให้ผู้ใช้งานสามารถสามารถอ่านและเข้าใจข้อมูลต่างๆ ได้อย่างสะดวกและง่ายมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น. (2566). “Cloud Computing คืออะไร?” [Online]. เข้าถึงได้จาก: <https://sc2.kku.ac.th/office/sci-it/index.php/29-cloud-computing.html>.
- [2] อภิญญา พรหมนัส. (2563). “วิธีการใช้งาน Visual Studio Code.” [Online]. เข้าถึงได้จาก: <https://cs.bru.ac.th/สอนวิธีการใช้-visual-studio-code-2/>.
- [3] Art Kraisingkorn. (2566). “การนำ GPT มาใช้งานในองค์กรด้วย Retrieval Augmented Generation (RAG).” [Online]. เข้าถึงได้จาก: <https://medium.com/@touchaponk/การนำ-gpt-มาใช้งานในองค์กรด้วย-retrieval-augmented-generation-rag-a5b8a770cade>.
- [4] BorntoDev. (2563). “React 101.” [Online]. เข้าถึงได้จาก: <https://www.borntodev.com/2020/07/15/react-101/>.
- [5] Computer Institute of Ramkhamhaeng University. (2566). “draw.io เครื่องมือสร้าง diagram แบบฟรีเหมียม.” [Online]. Available: https://computer.ru.ac.th/images/Km2Article/16064444802_Article.pdf.
- [6] DUSADEEVIROJ. (2566). “AZURE PORTAL.” [Online]. เข้าถึงได้จาก: <https://www.fusionsol.com/blog/azure-portal/https://www.fusionsol.com/blog/ai/>.
- [7] khanacademy. (2566). “ทำความเข้าใจกับ Large Language Models (LLM).” [Online]. เข้าถึงได้จาก: <https://stepstraining.co/analytics/large-language-models>.
- [8] Kipakapron. (2561). “CSS คืออะไร มีประโยชน์ อย่างไร.” [Online]. เข้าถึงได้จาก: <https://blog.sogoodweb.com/Article/Detail/79237/CSS-คืออะไร-มีประโยชน์-อย่างไร>.
- [9] RealSmart Academy. (2566). “Prompt คืออะไร? และมีความสำคัญในการใช้งาน AI อย่างไร?” [Online]. เข้าถึงได้จาก: <https://academy.realsmart.co.th/prompt-คืออะไร-และมีความสำคัญในการใช้งาน-ai-อย่างไร/>.
- [10] Sarayut Nonsiri, PhD. (2566). “ภาษาโปรแกรม Python คืออะไร ?.” [Online]. เข้าถึงได้จาก: <https://www.9experttraining.com/articles/python-คืออะไร>.
- [11] Witsawa Chanton. (2566). “Azure AI ผู้ช่วยอัจฉริยะที่นักพัฒนาไม่ควรมองข้าม.” [Online]. เข้าถึงได้จาก: <https://monsterconnect.co.th/azure-ai/>.
- [12] Atlassian. (2023). “Introduction to Jira family.” [Online]. Available: <https://www.atlassian.com/software/jira/guides/more/jira-family#what-is-the-jira-family>.
- [13] Atlassian. (2023). “Jira Software boards introduction.” [Online]. Available: <https://www.atlassian.com/software/jira/guides/boards/overview#what-is-a-jira-board>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [14] AWS. (2023). “**What is GPT?**” [Online]. Available: <https://aws.amazon.com/what-is/gpt/>.
- [15] AWS. (2023). “**What is Natural Language Processing (NLP)?**” [Online Available: https://aws.amazon.com/what-is/nlp/?nc1=h_ls.
- [16] Dubey, et al. (2023). “**Introduction to Azure Blob Storage.**” [Online]. Available: <https://learn.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>.
- [17] Fatih Karabiber. (2024). “**Cosine Similarity.**” [Online]. Available: <https://www.learn datasci.com/glossary/cosine-similarity/>.
- [18] Figma Learn. (2023). “**What is Figma?**” [Online]. Available: <https://help.figma.com/hc/en-us/articles/14563969806359-What-is-Figma#article-container>.
- [19] Haas, et al. (2023). “**Azure Storage redundancy.**” [Online]. Available: <https://learn.microsoft.com/en-us/azure/storage/common/storage-redundancy>.
- [20] Jessicawrr. (2023). “**sample-app-aoai-chatGPT.**” [Online]. Available: <https://github.com/microsoft/sample-app-aoai-chatGPT>.
- [21] John Terra. (2023). “**How to Use GitHub Desktop: A GitHub Desktop Tutorial.**” [Online]. Available: <https://www.simplilearn.com/how-to-use-github-desktop-tutorial-article>.
- [22] Kavita Ganesan. (2017). “**An intro to ROUGE, and how to use it to evaluate summaries.**” [Online]. Available: <https://www.freecodecamp.org/news/what-is-rouge-and-how-it-works-for-evaluation-of-summaries-e059fb8ac840/>.
- [23] Leonie Monigatti. (2023). “**Easily Estimate Your OpenAI API Costs with Tiktoken.**” [Online]. Available: <https://towardsdatascience.com/easily-estimate-your-openai-api-costs-with-tiktoken-c17caf6d015e>.
- [24] Natalie Mickey. (2023). “**Explore the benefits of Azure OpenAI Service with Microsoft Learn.**” [Online]. Available: <https://azure.microsoft.com/en-us/blog/explore-the-benefits-of-azure-openai-service-with-microsoft-learn/>.
- [25] OpenAI. (2023). “**Best practices for prompt engineering with the OpenAI API.**” [Online]. Available: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api>.
- [26] OpenAI. (2023). “**GPT-3.5 Turbo.**” [Online]. Available: <https://platform.openai.com/docs/models/gpt-3-5-turbo>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [27] OpenAI. (2023). “**What are embeddings?**” [Online]. Available: <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>.
- [28] pamelafox et al. (2023). “**azure-search-openai-demo.**” [Online]. Available: <https://github.com/Azure-Samples/azure-search-openai-demo>.
- [29] Postman. (2023). “**About Postman.**” [Online]. Available: <https://www.postman.com/company/about-postman/>.
- [30] Prompt Engineering Guide. (2023). “**Retrieval Augmented Generation (RAG).**” [Online]. Available: <https://www.promptingguide.ai/techniques/rag>.
- [31] Rachael Tatman. (2019). “**Evaluating Text Output in NLP: BLEU at your own risk.**” [Online]. Available: <https://towardsdatascience.com/evaluating-text-output-in-nlp-bleu-at-your-own-risk-e8609665a213>.
- [32] suvratara06. (2024). “**NLP – BLEU Score for Evaluating Neural Machine Translation – Python.**” [Online]. Available: <https://www.geeksforgeeks.org/nlp-bleu-score-for-evaluating-neural-machine-translation-python/>.
- [33] TechTarget Contributor. (2020). “**DEFINITION GitLab.**” [Online]. Available: <https://www.techtarget.com/whatis/definition/GitLab>.
- [34] Zoumana Keita. (2023). “**Introduction to Text Embeddings with the OpenAI API.**” [Online]. Available: <https://www.datacamp.com/tutorial/introduction-to-text-embeddings-with-the-open-ai-api>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

การสร้าง Index

กระบวนการสร้าง index

การสร้าง index ใน Azure Cognitive Search เป็นขั้นตอนที่สำคัญในการค้นหาและจัดการกับข้อมูล ซึ่ง index จะเปรียบเสมือนที่เก็บข้อมูลของเอกสารที่ถูกประมวลผล เป็นที่เก็บของชุดข้อมูลที่จะนำไปใช้ในการค้นหา การสร้าง index จะช่วยให้สามารถค้นหาเอกสารที่มี content หรือเนื้อหาที่ตรงกับคำค้นหาของผู้ใช้งานได้ และทำให้สามารถพบเอกสารที่เกี่ยวข้องกับคำค้นหานั้นๆ ทำให้ระบบมีประสิทธิภาพในการค้นหาข้อมูลจากเอกสารที่มีปริมาณมากได้ดีขึ้น

Index จะเก็บข้อมูลของเอกสารในแต่ละหน้าในรูปแบบของ field และ content โดย field จะทำให้ข้อมูลสามารถค้นหาได้ง่ายและรวดเร็ว และ content หรือคำสำคัญที่ใช้ในการค้นหา ตัวอย่างโครงสร้างของ index เช่น ชื่อเอกสาร, วันที่สร้างเอกสาร หรือ เนื้อหาเอกสาร เป็นต้น โดยขั้นตอนในการสร้าง index สำหรับระบบแชทบอทมีขั้นตอนดังนี้

- 1) สร้าง index client เพื่อตรวจสอบข้อมูลใน Azure Cognitive Search โดยใช้ข้อมูลของ search service ที่มีอยู่ และ credential หรือข้อมูลที่รับรองความปลอดภัย

```
index_client = SearchIndexClient(
    endpoint=f'https://{args.searchservice}.search.windows.net/', credential=search_creds
)
```

รูปที่ ก.1 การสร้าง index client

- 2) ตรวจสอบว่าชื่อของ index ที่กำหนดไว้มีอยู่ใน index client หรือไม่ ถ้าชื่อ index นั้นมีอยู่แล้วจะแสดงผลชื่อ index นั้น แต่ถ้า index นั้นยังไม่มีอยู่ใน index client จะทำการสร้าง index โดยกำหนดข้อมูลดังนี้
 - name: ชื่อของ index
 - fields: กำหนดประเภทข้อมูลของ index
 - semantic_settings: กำหนดตัวเลือกในการค้นหา ว่าต้องการให้ค้นหาโดยใช้ semantic หรือไม่
 - vector_search: เป็นการกำหนดว่าต้องการใช้ข้อมูลเวกเตอร์ในการค้นหาหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

if args.index not in index_client.list_index_names():
    index = SearchIndex(
        name=args.index,
        fields=fields,
        semantic_settings=SemanticSettings(
            configurations=[
                SemanticConfiguration(
                    name="default",
                    prioritized_fields=PrioritizedFields(
                        title_field=None, prioritized_content_fields=[SemanticField(field_name="content")]
                    ),
                )
            ]
        ),
        vector_search=VectorSearch(
            algorithm_configurations=[
                VectorSearchAlgorithmConfiguration(
                    name="default", kind="hns", hns_parameters=HnsParameters(metric="cosine")
                )
            ]
        ),
    )
    if args.verbose:
        print(f"Creating {args.index} search index")
    index_client.create_index(index)
else:
    if args.verbose:
        print(f"Search index {args.index} already exists")

```

รูปที่ ก.2 การตรวจสอบรายการ index และการสร้าง index

3) ในการกำหนด field หรือประเภทข้อมูลของ index สามารถกำหนดได้ 3 แบบ คือ

- 1) สร้าง field โดยใช้ SimpleField เป็นการกำหนดประเภทของ field ที่ใช้ในการเก็บข้อมูล
- 2) สร้าง field โดยใช้ SearchableField เป็นการรวมรูปแบบของ field ที่มีการกำหนดคุณสมบัติ retrievable, filterable, sortable, facetable และ searchable เพื่อให้ field นั้นเหมาะสมสำหรับใช้งานในทุกกรณี
- 3) สร้าง field โดยใช้ SearchField เป็น field ที่ใช้ในการค้นหาข้อมูล โดยมีคุณสมบัติที่มีการกำหนดเกี่ยวกับการค้นหาเพิ่มเติม เช่น analyzer_name เป็นการกำหนดชื่อของ analyzer ที่ใช้ในการแยกคำเพื่อค้นหา

ข้อมูลที่ใช้ในการสร้าง field จะมีการระบุข้อมูลดังนี้

- name: ชื่อ field
- type: ประเภทของข้อมูลใน field
- retrievable: กำหนดคุณสมบัติของ field ว่า field นั้นสามารถถูกส่งคืนค่ากลับมาในผลลัพธ์ของการค้นหาได้หรือไม่
- filterable: กำหนดคุณสมบัติของ field ว่าสามารถใช้ในการกรองร่วมกับ field นั้นในการค้นหาได้หรือไม่

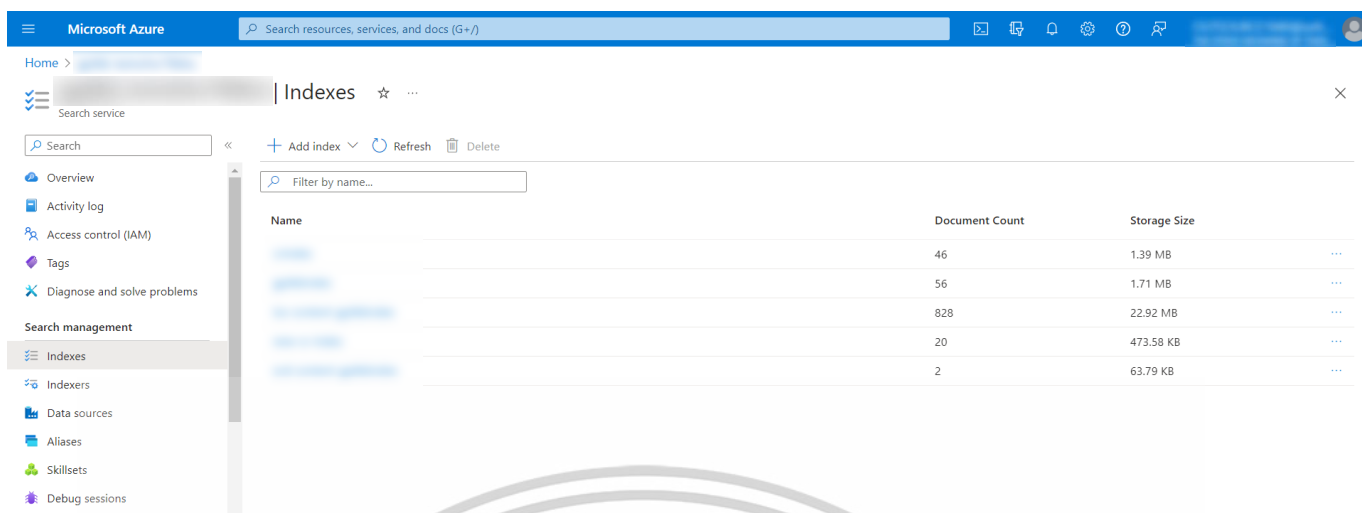
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- sortable: กำหนดคุณสมบัติของ field ว่า field นั้นสามารถใช้ในการเรียงลำดับหรือ sort ผลลัพธ์ของการค้นหาได้หรือไม่
 - facetable: กำหนดคุณสมบัติของ field ว่าสามารถใช้ในการสร้างตารางเรียกดูผลลัพธ์ และแยกแยะข้อมูลของการค้นหาได้หรือไม่
 - searchable: กำหนดคุณสมบัติของ field ว่า field นั้นสามารถใช้ในการค้นหาข้อมูลได้หรือไม่
 - type: ประเภทของการเก็บข้อมูลในรูปแบบ collection จากตัวอย่าง ถ้ากำหนดเป็น Single จะหมายถึงให้เก็บข้อมูลค่าเดียวเท่านั้น หรือการกำหนดเป็น String คือ ประเภทของ collection ที่เก็บข้อมูลจะเป็น string สามารถเก็บข้อมูลที่เป็น string ที่มีหลายค่าได้
 - hidden: กำหนดว่า field นั้นจะถูกซ่อนหรือจะถูกแสดงในผลลัพธ์ของการค้นหาหรือไม่
 - vector_search_dimensions: กำหนดว่า field นั้นเก็บข้อมูลเวกเตอร์กี่มิติ (dimensions)
 - vector_search_configuration: กำหนดว่าในการค้นหาเวกเตอร์จะใช้ค่าตั้งต้นหรือกำหนดค่าความคล้ายของเวกเตอร์เพิ่มเติม
- 4) เมื่อสร้าง index เสร็จแล้ว สามารถดูรายการ index ที่ถูกสร้างขึ้นใหม่หรือ index ที่มีอยู่ทั้งหมดได้จากรายการ indexes ในเมนู search service ของ azure portal

```
fields = [
  SimpleField(name="id", type="Edm.String", key=True),
  SearchableField(name="content", type="Edm.String", analyzer_name="en.microsoft"),
  SearchField(
    name="embedding",
    type=SearchFieldType.Collection(SearchFieldType.Single),
    hidden=False,
    searchable=True,
    filterable=False,
    sortable=False,
    facetable=False,
    vector_search_dimensions=1536,
    vector_search_configuration="default",
  ),
  SimpleField(name="category", type="Edm.String", filterable=True, facetable=True),
  SimpleField(name="sourcepage", type="Edm.String", filterable=True, facetable=True),
  SimpleField(name="sourcefile", type="Edm.String", filterable=True, facetable=True),
]
```

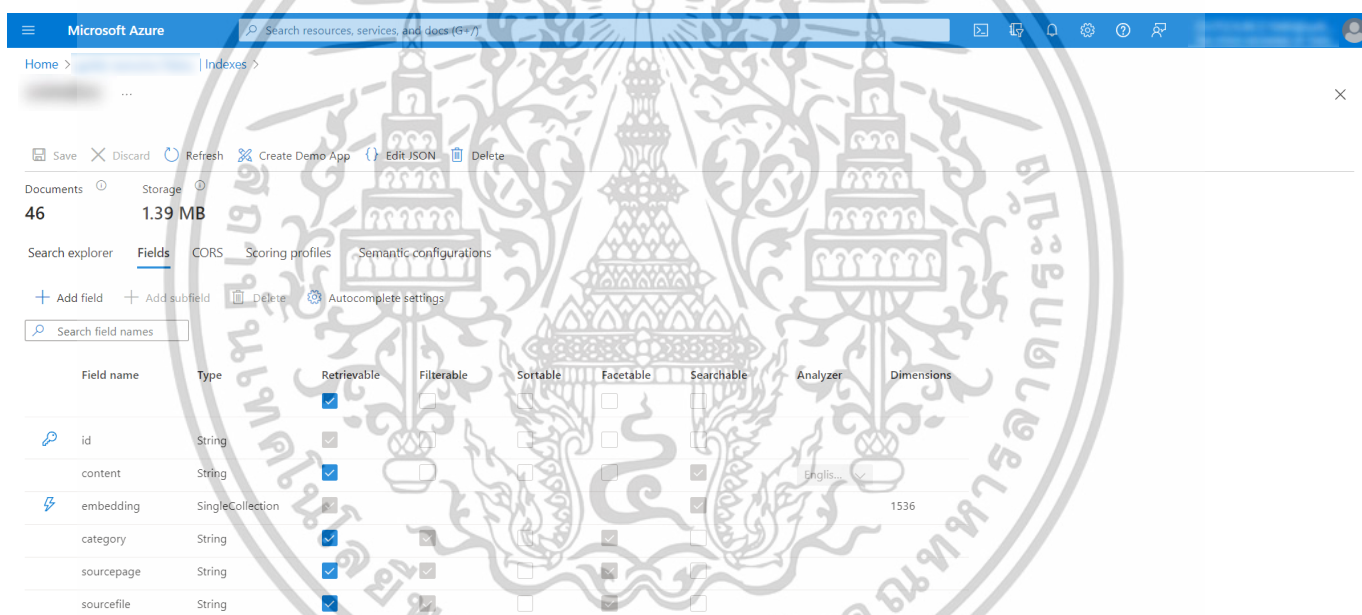
รูปที่ ก.3 การสร้าง field ของ index

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Name	Document Count	Storage Size
	46	1.39 MB
	56	1.71 MB
	828	22.92 MB
	20	473.58 KB
	2	63.79 KB

รูปที่ ก.4 รายชื่อ index ที่แสดงใน azure portal



Field name	Type	Retrievable	Filterable	Sortable	Facetable	Searchable	Analyzer	Dimensions
id	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
content	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Englis...	
embedding	SingleCollection	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		1536
category	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
sourcepage	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
sourcefile	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		

รูปที่ ก.5 รายการ field ใน index ที่ถูกสร้างขึ้นจากโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

การนำเอกสารเข้าสู่กระบวนการค้นหา

กระบวนการนำเอกสารเข้า container

กระบวนการนำเอกสารเข้า container เป็นส่วนที่ใช้ในการอัปโหลดเอกสารไปยัง container ใน Azure Blob Storage ของ Microsoft Azure โดยใช้ Azure SDK for Python

การนำเอกสารเข้า container จะเริ่มจากการสร้างอ็อบเจกต์ BlobServiceClient โดยระบุ URL ของ container และใส่ข้อมูลยืนยันตัวตนเป็นพารามิเตอร์

```
blob_service = BlobServiceClient(
    account_url=f"https://{args.storageaccount}.blob.core.windows.net", credential=storage_creds
)
```

รูปที่ ข.1 การเชื่อมต่อกับ container ใน Azure Blob Storage

จากนั้นจะสร้างอ็อบเจกต์ ContainerClient โดยใช้ get_container_client จาก blob_service และระบุชื่อของ container ที่ต้องการใช้ในพารามิเตอร์ args.container และจะทำการตรวจสอบว่า container นั้นมีอยู่ใน Azure Blob Storage แล้วหรือไม่ หากไม่มี จะสร้าง container ใหม่

```
blob_container = blob_service.get_container_client(args.container)
if not blob_container.exists():
    blob_container.create_container()
```

รูปที่ ข.2 การสร้าง ContainerClient

โดยการทำงานจะมีการตรวจสอบนามสกุลของไฟล์ที่ต้องการอัปโหลด ถ้าเป็นไฟล์ PDF จะใช้ PdfReader จาก PyPDF2 เพื่ออ่านไฟล์ PDF และทำการแยกหน้าใน PDF จากนั้นจะอัปโหลดแต่ละหน้าเป็น blob แยกต่างหากใน container และจะสร้างชื่อของ blob โดยใช้ฟังก์ชัน blob_name_from_file_page() และใช้ upload_blob() เพื่ออัปโหลดข้อมูลแต่ละหน้าใน blob ของ container แต่ถ้าหากไฟล์ไม่ใช่ PDF จะทำการสร้างชื่อของ blob สำหรับไฟล์ทั้งหน้าโดยใช้ฟังก์ชัน blob_name_from_file_page() และใช้ upload_blob() เพื่ออัปโหลดไฟล์ทั้งหน้าใน blob ของ container

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# if file is PDF split into pages and upload each page as a separate blob
if os.path.splitext(filename)[1].lower() == ".pdf":
    reader = PdfReader(filename)
    pages = reader.pages
    for i in range(len(pages)):
        blob_name = blob_name_from_file_page(filename, i)
        if args.verbose:
            print(f"\tUploading blob for page {i} -> {blob_name}")
        f = io.BytesIO()
        writer = PdfWriter()
        writer.add_page(pages[i])
        writer.write(f)
        f.seek(0)
        blob_container.upload_blob(blob_name, f, overwrite=True)
else:
    blob_name = blob_name_from_file_page(filename)
    with open(filename, "rb") as data:
        blob_container.upload_blob(blob_name, data, overwrite=True)

```

รูปที่ ข.3 การอัปโหลดเอกสารเข้า Blob Storage

ฟังก์ชัน `blob_name_from_file_page()` ใช้ในการสร้างชื่อของ blob สำหรับอัปโหลดเอกสารลงใน Azure Blob Storage container โดยหากเป็นไฟล์ PDF จะสร้างชื่อ blob จากไฟล์ที่กำลังอัปโหลดและหน้าที่ต้องการอัปโหลด แต่หากไม่ใช่ไฟล์ PDF จะใช้ชื่อเดิมของไฟล์ และอัปโหลดเอกสารลงใน Azure Blob Storage container

```

def blob_name_from_file_page(filename, page=0):
    if os.path.splitext(filename)[1].lower() == ".pdf":
        return os.path.splitext(os.path.basename(filename))[0] + f"-{page}" + ".pdf"
    else:
        return os.path.basename(filename)

```

รูปที่ ข.4 การสร้างชื่อของ blob สำหรับอัปโหลดเอกสาร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The screenshot shows the Microsoft Azure portal interface. The main content area displays a list of containers under the heading "Containers". The list includes columns for Name, Last modified, Anonymous access level, and Lease state. The containers listed are:

Name	Last modified	Anonymous access level	Lease state
[Redacted]	7/25/2023, 10:14:54 AM	Private	Available
[Redacted]	10/9/2023, 9:47:04 PM	Private	Available
[Redacted]	8/15/2023, 10:14:32 AM	Private	Available
[Redacted]	8/29/2023, 12:55:10 AM	Private	Available
[Redacted]	9/6/2023, 11:58:03 AM	Private	Available
[Redacted]	10/6/2023, 10:49:58 AM	Private	Available
[Redacted]	9/11/2023, 2:48:09 PM	Private	Available
[Redacted]	9/28/2023, 9:00:07 AM	Private	Available

รูปที่ ข.5 รายชื่อ container ที่แสดงใน azure portal

The screenshot shows the Microsoft Azure portal interface for a specific container. The main content area displays a list of blobs under the heading "Overview". The list includes columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. The blobs listed are:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[Redacted]	8/15/2023, 10:18:40 ...	Hot (Inferred)		Block blob	434.56 KiB	Available
[Redacted]	8/15/2023, 10:18:39 ...	Hot (Inferred)		Block blob	96.71 KiB	Available
[Redacted]	8/15/2023, 10:18:39 ...	Hot (Inferred)		Block blob	178.26 KiB	Available
[Redacted]	8/28/2023, 10:10:49 ...	Hot (Inferred)		Block blob	86.01 KiB	Available
[Redacted]	8/28/2023, 10:10:50 ...	Hot (Inferred)		Block blob	114.56 KiB	Available
[Redacted]	8/28/2023, 10:10:51 ...	Hot (Inferred)		Block blob	39.56 KiB	Available
[Redacted]	9/11/2023, 2:57:49 PM	Hot (Inferred)		Block blob	148.65 KiB	Available

รูปที่ ข.6 รายการไฟล์ใน container ที่ถูกอัปโหลดจากโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค

แผนการดำเนินงาน

แผนการดำเนินงานระบบแชทบอทสำหรับองค์กร

การดำเนินงานจะมีการวางแผนการทำงานทุกๆ 2 สัปดาห์ หรือเรียกว่า Sprint โดยระยะเวลาดำเนินงานเริ่มตั้งแต่ 2 มิถุนายน 2566 – 24 พฤศจิกายน 2566 มีการวางแผนดังนี้

TASK	ASSIGNED TO	PROGRESS	START	END	DAY
Sprint 1 Conduct research on Azure services					
Task 1 Learning AZ-900	Piyaporn, Prueksa	100%	2-Jun	6-Jun	4
Task 2 Research Azure Cognitive Search	Piyaporn, Prueksa	100%	6-Jun	8-Jun	2
Task 3 Research Blob Storage	Piyaporn, Prueksa	100%	8-Jun	12-Jun	4
Task 4 Research OpenAI	Piyaporn, Prueksa	100%	12-Jun	15-Jun	3
Sprint 2 Design Application Architecture					
Task 1 List Azure Service	Piyaporn, Prueksa	100%	15-Jun	16-Jun	1
Task 2 Selecting Front-end Language	Piyaporn, Prueksa	100%	19-Jun	21-Jun	2
Task 3 Selecting Back-end Language	Piyaporn, Prueksa	100%	21-Jun	23-Jun	2
Task 4 Design Flow Application	Piyaporn, Prueksa	100%	26-Jun	28-Jun	2
Task 5 Identify Use Case Scenario	Piyaporn, Prueksa	100%	28-Jun	30-Jun	2
Sprint 3 Set up Azure account and provision resources					
Task 1 Flow Chart Service	Piyaporn, Prueksa	100%	3-Jul	6-Jul	3
Task 2 Planning Project Timeline to deployment	Piyaporn, Prueksa	100%	7-Jul	11-Jul	4
Task 3 Learning HTML	Piyaporn, Prueksa	100%	10-Jul	11-Jul	1
Task 4 Learning CSS	Piyaporn, Prueksa	100%	11-Jul	11-Jul	0.5
Task 5 Learning JavaScript	Piyaporn, Prueksa	100%	12-Jul	13-Jul	1
Task 6 Learning React (Type Script)	Piyaporn, Prueksa	100%	13-Jul	14-Jul	1
Task 7 Learning API	Piyaporn, Prueksa	100%	17-Jul	18-Jul	1
Task 8 Prepare Azure Services	Arkaphat, Piyaporn, Prueksa	100%	19-Jul	21-Jul	2
8.1) Azure Service OpenAI					
8.2) Azure Service Cognitive Search					
8.3) Azure Service Blob					
8.4) Database MySQL					
8.5) Azure Service Web App (Hosting)					
8.6) Web Server					
Sprint 4 Development Phase 1					
Task 1 Prove Azure Service By OpenAI Demo	Arkaphat, Piyaporn, Prueksa	100%	24-Jul	27-Jul	3
1.1) Run Server Localhost					
1.2) ศึกษาตัวอย่างโค้ด					
1.3) ศึกษาการสร้าง prompt					
1.4) ไล่ flow การทำงานของโปรแกรม					
Task 2 Setup Project Front & Back End & Database	Arkaphat, Piyaporn, Prueksa	100%	31-Jul	2-Aug	2
Task 3 เชื่อมต่อ azure service	Arkaphat, Piyaporn, Prueksa	100%	2-Aug	4-Aug	2
Sprint 5 Development Phase 2					
Task 1 Chat Backend : รับข้อมูลคำถามผ่าน api, call api cognitive search, call api blob storage, call api openai และแสดงผลลัพธ์	Piyaporn, Prueksa	100%	7-Aug	8-Aug	1
Task 2 ทดลองนำ file ข้อมูลต่างๆ เข้า blob storage	Piyaporn, Prueksa	100%	8-Aug	10-Aug	2
Task 3 แสดง file ที่ได้มาจากการอ้างอิงในคำตอบของ chatbot (ชื่อ file)	Piyaporn, Prueksa	100%	11-Aug	15-Aug	2
Task 4 Index	Piyaporn, Prueksa	100%	11-Aug	16-Aug	3
4.1) สร้าง index					
4.2) ทดลองใช้งาน index ที่ปรับแต่งใหม่					
Task 5 Backend : แก้ไขการทำงานการนำไฟล์เข้า blob storage	Piyaporn, Prueksa	100%	15-Aug	16-Aug	1
Function Chat					
Task 1 Design Prototype	Piyaporn, Prueksa	100%	16-Aug	17-Aug	1
Task 2 Chat Frontend : Nav Bar (แสดง username, ปุ่มรายละเอียด, ปุ่ม logout), ช่องใส่คำถาม, ปุ่มคำถาม, ปุ่ม clear chat	Piyaporn, Prueksa	100%	17-Aug	18-Aug	1

รูปที่ ค.1 แผนการดำเนินงาน Sprint 1-5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาเท่านั้น เมื่ออนุญาตเห็นนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sprint 6 Development Phase 3 (Index & Compare code)						
Task 1 Index	Piyaporn, Prueksa	100%	21-Aug	25-Aug	4	
1.1) สร้าง skillset ใน index						
1.2) เชื่อมต่อและใช้งาน index ผ่านโค้ด						
1.3) สร้าง indexer						
Task 2 เชื่อมต่อ service และทดลองใช้งานโปรแกรมตัวใหม่	Piyaporn, Prueksa	100%	25-Aug	31-Aug	4	
Task 3 เปรียบเทียบการทำงานของโค้ดทั้ง 2 โปรแกรม	Piyaporn, Prueksa	100%	31-Aug	4-Sep	2	
3.1) ศึกษา data flow ของแต่ละโปรแกรม						
3.2) เปรียบเทียบการทำงานและความแม่นยำในการตอบ						
3.3) ทดลองนำไฟลัสมารายชื่อผู้สมัครเข้าและเปรียบเทียบการตอบคำถาม						
Sprint 7 Development Phase 4						
Task 1 POC ISO document with OpenAI	Arkaphat, Piyaporn, Prueksa	100%	6-Sep	10-Sep	5	
1.1) Research limitation of OpenAI API						
Task 2 Update & Investigate preproc.py	Piyaporn, Prueksa	100%	13-Sep	15-Sep	2	
2.1) Know-well with indexing flow Git1, Git2						
2.2) Check Duplicate / Incremental when add more data to index						
Task 3 ทำบทที่ 1,2,3 ของเล่มโปรเจค	Piyaporn, Prueksa	100%	15-Sep	17-Sep	2	
Sprint 8 ทำเล่มโครงการ						
Task 1 ทำบทที่ 1,2,3 ของเล่มโปรเจค	Piyaporn, Prueksa	100%	19-Sep	21-Sep	2	
Task 2 Presentation for BearyGPT (1-3)	Piyaporn, Prueksa	100%	21-Sep	22-Sep	1	
Task 3 Presentation Project to instructor	Piyaporn, Prueksa	100%	22-Sep	25-Sep	1	
Sprint 9 เก็บ Log ของระบบ						
Task 1 Research Best practice Prompt for AI	Prueksa	100%	27-Sep	9-Oct	9	
1.1) Research Best practice Prompt for AI						
1.2) แก้ไขการส่ง Prompt ให้ OpenAI						
Task 2 Log file to keep Answer and Question	Piyaporn	100%	27-Sep	9-Oct	9	
Sprint 10 Test and Integration 1						
Task 1 Log file to keep Answer and Question	Piyaporn	100%	10-Oct	12-Oct	2	
Task 2 Research Validate Method for Answer of Beary GPT	Piyaporn, Prueksa	100%	16-Oct	18-Oct	3	
2.1) Research ROUGE Score						
2.2) Research BLEU Score						
2.3) Research Cosine similarity						
Task 3 Create Validate program Using Cosine Similarity	Piyaporn, Prueksa	100%	19-Oct	20-Oct	2	
Sprint 11 Test and Integration 2						
Task 1 Create Validate program Using Euclidean Distance	Piyaporn, Prueksa	100%	24-Oct	27-Oct	3	
Task 2 ทำเล่มโครงการ	Piyaporn, Prueksa	100%	24-Oct	6-Nov	10	
2.1) บทที่1-5						
2.2) เตรียมสไลด์นำเสนอ						
2.3) เตรียมคำถามที่ไขในการทดสอบ						
Sprint 12 Project Documentation						
Task 1 ทำเล่มโครงการบทที่ 4	Piyaporn, Prueksa	100%	7-Nov	10-Nov	4	
Task 2 ทำเล่มโครงการบทที่ 5	Piyaporn, Prueksa	100%	13-Nov	15-Nov	3	
Task 3 ทำสไลด์นำเสนอ	Piyaporn, Prueksa	100%	16-Nov	20-Nov	3	

รูปที่ ค.2 แผนการดำเนินงาน Sprint 6-12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำรับรองเล่มสหกิจศึกษาโดยสถานประกอบการ

วันที่ 8 เดือน เมษายน พ.ศ. 2567

Senior Officer - Senior

ข้าพเจ้า ดร.กรภัทร์ ปทุมวงศ์ ตำแหน่ง Application Engineer เป็นตัวแทนของสถานประกอบการ สหกรณ์การเกษตรแห่งประเทศไทย ขอรับรองว่าทางสถานประกอบการได้ตรวจสอบเล่มสหกิจศึกษา เรื่อง Beary OPT: ระบบแม่ข่ายคอมพิวเตอร์ ของนักศึกษาชื่อ อภิชากรณ์ ด้วงจิตรวิวัฒน์ ซึ่งเป็นนักศึกษา ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เรียบร้อยแล้ว และไม่มีส่วนหนึ่งส่วนใดในเล่มสหกิจศึกษาที่มีข้อมูลอ่อนไหว และ/หรือ ข้อมูลอันเป็นความลับอัน จะก่อให้เกิดความเสียหายต่อสถานประกอบการ รวมทั้งอนุญาตให้สามารถเผยแพร่ต่อสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังได้ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ ดร.กรภัทร์ ปทุมวงศ์

(ดร.กรภัทร์ ปทุมวงศ์)

ตัวแทนสถานประกอบการ

ข้าพเจ้า ดร.จักรพันธ์ เตโช อาจารย์ที่ปรึกษาสหกิจศึกษา ได้ตรวจสอบเล่มสหกิจศึกษาแล้วและรับทราบ ว่าสถานประกอบการดำเนินการตรวจสอบเล่มสหกิจศึกษาแล้ว จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ ดร.จักรพันธ์ เตโช

(ดร.จักรพันธ์ เตโช)

อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



งานทะเบียนคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

คำรับรองเล่มโครงการพิเศษ/ปัญหาพิเศษ/สหกิจศึกษา

วันที่...5...เดือน.....เมษายน.....พ.ศ..2567...

ข้าพเจ้า นางสาว...ปิยาภรณ์ ตั้งจิตสิริสิน..... รหัสประจำตัว...63050157.....

นางสาว...พุกษา ขำแป้น..... รหัสประจำตัว...63050160.....

นักศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา...วิทยาการคอมพิวเตอร์..... ภาควิชา...วิทยาการคอมพิวเตอร์.....

ขอรับรองว่าสหกิจศึกษา เรื่อง

ชื่อภาษาไทย...ระบบเซทบอทสำหรับองค์กร.....

ชื่อภาษาอังกฤษ... Beary GPT: Generative AI for SET.....

ปีการศึกษา...2566.....

เป็นผลงานวิจัยที่มีได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อนเรียบร้อยแล้ว และได้แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่มสหกิจศึกษามบับสมบูรณ์แล้ว

โปรแกรมอักขราวิสุทธิ์...0.28.....% หรือโปรแกรม Turnitin.....%

ลงชื่อ...ปิยาภรณ์ ตั้งจิตสิริสิน.....

ลงชื่อ...พุกษา ขำแป้น.....

(ปิยาภรณ์ ตั้งจิตสิริสิน)

(พุกษา ขำแป้น)

นักศึกษา

นักศึกษา

ข้าพเจ้า ดร. ...จักรพันธ์ เตโชยา..... อาจารย์ที่ปรึกษา สหกิจศึกษา ได้ตรวจสอบสหกิจศึกษาของนักศึกษาข้างต้น แล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ...จักรพันธ์ เตโชยา.....

อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้