ระบบวิเคราะห์การประชุมสำหรับการติดตามผู้พูดและการสรุปผลด้วยโมเดลภาษาขนาดใหญ่
A Meeting Analytics System for Speaker Tracking and LLM-Based Summarization

รัตนชาติ ทองมนต์[1], รัศมี แสนอุบล[1], สุระเจตน์ อ่อนฤทธิ์[1], เอกรินทร์ วทัญญูเลิศสกุล[1] และ พัฒนา ศรีชาลี[1*]

[1]คณะเทคโนโลยีอุตสาหกรรม, มหาวิทยาลัยราชภัฏอุบลราชธานี

Rattanachad Thongmon[1], Radsamee Sanubon[1], Surajate On-rit[1],

Egkarin Watanyulertsakul[1] and Pattana Srichalee[1*]

[1]Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University

Corresponding Author Email : Pattana.s@ubru.ac.th

Received 05 January 2026

Revised 30 April 2026

Accepted 11 May 2026

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบวิเคราะห์การประชุมสำหรับการติดตามผู้พูดและสรุปผลการประชุมแบบอัตโนมัติ โดยใช้การระบุทิศทางเสียงเพื่อควบคุมการหมุนกล้อง และประยุกต์ใช้โมเดลภาษาแบบขนาดใหญ่ สำหรับสังเคราะห์เนื้อหา ระบบต้นแบบประกอบด้วยไมโครโฟนทิศทางร่วมกับกลไกหมุนกล้อง 360 องศา พัฒนาด้วยภาษา Python บนบอร์ด Raspberry Pi 5 และออกแบบส่วนติดต่อผู้ใช้ด้วย PyQt5 ระบบเชื่อมต่อกับ Microsoft Azure Speech Service สำหรับการถอดความเสียง และใช้โมเดลภาษา DeepSeek ร่วมกับเทคนิคการสร้างข้อความโดยใช้การค้นคืนสารสนเทศ (RAG) และการค้นหาเชิงความหมาย เพื่อเพิ่มความแม่นยำและความถูกต้องของข้อมูลอ้างอิงจากฐานข้อมูลการประชุม

การประเมินประสิทธิภาพของระบบดำเนินการในห้องประชุมจำลองที่มีระดับเสียงรบกวนต่ำ มีผู้เข้าร่วม 8-10 คน อยู่ในระยะห่างจากอุปกรณ์ 1–2 เมตร ผลการทดสอบการระบุทิศทางเสียงจำนวน 200 ครั้ง พบว่าระบบมีความถูกต้องเฉลี่ยร้อยละ 98.50 สำหรับการถอดความบทสนทนาทั่วไปมีค่าอัตราคำผิดพลาด (WER) ร้อยละ 5.56 ขณะที่ประโยคที่มีศัพท์เทคนิคและมีความซับซ้อนทางภาษาสูงมีค่าอัตราคำผิดพลาดร้อยละ 15.82 และ 17.88 ตามลำดับ ในส่วนของการประเมินความถูกต้องของการติดตามผู้พูดมีค่าอัตราความผิดพลาดในการแยกแยะผู้พูด (DER) โดยรวมที่ร้อยละ 14.54 ± 1.85 นอกจากนี้ การประเมินคุณภาพบทสรุปด้วยเกณฑ์จีอีแวล โดยใช้โมเดลภาษาเป็นผู้ประเมิน ครอบคลุมมิติด้านความสอดคล้อง ความเชื่อมโยง ความลื่นไหลทางภาษา และความเกี่ยวข้องกับต้นฉบับ ให้คะแนนเฉลี่ย 4.55 ± 0.53 จาก 5 คะแนน ผลการทดลองสะท้อนให้เห็นถึงประสิทธิภาพของระบบภายใต้สภาวะควบคุม และบ่งชี้ถึงศักยภาพการใช้เพื่อการจัดการความรู้และการประชุมอัจฉริยะในบริบทการใช้งานจริงได้

คำสำคัญ: ระบบวิเคราะห์การประชุม, การติดตามผู้พูด, โมเดลภาษาขนาดใหญ่, การสรุปผลการประชุม

Abstract

This research aimed to develop a meeting analysis system for automated speaker tracking and meeting summarization. The system employed sound direction identification to control a motorized 360-degree camera system and utilizes a large language model (LLM) for content synthesis. The prototype integrated directional microphones with a camera rotation module and was developed in Python on a Raspberry Pi 5 platform, with a graphical user interface developed using PyQt5. The system was connected to Microsoft Azure Speech Service for speech transcription and incorporates the DeepSeek large language model combined with a Retrieval-Augmented Generation (RAG) framework and semantic search to enhance factual accuracy and reference consistency from a meeting knowledge base.

System performance was evaluated in a simulated low-noise meeting environment involving 8-10 participants positioned 1–2 meters from the device. Experimental results from 200 sound direction identification trials demonstrate an average accuracy of 98.50%. For speech transcription, the word error rate (WER) for general conversational speech was 5.56%, whereas speech containing technical terminology and higher linguistic complexity yielded WERs of 15.82% and 17.88%, respectively. Speaker tracking evaluation achieved an overall diarization error rate (DER) of 14.54 ± 1.85%. Summary quality was assessed using the G-Eval metric, with a language model serving as the evaluator, and achieved an average score of 4.55 ± 0.53 out of 5 across coherence, relevance, fluency, and content consistency dimensions. These results highlighted the effectiveness of the proposed approach under controlled experimental conditions and indicated its potential for knowledge management and intelligent meeting applications in real-world scenarios.

Keywords: Meeting Analysis System, Speaker Tracking, Large Language Models, Meeting Summarization

1. บทนำ

การประชุมเป็นกระบวนการดำเนินงานที่มีความสำคัญในทุกองค์กร เนื่องจากเป็นเวทีสำหรับการสื่อสาร การแลกเปลี่ยนข้อมูล และการระดมความคิดเห็นของบุคลากรในรูปแบบกลุ่ม เพื่อกำหนดทิศทางการทำงาน การตัดสินใจเชิงนโยบาย และการประสานความร่วมมือในระดับต่าง ๆ ดังนั้น การจัดทำรายงานการประชุมที่ถูกต้อง ครบถ้วน และทันเวลา จึงมีความสำคัญยิ่งต่อการบริหารจัดการและการติดตามผลการดำเนินงานขององค์กร อย่างไรก็ตาม วิธีการบันทึกแบบดั้งเดิมซึ่งอาศัยการจดบันทึกด้วยมือมักเผชิญข้อจำกัดหลากหลายประการ เช่น การสูญเสียข้อมูลสำคัญอันเนื่องจากความเร็วในการพูดสูงกว่าความเร็วที่ใช้ในการเขียน รวมถึงความไม่สอดคล้องของรูปแบบการบันทึก และภาระงานที่เพิ่มขึ้นของผู้จดบันทึก ซึ่งส่งผลทำให้ไม่สามารถมีส่วนร่วมในการประชุมได้อย่างเต็มที่ [1] แม้ภายหลังจะมี

การนำอุปกรณ์บันทึกเสียงมาใช้เพื่อแก้ไขปัญหาดังกล่าว แต่กระบวนการถอดความจากไฟล์เสียงยังคงต้องดำเนินการด้วยมือหรือใช้เครื่องมือแปลงข้อความเพิ่มเติม ทำให้เกิดภาระงานและค่าใช้จ่ายโดยรวมเพิ่มขึ้น

ในปัจจุบัน เทคโนโลยีปัญญาประดิษฐ์ได้มีการพัฒนาความสามารถและมีบทบาทที่สำคัญในกิจกรรมด้านการสื่อสารและการประมวลผลภาษา ทั้งด้านการแปลภาษา การสรุปเนื้อหา และการวิเคราะห์ข้อมูลเชิงความหมาย [2] โดยสามารถเรียนรู้ ทำความเข้าใจ และสร้างภาษาธรรมชาติได้ในระดับใกล้เคียงมนุษย์ เทคโนโลยีดังกล่าวช่วยเพิ่มประสิทธิภาพในการสรุปเนื้อหา การวิเคราะห์ความรู้สึก และการตอบคำถาม ซึ่งเป็นเครื่องมือสำคัญต่อการจัดการข้อมูลปริมาณมาก [3] โดยเฉพาะเทคโนโลยีที่ใช้โมเดลภาษาขนาดใหญ่ (Large Language Models: LLMs) เช่น GPT, BERT และ T5 ซึ่งมีศักยภาพที่โดดเด่นทั้งในงานประมวลผลข้อความและการวิเคราะห์ข้อมูลเชิงบริบท อย่างไรก็ตาม การนำโมเดลเหล่านี้ไปใช้งานบนอุปกรณ์ที่มีทรัพยากรจำกัด เช่น ไมโครคอนโทรลเลอร์ และคอมพิวเตอร์บอร์ดเดี่ยว จำเป็นต้องอาศัยเทคนิคการลดขนาดโมเดล เพื่อให้สามารถใช้งานได้อย่างมีประสิทธิภาพ ขณะเดียวกันยังมีข้อได้เปรียบด้านความเป็นส่วนตัวของข้อมูล เนื่องจากไม่จำเป็นต้องส่งข้อมูลเสียงหรือข้อความไปประมวลผลบนระบบคลาวด์ รวมถึงยังช่วยลดต้นทุนด้านโครงสร้างพื้นฐานที่ใช้ในการออกแบบและพัฒนา [4]

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบวิเคราะห์การประชุมที่สามารถติดตามผู้พูดโดยอาศัยการประมวลผลทิศทางเสียงร่วมกับภาพ และควบคุมการหันกล้องอัตโนมัติ พร้อมทั้งถอดความและสรุปผลการประชุมอัตโนมัติด้วยโมเดลภาษาขนาดใหญ่ โดยประยุกต์ใช้เทคนิคการสร้างข้อความโดยใช้การค้นคืนสารสนเทศ (Retrieval-Augmented Generation: RAG) เพื่อช่วยให้โมเดลสามารถเข้าถึงฐานข้อมูลภายนอกหรือคลังข้อมูลขนาดใหญ่ และสร้างคำตอบที่อ้างอิงจากแหล่งข้อมูลจริงได้อย่างแม่นยำ [5] นอกจากนี้ ระบบยังได้รับการออกแบบให้รองรับการทำงานแบบผสานรวมระหว่างการประมวลผลบนคลาวด์สาธารณะกับโครงสร้างพื้นฐานบนอุปกรณ์ขอบเครือข่าย (Edge Computing) โดยใช้ Raspberry Pi 5 เป็นต้นแบบของระบบ เพื่อเพิ่มความปลอดภัยและรักษาความเป็นส่วนตัวของข้อมูลการประชุม เพื่อยกระดับประสิทธิภาพการจัดการข้อมูลการประชุม ลดภาระงานของบุคลากร และเพิ่มความสะดวกในการสืบค้นสารสนเทศย้อนหลังสำหรับองค์กร

2. ขอบเขตงานวิจัย

งานวิจัยนี้มุ่งเน้นการพัฒนาระบบต้นแบบและประเมินประสิทธิภาพภายใต้ขอบเขตดังนี้

2.1 ขอบเขตด้านฮาร์ดแวร์ พัฒนาระบบโดยใช้คอมพิวเตอร์บอร์ดเดี่ยว Raspberry Pi 5 หน่วยความจำ 8 กิกะไบต์ เป็นหน่วยประมวลผลหลัก เชื่อมต่อกับกล้อง Camera Module 3 ที่ติดตั้งบนชุดขับเคลื่อนเซอร์โวมอเตอร์ รุ่น MG996R จำนวน 2 ตัว เพื่อควบคุมทิศทางกล้องแบบ 360 องศา ระบบรับเสียงใช้ไมโครโฟนประชุม Poly Sync 10 ร่วมกับโมดูลไมโครคอนโทรลเลอร์ ESP32 สำหรับประมวลผลทิศทางเสียง

2.2 ขอบเขตด้านซอฟต์แวร์และฟังก์ชัน ระบบถูกพัฒนาด้วยภาษา Python บนระบบปฏิบัติการ Raspberry Pi OS (64-bit) ครอบคลุมฟังก์ชันหลัก ได้แก่ 1) การติดตามผู้พูดอัตโนมัติด้วยเสียง 2) การระบุตัวตนผู้เข้าร่วมประชุมด้วยใบหน้า 3) การถอดความเสียงเป็นข้อความ รองรับภาษาไทยและอังกฤษผ่านบริการ Azure Speech Service เพื่อแปลงสัญญาณเสียงการสนทนาเป็นข้อความแบบเรียลไทม์

ส่วนการจัดเก็บข้อมูลที่มีโครงสร้าง เช่น ทะเบียนประวัติผู้เข้าร่วมประชุมและบันทึกเวลาการประชุม ใช้ฐานข้อมูล SQLite และประยุกต์ใช้ระบบการค้นหาข้อความเต็ม (Full-Text Search: FTS) ของ SQLite เพื่อสร้างดัชนีข้อความ (Indexing) โดยใช้ DeepSeek API ซึ่งเป็นโมเดลภาษาขนาดใหญ่ (LLM) ทำหน้าที่หลักในการประมวลผลสรุปสาระสำคัญจากการประชุม ร่วมกับเทคนิค RAG ในรูปแบบวิธีการค้นหาเชิงความหมาย (Semantic Search) เพื่อเพิ่มความแม่นยำของข้อมูล

2.3 ขอบเขตด้านการทดสอบ ดำเนินการทดสอบในห้องประชุมจำลองภายใต้สภาพแวดล้อมควบคุมขนาดเล็กถึงขนาดกลาง ที่มีพื้นที่ประมาณ 20–30 ตารางเมตร โดยมีระดับเสียงพื้นหลังไม่เกิน 35 dBA เพื่อจำกัดผลกระทบจากเสียงรบกวนภายนอก กำหนดให้มีจำนวนผู้เข้าร่วมประชุมไม่เกิน 10 คน และแต่ละคนนั่งห่างจากระบบบันทึกเสียงไม่เกิน 2 เมตร การประเมินความแม่นยำของการถอดความเสียงใช้อัตราคำผิดพลาด (Word Error Rate: WER) โดยเปรียบเทียบผลลัพธ์กับข้อความอ้างอิงจากมนุษย์ ขณะที่การประเมินความถูกต้องของการติดตามผู้พูดใช้ อัตราความผิดพลาดในการแยกแยะผู้พูด (Diarization Error Rate: DER) เพื่อวัดประสิทธิภาพการแยกและระบุผู้พูดของระบบ

3. ทฤษฎีที่เกี่ยวข้อง

งานวิจัยนี้ได้ประยุกต์ใช้องค์ความรู้และทฤษฎีทางวิศวกรรมคอมพิวเตอร์และปัญญาประดิษฐ์ เพื่อพัฒนาระบบวิเคราะห์การประชุมที่มีประสิทธิภาพ โดยมีรายละเอียดดังนี้

3.1 การระบุทิศทางของแหล่งกำเนิดเสียง

การระบุตำแหน่งของผู้พูดอาศัยหลักการความแตกต่างของเวลาที่เสียงเดินทางมาถึงไมโครโฟนแต่ละตัว (Time Difference of Arrival: TDOA) ซึ่งเป็นเทคนิคที่นิยมใช้ในระบบอาร์เรย์ไมโครโฟน เมื่อมีแหล่งกำเนิดเสียงเกิดขึ้น คลื่นเสียงจะเดินทางมาถึงไมโครโฟนแต่ละตัวในเวลาที่ต่างกันขึ้นอยู่กับระยะทาง ความสัมพันธ์นี้สามารถคำนวณหามุมตกกระทบของเสียง (θ) โดยใช้ความสัมพันธ์ระหว่างระยะห่างของไมโครโฟน (d) ความเร็วเสียง (c) และความต่างของเวลา (Δt) ดังสมการที่ 1

$$\theta = arcsin\left(\frac{c \cdot \Delta t}{d}\right) \tag{1}$$

3.2 การรู้จำใบหน้า

เพื่อระบุว่าใบหน้าที่ตรวจพบตรงกับบุคคลใด ผลลัพธ์ที่ได้จะถูกส่งต่อไปยังระบบหรือโปรแกรมอื่นเพื่อใช้ในการประมวลผลเพิ่มเติม เช่น การระบุตัวตนเพื่อความมั่นคงปลอดภัยของระบบเครือข่าย การจัดทำดัชนีสำหรับการสืบค้นข้อมูลเนื้อหา และการบีบอัดวิดีโอ [6] กระบวนการดังกล่าวดำเนินการโดยการแปลงภาพใบหน้าให้อยู่ในรูปของเวกเตอร์เชิงตัวเลขเพื่อใช้เป็นตัวแทนเอกลักษณ์ของใบหน้านั้น จากนั้นคำนวณระยะห่างแบบยูคลิด (Euclidean Distance) ระหว่างเวกเตอร์ของใบหน้าที่ตรวจจับได้กับเวกเตอร์ใบหน้าต้นแบบในฐานข้อมูล หากค่าระยะห่างน้อยกว่าเกณฑ์ที่กำหนด (Threshold) จะทำให้สามารถระบุว่าเป็นบุคคลเดียวกันได้ โดยคำนวณระยะห่างเวกเตอร์ได้จากสมการที่ 2 [7]

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

เมื่อ p คือ เวกเตอร์คุณลักษณะของใบหน้าต้นแบบ

q คือ เวกเตอร์คุณลักษณะของใบหน้าที่ตรวจจับได้

3.3 การสร้างข้อความโดยใช้การค้นคืนสารสนเทศ

การสร้างข้อความโดยใช้การค้นคืนสารสนเทศ หรือ RAG เป็นเทคนิคที่ช่วยเพิ่มความแม่นยำและความน่าเชื่อถือของโมเดลภาษาขนาดใหญ่ โดยการเชื่อมต่อโมเดลเข้ากับฐานข้อมูลความรู้ภายนอก กระบวนการทำงานของ RAG ประกอบด้วยสองส่วนหลัก ได้แก่ 1) กระบวนการค้นคืน ซึ่งทำหน้าที่ค้นหาเอกสารหรือข้อความที่มีความเกี่ยวข้องกับคำถามหรือบริบทจากฐานข้อมูล และ 2) กระบวนการสร้างข้อความ ซึ่งนำข้อมูลที่ค้นคืนได้มาใช้เป็น

บริบท เพื่อให้โมเดลภาษาขนาดใหญ่ประมวลผลและสร้างคำตอบที่มีความถูกต้อง สอดคล้องกับข้อมูลอ้างอิง และสามารถลดปัญหาการสร้างข้อมูลเท็จ (Hallucination) ของระบบปัญญาประดิษฐ์ โดยกระบวนการ RAG เริ่มต้นจากผู้ใช้งานป้อนคำถามหรือคำสั่งซึ่งเป็นอินพุตเข้าไปในระบบ จากนั้นอินพุตจะถูกแปลงให้อยู่ในรูปเวกเตอร์และนำไปเปรียบเทียบกับข้อมูลที่จัดเก็บไว้ในฐานความรู้หรือฐานข้อมูลเวกเตอร์ เพื่อค้นหาข้อมูลที่มีความหมายใกล้เคียงกับคำถามด้วยวิธีการค้นหาเชิงความหมาย เมื่อได้ข้อมูลที่เกี่ยวข้องแล้ว ระบบจะดำเนินการเตรียมข้อมูล เช่น การคัดกรอง การตัดทอน หรือการจัดรูปแบบเนื้อหา เพื่อให้เหมาะสมต่อการนำไปใช้ร่วมกับโมเดลภาษา ขั้นตอนสุดท้ายคือการนำคำถามเดิมรวมกับข้อมูลอ้างอิงที่ผ่านการเตรียมแล้วส่งเข้าสู่โมเดลภาษา เพื่อสร้างคำตอบที่มีความถูกต้อง ชัดเจน และอ้างอิงจากแหล่งข้อมูลจริง [8]

4. การดำเนินการวิจัย

การวิจัยนี้เป็นใช้รูปแบบการวิจัยและพัฒนา มีวัตถุประสงค์เพื่อพัฒนาระบบผู้ช่วยการประชุมอัจฉริยะ และประเมินประสิทธิภาพการทำงานของระบบในสภาพแวดล้อมการใช้งานจริง คณะผู้วิจัยได้กำหนดระเบียบวิธีวิจัยที่ครอบคลุมองค์ประกอบสำคัญ ได้แก่ ประชากรและกลุ่มตัวอย่าง เครื่องมือที่ใช้ในการวิจัย วิธีการเก็บรวบรวมข้อมูล และการวิเคราะห์ผลการทดลอง ซึ่งมีรายละเอียดดังต่อไปนี้

4.1 ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการศึกษาคือบุคคลทั่วไปที่สื่อสารด้วยภาษาไทยและภาษาอังกฤษ สำหรับกลุ่มตัวอย่างที่ใช้ในการทดสอบระบบ คัดเลือกด้วยวิธีการสุ่มแบบเจาะจง จำนวน 10 คน ประกอบด้วยเพศชาย 5 คน และเพศหญิง 5 คน อายุระหว่าง 20–35 ปี เพื่อให้ครอบคลุมความหลากหลายของระดับเสียง สำเนียงการพูด และลักษณะทางกายภาพสำหรับการทดสอบระบบรู้จำใบหน้า

4.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการพัฒนาและทดสอบระบบ แบ่งออกเป็น 2 ส่วนหลัก ได้แก่

4.2.1 สถาปัตยกรรมฮาร์ดแวร์

ระบบต้นแบบถูกพัฒนาให้อยู่ในรูปแบบอุปกรณ์ประมวลผลแบบ Hybrid Edge-Cloud เพื่อรักษาสมดุลระหว่างประสิทธิภาพและความปลอดภัยของข้อมูล โดยมีองค์ประกอบการเชื่อมต่อดังแสดงในรูปที่ 1 ประกอบด้วยหน่วยประมวลผลหลักใช้บอร์ด Raspberry Pi 5 เป็นอุปกรณ์ขอบเครือข่าย ทำหน้าที่เป็นศูนย์กลางในการประมวลผลข้อมูล ระบบรับภาพใช้กล้อง Raspberry Pi Camera Module 3 เชื่อมต่อผ่านพอร์ต CSI ส่วนรับเสียงใช้ไมโครโฟนสำหรับการประชุมรุ่น Poly Sync 10 เชื่อมต่อผ่านพอร์ต USB 3.0 เพื่อรับสัญญาณเสียงหลักของการประชุม ร่วมกับชุดไมโครโฟนรุ่น MAX4466 จำนวน 4 ตัว ซึ่งเชื่อมต่อกับไมโครคอนโทรลเลอร์ ESP32 ทำหน้าที่ประมวลผลข้อมูลทิศทางแหล่งกำเนิดเสียง และส่งพิกัดตำแหน่งผู้พูดไปยังหน่วยประมวลผลหลัก ข้อมูลดังกล่าวถูกนำไปใช้ควบคุมการปรับทิศทางการรับภาพและเสียงของระบบผ่านเซอร์โวมอเตอร์จำนวน 2 ตัว เพื่อให้สามารถติดตามผู้พูดได้อย่างอัตโนมัติ

4.2.2 สถาปัตยกรรมซอฟต์แวร์

ระบบวิเคราะห์การประชุมสำหรับการติดตามผู้พูดและการสรุปผลด้วยโมเดลภาษาขนาดใหญ่ พัฒนาด้วยโปรแกรมภาษา Python คณะผู้วิจัยได้แบ่งกระบวนการออกแบบและพัฒนาซอฟต์แวร์ออกเป็นการทำงานหลัก 3 ส่วน ดังนี้

1) ส่วนติดตามและระบุตัวตนผู้พูด ทำหน้าที่ประมวลผลและผสานข้อมูลพิกัดทิศทางเสียงที่ได้รับจากไมโครคอนโทรลเลอร์ ESP32 ร่วมกับอัลกอริทึมการรู้จำใบหน้า โดยใช้การคำนวณระยะห่างแบบยูคลิด และกำหนดค่าเกณฑ์ตัดสิน (Threshold) ที่ระดับ 0.6 เพื่อระบุและติดตามผู้พูดในระหว่างการประชุมแบบอัตโนมัติ

2) ส่วนการถอดความและสรุปผลการประชุม ทำหน้าที่แปลงสัญญาณเสียงเป็นข้อความโดยใช้บริการ Azure Speech Service จากนั้นนำข้อความที่ได้ไปประมวลผลร่วมกับโมเดลภาษาขนาดใหญ่ผ่าน DeepSeek API โดยประยุกต์ใช้เทคนิค RAG ในส่วนการสรุปผล เพื่อดึงข้อมูลบริบท (Context) จากฐานข้อมูลมาประกอบกับคำสั่ง (Prompt) ในรูปแบบโครงสร้างลำดับขั้น (Structured Prompting) ก่อนส่งให้โมเดลภาษา DeepSeek ดำเนินการสรุปสาระสำคัญของการประชุม

3) ส่วนต่อประสานกราฟิกกับผู้ใช้ พัฒนาด้วยไลบรารี PyQt5 ทำหน้าที่แสดงผลข้อมูลการประชุม การติดตามผู้พูด และผลการสรุปเนื้อหา รวมถึงรองรับการควบคุมและตั้งค่าการทำงานของระบบโดยผู้ใช้งาน

4.3 ส่วนการดำเนินการวิจัย

คณะผู้วิจัยได้กำหนดกระบวนการดำเนินงานโดยอาศัยกรอบแนวคิดวงจรการพัฒนาระบบ (System Development Life Cycle: SDLC) ในรูปแบบของแบบจำลองน้ำตก ซึ่งเป็นลำดับขั้นที่ต่อเนื่องและมีโครงสร้างชัดเจนเหมือนน้ำตกที่ไหลจากบนลงล่าง ด้วยการกำหนดความต้องการและขอบเขตของระบบชัดเจน เน้นการวางแผนอย่างเป็นระบบ โดยสามารถแบ่งขั้นตอนการดำเนินงานออกเป็น 5 ระยะ [9] ดังนี้

4.3.1 ระยะการวิเคราะห์ความต้องการ

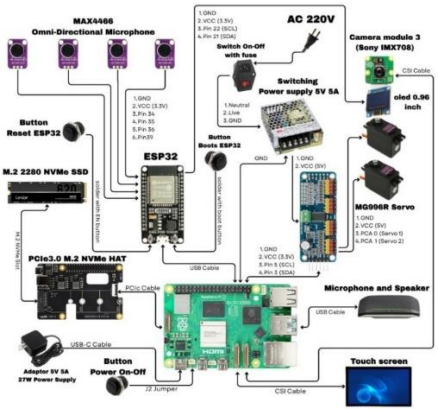
ขั้นตอนนี้จะมุ่งเน้นการกำหนดขอบเขตและข้อกำหนดของระบบให้มีความชัดเจน โดยคณะผู้วิจัยได้รวบรวมข้อมูลจากทฤษฎีที่เกี่ยวข้อง เช่น การระบุทิศทางเสียง เทคโนโลยีโมเดลภาษาขนาดใหญ่ และขีดจำกัดทางเทคนิคของบอร์ด Raspberry Pi 5 เพื่อทำเป็นข้อกำหนดความต้องการทางเทคนิคสำหรับส่วนฮาร์ดแวร์และซอฟต์แวร์

4.3.2 ระยะการออกแบบระบบ

คณะผู้วิจัยได้นำข้อกำหนดความต้องการมาแปลงเป็นแบบจำลองทางเทคนิค โดยแบ่งการออกแบบออกเป็น 3 ส่วนหลัก ได้แก่ ฮาร์ดแวร์ ซอฟต์แวร์ และข้อมูล ดังนี้

1) การออกแบบสถาปัตยกรรมฮาร์ดแวร์

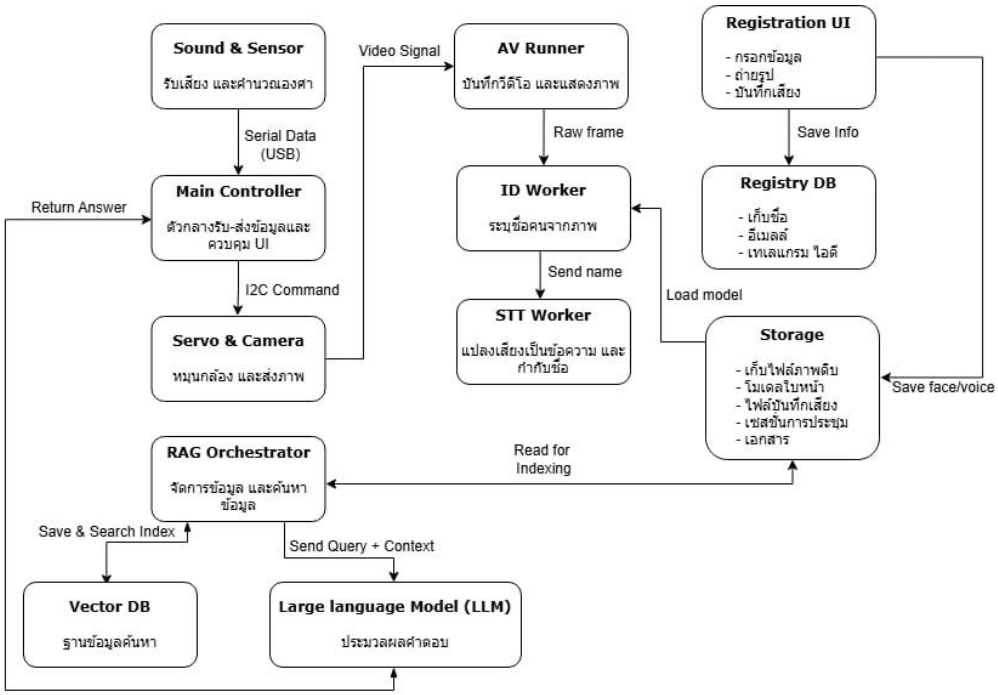
การออกแบบสถาปัตยกรรมฮาร์ดแวร์ของระบบใช้คอมพิวเตอร์บอร์ดเดี่ยว Raspberry Pi 5 เป็นหน่วยประมวลผลหลัก ทำหน้าที่ควบคุมการทำงานโดยรวมของระบบ ภาคขับเคลื่อนมอเตอร์ใช้บอร์ดควบคุมรุ่น PCA9685 สำหรับสั่งงานเซอร์โวมอเตอร์อย่างมีเสถียรภาพ นอกจากนี้ ได้ติดตั้งบอร์ดไมโครคอนโทรลเลอร์ ESP32 เชื่อมต่อกับหน่วยประมวลผลหลักผ่านพอร์ต USB เพื่อทำหน้าที่รับและประมวลผลข้อมูลสัญญาณเสียงจากไมโครโฟนรุ่น MAX4466 จำนวน 4 ตัว และมีส่วนกลไกอัจฉริยะที่ติดตั้งชุดเซอร์โวมอเตอร์รุ่น MG996R ในลักษณะการวางซ้อนแบบสแต็ค เพื่อควบคุมการหมุนและการปรับมุมของกล้อง พร้อมทั้งติดตั้งจอแสดงผลแบบ OLED ขนาด 0.96 นิ้ว สำหรับแสดงสถานะและระดับความดังของเสียง ระบบรองรับการแสดงผลข้อมูลบันทึกการประชุมและผลการสรุปการประชุมผ่านหน้าจอสัมผัสขนาด 7 นิ้ว ทั้งนี้ รายละเอียดการออกแบบวงจรและการเชื่อมต่ออุปกรณ์ แสดงดังรูปที่ 1



รูปที่ 1 โครงสร้างการออกแบบสถาปัตยกรรมฮาร์ดแวร์

2) การออกแบบกระบวนการทำงานของซอฟต์แวร์

การออกแบบและพัฒนาซอฟต์แวร์ของระบบนี้ใช้สถาปัตยกรรมการประมวลผลแบบหลายเธรด โดยพัฒนาด้วยภาษา Python ร่วมกับเฟรมเวิร์ก PyQt5 เพื่อให้ระบบสามารถดำเนินการหลายกระบวนการพร้อมกันได้อย่างมีประสิทธิภาพและลดความหน่วงในการทำงาน



รูปที่ 2 ผังงานกระบวนการทำงานของซอฟต์แวร์

จากรูปที่ 2 กระบวนการทำงานของซอฟต์แวร์ในงานวิจัยนี้แบ่งออกเป็นระบบย่อยที่ทำงานประสานกัน 3 ส่วน ดังรายละเอียดต่อไปนี้

2.1) ระบบติดตามและระบุตัวตนผู้พูด ทำหน้าที่รับข้อมูลทิศทางของแหล่งกำเนิดเสียงจากบอร์ด ESP32 ผ่านการเชื่อมต่อแบบ Serial เพื่อนำไปควบคุมการหมุนของเซอร์โวมอเตอร์ให้กล้องหันไปยังตำแหน่งของผู้พูด พร้อมทั้งประมวลผลภาพใบหน้าเพื่อใช้ในการระบุตัวตนของผู้พูด

2.2) ระบบบันทึกและสรุปผลการประชุม ทำหน้าที่บันทึกข้อมูลภาพและเสียงจากการประชุม พร้อมทั้งถอดความเสียงเป็นข้อความแบบเรียลไทม์ จากนั้นประยุกต์ใช้เทคนิค RAG แบบการค้นหาเชิงความหมายร่วมกับโมเดลภาษาขนาดใหญ่ เพื่อสรุปสาระสำคัญของการประชุม

2.3) ระบบลงทะเบียนผู้ใช้งาน ทำหน้าที่เก็บรวบรวมข้อมูลอัตลักษณ์บุคคล ทั้งข้อมูลใบหน้าและเสียง เพื่อจัดเก็บเป็นฐานข้อมูลอ้างอิงสำหรับใช้ในการระบุตัวตนของผู้เข้าร่วมประชุมของระบบทั้งหมด

3) การออกแบบโครงสร้างการจัดเก็บข้อมูล

คณะผู้วิจัยได้เลือกใช้รูปแบบการจัดเก็บข้อมูลแบบผสมผสาน เพื่อเพิ่มประสิทธิภาพในการเข้าถึงข้อมูล ความยืดหยุ่นในการบริหารจัดการ และความเหมาะสมกับลักษณะข้อมูลที่หลากหลายภายในระบบ โดยโครงสร้างการจัดเก็บข้อมูลแบ่งออกเป็น 2 ส่วนหลัก ดังนี้

3.1) ฐานข้อมูลเชิงสัมพันธ์ ใช้ระบบจัดการฐานข้อมูล SQLite สำหรับจัดเก็บข้อมูลที่มีโครงสร้างชัดเจนและต้องการความรวดเร็วในการสืบค้น โดยประกอบด้วยฐานข้อมูลทะเบียนผู้ใช้งาน (User Registry) ซึ่งใช้จัดเก็บข้อมูลระบุตัวตนและข้อมูลการติดต่อของผู้ใช้งาน เช่น ชื่อที่แสดงและที่อยู่อีเมล นอกจากนี้ ยังมีฐานข้อมูลดัชนีการค้นหา (RAG Indexing) ซึ่งประยุกต์ใช้ตารางเสมือนแบบ FTS5 สำหรับการสร้างดัชนีข้อความเต็มจากบันทึกการประชุม เพื่อสนับสนุนการสืบค้นข้อมูลการสนทนาย้อนหลังตามบริบทได้อย่างมีประสิทธิภาพและรวดเร็ว

3.2) ระบบไฟล์ สำหรับจัดเก็บข้อมูลมัลติมีเดียและข้อมูลขนาดใหญ่ เพื่อลดภาระการประมวลผลของระบบฐานข้อมูล โดยข้อมูลทั้งหมดถูกจัดเก็บภายใต้โครงสร้างไดเรกทอรีที่กำหนดไว้อย่างเป็นระบบ ประกอบด้วยไฟล์วิดีโอบันทึกการประชุม (.mkv) ซึ่งบันทึกทั้งภาพและเสียงตลอดระยะเวลาการประชุม ไฟล์ข้อมูลใบหน้า (.json) ใช้เก็บเวกเตอร์คุณลักษณะใบหน้าสำหรับการเปรียบเทียบในกระบวนการรู้จำใบหน้า และไฟล์โมเดลเสียง (.onnx) สำหรับกระบวนการระบุและแยกผู้พูด (Speaker Diarization)

4.3.3 ระยะการสร้างและพัฒนาระบบ

คณะผู้วิจัยได้ดำเนินการพัฒนาซอฟต์แวร์และติดตั้งระบบลงบนอุปกรณ์ตามโครงสร้างการออกแบบสถาปัตยกรรมฮาร์ดแวร์ที่กำหนดไว้ โดยมุ่งเน้นการนำแบบจำลองและแนวคิดที่ได้จากการออกแบบทั้งด้านฮาร์ดแวร์และซอฟต์แวร์มาประยุกต์ใช้ในการพัฒนาให้เป็นระบบที่สามารถใช้งานได้จริง การพัฒนาระบบดำเนินการโดยใช้ภาษา Python สำหรับควบคุมและประสานการทำงานของอุปกรณ์ฮาร์ดแวร์อย่างเป็นระบบ ร่วมกับไมโครคอนโทรลเลอร์ ESP32 เพื่อรับสัญญาณเสียงและคำนวณตำแหน่งของแหล่งกำเนิดเสียง จากนั้นจึงนำข้อมูลดังกล่าวไปใช้ควบคุมการทำงานของเซอร์โวมอเตอร์ ส่งผลให้กล้องสามารถหมุนและติดตามตำแหน่งของผู้พูดได้โดยอัตโนมัติ

ในส่วนของการประมวลผลและการจัดเก็บข้อมูลได้พัฒนาด้วย PyQt5 ให้สามารถระบุตัวตนผู้ใช้งาน ด้วยการรู้จำใบหน้าที่เปรียบเทียบกับฐานข้อมูลอ้างอิงสำหรับการระบุชื่อผู้ใช้งาน ควบคู่กับโมดูลการถอดความเสียงเป็นข้อความ ซึ่งทำงานร่วมกับบริการ Azure Speech Service เพื่อแปลงสัญญาณเสียงการสนทนาเป็นข้อความแบบเรียลไทม์พร้อมระบุชื่อผู้พูด ข้อมูลที่ได้จากกระบวนการดังกล่าวจะถูกจัดเก็บตามโครงสร้างการจัดเก็บข้อมูลแบบที่ได้ออกแบบไว้ในขั้นตอน ระยะการออกแบบระบบ

นอกจากนี้ ในชั้นการประมวลผลปัญญาประดิษฐ์และการสังเคราะห์ข้อมูล ซึ่งเป็นส่วนสำคัญของระบบได้มีการพัฒนาโมดูลสำหรับจัดทำดัชนีข้อความจากบันทึกการประชุม และเชื่อมต่อกับโมเดลภาษาขนาดใหญ่ ผ่าน DeepSeek API โดยประยุกต์ใช้ RAG เพื่อสังเคราะห์และสรุปสาระสำคัญของการประชุมอย่างเป็นระบบ นอกจากนั้นการพัฒนาในระยะนี้ยังครอบคลุมถึงการทดสอบการทำงานของโมดูลย่อย การทดสอบการทำงานร่วมกันของระบบโดยรวม และการแก้ไขข้อผิดพลาดที่ตรวจพบ เพื่อให้ระบบมีความถูกต้อง เสถียร และพร้อมสำหรับการนำไปใช้งานจริงในสภาพแวดล้อมการประชุม [8]

4.3.4 ระยะการทดสอบและตรวจสอบ

ในระยะทดสอบ คณะผู้วิจัยได้ประเมินประสิทธิภาพและความถูกต้องของระบบตามระเบียบวิธีที่กำหนดเพื่อยืนยันว่าระบบทำงานสอดคล้องกับวัตถุประสงค์ที่ออกแบบไว้ การทดลองดำเนินการกับกลุ่มตัวอย่าง 10 คน โดยทดสอบในแต่ละทิศทางจำนวน 50 ครั้ง (รวม n = 200) เพื่อให้ได้ค่าสถิติที่มั่นคงและสะท้อนการใช้งานจริง การทดสอบครอบคลุมองค์ประกอบหลักของระบบ ได้แก่ การควบคุมกลไกและทิศทาง การระบุตัวตนผู้ใช้ การถอดเสียงเป็นข้อความ และการสรุปผลการประชุม โดยใช้ตัวชี้วัดที่เหมาะสมกับแต่ละโมดูลเพื่อประเมินความแม่นยำและคุณภาพของผลลัพธ์ และเพื่อพิสูจน์ผลกระทบของการประยุกต์ใช้โมเดลภาษาขนาดใหญ่และ RAG ต่อคุณภาพของ

บทสรุป คณะผู้วิจัยได้ออกแบบการทดลองเชิงเปรียบเทียบ โดยกำหนดวิธีการเปรียบเทียบ 3 รูปแบบ ดังนี้

1) Baseline 1: ไม่ใช้โมเดลภาษา (Extractive Summarization) คือ การสรุปแบบดึงข้อความสำคัญจากบทถอดความ โดยไม่ใช้โมเดลภาษา

2) Baseline 2: ใช้โมเดลภาษาอย่างเดียว (LLM only) ใช้โมเดลภาษา DeepSeek สร้างบทสรุปจากข้อความถอดความโดยตรง โดยไม่เชื่อมต่อฐานข้อมูลภายนอกหรือกระบวนการค้นคืนข้อมูลเพิ่มเติม

3) Baseline 3: ใช้โมเดลภาษาร่วมกับ RAG (Proposed Method) ใช้โมเดลภาษา DeepSeek ร่วมกับเทคนิค RAG โดยค้นคืนข้อมูลจากฐานข้อมูล FTS เพื่อใช้เป็นบริบทเพิ่มเติมในการสร้างบทสรุป

4.3.5 ระยะการประเมินผลการใช้งานระบบ

คณะผู้วิจัยได้กำหนดกรอบการประเมินผลการใช้งานระบบโดยอาศัยเกณฑ์ทางสถิติเชิงพรรณนา และตัวชี้วัดมาตรฐานที่สอดคล้องกับลักษณะข้อมูลของแต่ละโมดูล เพื่อวิเคราะห์ประสิทธิภาพและความถูกต้องของระบบในเชิงปริมาณ โดยการประเมินแบ่งออกเป็น 3 ด้านหลัก ดังนี้

1) การประเมินระบบฮาร์ดแวร์และการระบุตัวตนผู้ใช้งาน

การประเมินส่วนนี้ใช้ค่าความถูกต้อง โดยคำนวณจากสัดส่วนจำนวนครั้งที่ระบบสามารถควบคุมทิศทางกล้องตามแหล่งกำเนิดเสียงและระบุตัวตนผู้ใช้งานได้ถูกต้อง เทียบกับจำนวนการทดสอบทั้งหมด และรายงานผลในรูปแบบร้อยละ สำหรับการติดตามและระบุตัวตนผู้พูด ใช้อัตราความผิดพลาดในการแยกแยะผู้พูด ซึ่งคำนวณจากสัดส่วนเวลาความผิดพลาดเทียบกับเวลาทั้งหมด โดยอ้างอิงข้อมูลจริง (Ground Truth) จากการประชุมภาษาไทยที่มีบทสนทนาความยาว 4,071–4,327 อักษร พร้อมบันทึกเวลาเริ่มต้น เวลาสิ้นสุด และตัวตนผู้พูดอย่างละเอียด โดยพิจารณาจาก 3 องค์ประกอบ ได้แก่ 1) การไม่ตรวจจับเสียงพูดทั้งที่มีผู้พูด (Missed Speech) 2) การตรวจจับว่าเป็นเสียงพูดทั้งที่ไม่มีเสียงพูด (False Alarm) และ 3) การระบุผู้พูดผิดคน (Speaker Confusion) [10]

2) การประเมินประสิทธิภาพการถอดความเสียงเป็นข้อความ

การประเมินประสิทธิภาพของระบบถอดความเสียงใช้เกณฑ์อัตราคำผิดพลาด ซึ่งเป็นตัวชี้วัดเชิงสถิติที่ใช้วิเคราะห์ความคลาดเคลื่อนระหว่างผลลัพธ์ของระบบกับข้อมูลอ้างอิง โดยพิจารณาจากจำนวนการแทนที่คำ การลบคำ และการเติมคำ ผลการประเมินถูกสรุปด้วยค่าสถิติเชิงพรรณนา ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน เพื่อแสดงระดับความแม่นยำและความสม่ำเสมอของระบบในบริบทการใช้งานจริง

3) การประเมินคุณภาพการสรุปผลการประชุม

การประเมินคุณภาพบทสรุปใช้เกณฑ์จีอีแวล (G-Eval) ซึ่งเป็นการให้คะแนนเชิงปริมาณโดยโมเดลภาษาขนาดใหญ่ ภายใต้เกณฑ์ที่กำหนดชัดเจนร่วมกับการให้เหตุผลแบบลำดับขั้น (chain-of-thought) และการให้คะแนนแบบกรอกฟอร์ม (form-filling paradigm) เพื่อให้ผลลัพธ์สอดคล้องกับการประเมินของมนุษย์ ได้แก่ ความสอดคล้องของเนื้อหา ความเชื่อมโยงของข้อความ ความลื่นไหลทางภาษา และความเกี่ยวข้องกับต้นฉบับ [11] โดยนำผลคะแนนมาวิเคราะห์ด้วยสถิติเชิงพรรณนา เช่น ค่าเฉลี่ยและช่วงคะแนน เพื่อสะท้อนคุณภาพโดยรวมของบทสรุป

นอกจากการประเมินคุณภาพของระบบที่พัฒนาขึ้น คณะผู้วิจัยยังการวิเคราะห์เชิงเปรียบเทียบระหว่างวิธีการสรุปผลแบบต่าง ๆ ผ่านกระบวนการทดลองแบบตัดองค์ประกอบ (Ablation Study) เพื่อศึกษาผลกระทบของแต่ละองค์ประกอบ โดยเฉพาะบทบาทของโมเดลภาษาขนาดใหญ่และเทคนิค RAG อย่างเป็นระบบ

5. ผลการวิจัย

จากการดำเนินการวิจัยและพัฒนาระบบวิเคราะห์การประชุมสำหรับการติดตามผู้พูดและการสรุปผลด้วยโมเดลภาษาขนาดใหญ่ มีผลการวิจัยดังนี้

5.1 ผลการออกแบบและพัฒนาระบบต้นแบบ

คณะผู้วิจัยได้ดำเนินการสร้างและประกอบเครื่องต้นแบบตามสถาปัตยกรรมที่ได้ออกแบบไว้ โดยแบ่งส่วนประกอบตามหน้าที่การทำงานออกเป็น 2 ส่วนหลัก ได้แก่

5.1.1 โครงสร้างและส่วนประกอบของระบบฮาร์ดแวร์

ผลจากการศึกษาและออกแบบถูกพัฒนาเป็นเครื่องบันทึกการประชุมสำหรับติดตามผู้พูด โดยประกอบด้วยองค์ประกอบหลัก ได้แก่ หน้าจอแสดงผลของระบบ ปุ่มควบคุมการทำงาน ชุดควบคุมทิศทางที่ทำงานร่วมกับไมโครโฟนทิศทาง 4 ทิศทาง ส่วนแสดงระดับสัญญาณเสียง ตลอดจนลำโพงและไมโครโฟนสำหรับ รับเสียงสนทนาหลักของการประชุม รายละเอียดโครงสร้างและส่วนประกอบฮาร์ดแวร์แสดงดังรูปที่ 3



รูปที่ 3 โครงสร้างและส่วนประกอบของระบบฮาร์ดแวร์เครื่องบันทึกการประชุมสำหรับการติดตามผู้พูด

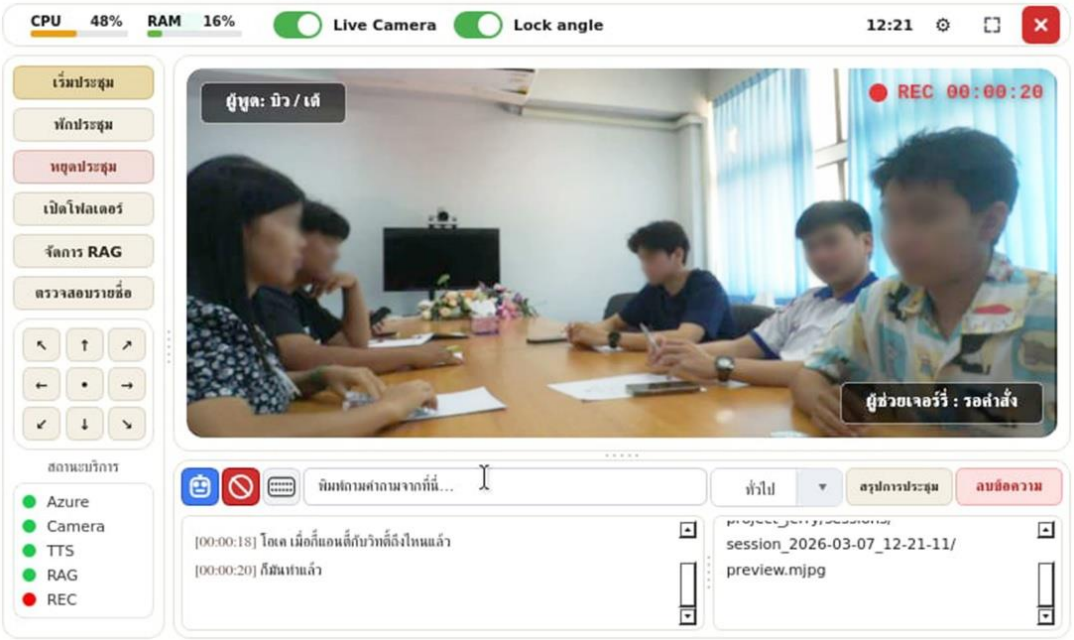
5.1.2 ด้านซอฟต์แวร์และส่วนต่อประสานกราฟิกผู้ใช้

การพัฒนาซอฟต์แวร์ควบคุมระบบวิเคราะห์การประชุมสำหรับการติดตามผู้พูดและการสรุปผลด้วยโมเดลภาษาขนาดใหญ่ ดำเนินการพัฒนาด้วยภาษา Python โดยออกแบบส่วนต่อประสานกราฟิกผู้ใช้ด้วย PyQt5 ซึ่งทำงานบนบอร์ด Raspberry Pi 5 โดยซอฟต์แวร์ของระบบถูกออกแบบในลักษณะเชิงโมดูล เพื่อให้สามารถบริหารจัดการการทำงานของระบบได้อย่างเป็นระบบและมีประสิทธิภาพ สามารถแบ่งการทำงานออกเป็น 3 ส่วน ดังนี้

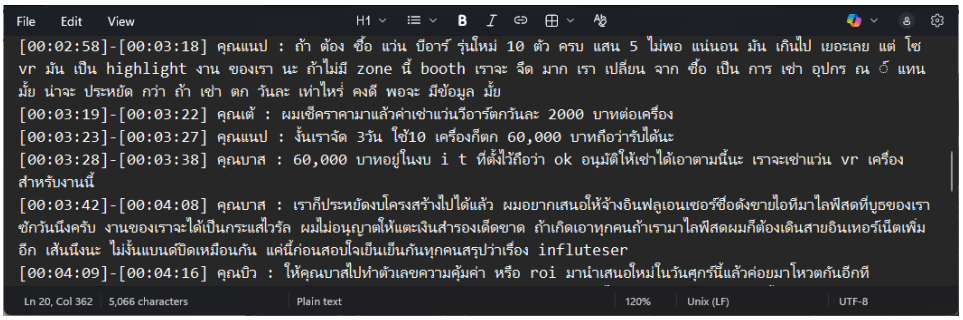
1) ส่วนควบคุมการทำงานฮาร์ดแวร์ ทำหน้าที่สื่อสารระหว่างซอฟต์แวร์หลักกับบอร์ด ESP32 ผ่านการสื่อสารแบบอนุกรม เพื่อรับข้อมูลทิศทางเสียง แล้วส่งต่อไปยังโมดูลคำนวณพิกัดเพื่อแปลงเป็นคำสั่งควบคุมเซอร์โวมอเตอร์ 2 ตัว ให้กล้องหมุนติดตามผู้พูดแบบครอบคลุม 360 องศา โดยเซอร์โวตัวที่ 1 ควบคุมช่วง 0–180 องศา และตัวที่ 2 ควบคุมช่วง 181–360 องศา เพื่อลดเวลาการหมุนและเพิ่มความแม่นยำ

2) ส่วนประมวลผลปัญญาประดิษฐ์ ทำหน้าที่จัดการข้อมูลเสียงและข้อความจากการประชุม เริ่มจากการถอดความเสียงผ่าน Azure Speech Service ร่วมกับโมดูลระบุตัวตนผู้พูด จากนั้นใช้โมเดลภาษาขนาดใหญ่ DeepSeek และฐานข้อมูลเชิงสัมพันธ์ โดยประยุกต์ใช้ RAG ในการสร้างดัชนีและค้นคืนข้อมูลประกอบการประชุมเพื่อใช้เป็นบริบทในการสรุปผลอย่างเป็นระบบ

3) ส่วนต่อประสานกราฟิกผู้ใช้ ออกแบบให้รองรับการบริหารจัดการการประชุมผ่านหน้าจอหลัก โดยประกอบด้วยส่วนแสดงสถานะระบบ เช่น หน่วยประมวลผลกลาง และหน่วยความจำ พร้อมปุ่มควบคุมการเปิด–ปิดกล้อง และการหมุนกล้องแบบอัตโนมัติหรือควบคุมด้วยตนเอง รวมถึงนาฬิกาและเมนูตั้งค่าระบบ ส่วนควบคุมการประชุมการเริ่ม-หยุดบันทึกการประชุม รวมถึงส่วนแสดงผลวิดีโอถ่ายทอดสดจากกล้อง พร้อมแสดงสถานะผู้ช่วยระบบและชื่อผู้พูดที่ตรวจจับได้ ดังรูปที่ 4 และผลลัพธ์การถอดเสียงพูดเป็นข้อความจะถูกจัดเก็บเป็นไฟล์ โดยมีรายละเอียดข้อมูลการบันทึกการประชุมที่ได้ ดังแสดงในรูปที่ 5

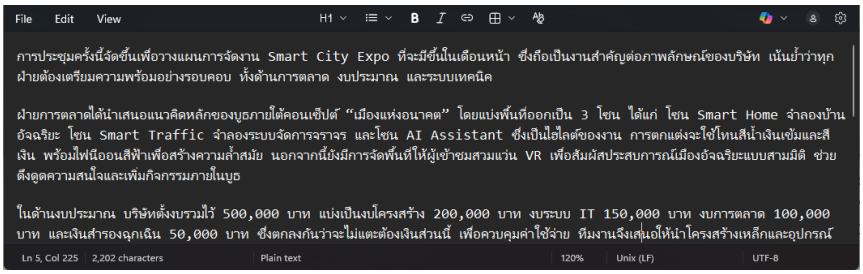


รูปที่ 4 ส่วนต่อประสานผู้ใช้และการตั้งค่าระบบ



รูปที่ 5 หน้าจอแสดงข้อมูลจากถอดความเสียงพูด

สำหรับกระบวนการสรุปผลการประชุม ระบบจะดึงข้อมูลข้อความฉบับสมบูรณ์ของการประชุมจากไฟล์ข้อมูล ซึ่งจัดเก็บอยู่ภายในโฟลเดอร์ของการประชุมแต่ละรายการ จากนั้นข้อมูลดังกล่าวจะถูกส่งไปยังโมเดลภาษาขนาดใหญ่ DeepSeek พร้อมคำสั่งกำหนดรูปแบบการสรุปผลที่ระบบต้นทางกำหนดไว้ในรูปแบบของพร็อมท์คำสั่ง เมื่อโมเดลภาษาประมวลผลเสร็จสิ้นและได้ผลลัพธ์การสรุป และจัดเก็บในรูปแบบไฟล์ข้อความ ซึ่งมีข้อความดังแสดงในรูปที่ 6



รูปที่ 6 รายละเอียดในไฟล์สรุปผลการประชุม

5.2 ผลการทดสอบประสิทธิภาพของระบบ

คณะผู้วิจัยได้ดำเนินการประเมินประสิทธิภาพของระบบตามระเบียบวิธีวิจัยที่กำหนด เพื่อตรวจสอบความถูกต้องและความแม่นยำในการทำงานของระบบต้นแบบ โดยแบ่งการทดสอบออกเป็น 3 ส่วนหลัก ซึ่งมีรายละเอียดผลการทดสอบดังนี้

5.2.1 ผลการทดสอบความถูกต้องของการระบุทิศทางเสียงและการระบุตัวตนผู้ใช้งาน

การประเมินประสิทธิภาพในส่วนนี้ดำเนินการภายในห้องเงียบที่มีระดับเสียงรบกวนต่ำ โดยกำหนดให้ผู้พูดล้อมรอบอุปกรณ์ในรัศมีไม่เกิน 2 เมตร เพื่อจำลองสภาพแวดล้อมการประชุมจริง คณะผู้วิจัยได้กำหนดจุดทดสอบรอบตัวเครื่องใน 4 ทิศทางหลัก ได้แก่ ด้านหน้า ด้านขวา ด้านซ้าย และด้านหลัง โดยทำการทดสอบเรียกชื่อระบบทิศทางละ 50 ครั้ง รวมจำนวนการทดสอบทั้งสิ้น 200 ครั้ง ผลการทดสอบถูกนำเสนอในรูปแบบตารางคอนฟิวชันเมตริก (Confusion Matrix) ดังแสดงในตารางที่ 1

จากผลในตารางที่ 1 คณะผู้วิจัยได้คำนวณค่าความถูกต้องเชิงสถิติ ประกอบด้วย ผลบวกจริง (True Positive: TP) ผลลบเท็จ (False Negative: FN) ผลบวกเท็จ (False Positive: FP) และผลลบจริง (True Negative: TN) แยกตามแต่ละทิศทาง จากนั้นนำมาคำนวณหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าคะแนน F1 (F1-score) ด้วยสมการที่ 3 - 6 [12] ได้ผลดังแสดงในตารางที่ 2

ตารางที่ 1 ผลการทดสอบความถูกต้องของการระบุทิศทางเสียง (n = 200)

		ทิศที่กล้องหัน			
		ด้านหน้า	ด้านขวา	ด้านซ้าย	ด้านหลัง
ทิศที่เรียกจริง (ทิศละ 50 ครั้ง)	ด้านหน้า	50	0	0	0
	ด้านขวา	1	48	0	1
	ด้านซ้าย	0	0	49	1
	ด้านหลัง	0	0	0	50

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

ตารางที่ 2 ค่าความถูกต้องเชิงสถิติของการระบุทิศทางเสียง

ทิศทาง	TP	FN	FP	TN	Accuracy	Precision	Recall	F1-score
ด้านหน้า	50	0	1	149	99.50%	98.00%	100.00%	99.00%
ด้านขวา	48	2	0	150	99.00%	100.00%	96.00%	98.00%
ด้านซ้าย	49	1	0	150	99.50%	100.00%	98.00%	99.00%
ด้านหลัง	50	0	2	148	99.00%	96.20%	100.00%	98.00%

จากตารางที่ 2 พบว่า ผลรวมจาก Confusion Matrix โดยทุกทิศทางมีค่า Accuracy มากกว่า 99% แสดงให้เห็นว่าระบบสามารถควบคุมการหมุนกล้องให้ตรงกับแหล่งกำเนิดเสียงได้อย่างแม่นยำ และเมื่อพิจารณาเชิงรายทิศทาง พบว่า ทิศด้านหน้าและด้านซ้ายมีค่า F1-score สูงถึง 99.00% ขณะที่ทิศด้านขวาและด้านหลังมีค่า F1-score เท่ากับ 98.00% ซึ่งสะท้อนถึงความสมดุลระหว่างค่า Precision และ Recall ในขณะที่ค่า Recall ของทุกทิศทางมีค่ามากกว่า 96% แสดงถึงความสามารถของระบบในการติดตามผู้พูดโดยอาศัยการประมวลผลทิศทางเสียงได้ครบถ้วน

การประเมินประสิทธิภาพของระบบติดตามและระบุตัวตนผู้พูดดำเนินการกับกลุ่มตัวอย่างจำนวน 10 คน ซึ่งผ่านกระบวนการลงทะเบียนข้อมูลใบหน้าและเสียงล่วงหน้า ระบบถูกทดสอบภายใต้สถานการณ์การประชุมจำลองจำนวน 10 เซสชัน (sessions) ในสภาพแวดล้อมห้องประชุมขนาดเล็กถึงขนาดกลาง และมีผู้ร่วมสนทนา 7–10 คนต่อครั้ง ระยะเวลาการประชุมเฉลี่ยอยู่ที่ 325.40 วินาทีต่อเซสชัน

ตารางที่ 3 ผลการประสิทธิภาพของระบบติดตามและระบุตัวตนผู้พูด

เซสชัน	เวลาประชุม (วินาที)	ไม่ตรวจจับ (วินาที)	ตรวจจับผิด (วินาที)	ระบุผู้พูดผิด (วินาที)	รวม (วินาที)	DER (%)
1	326	3	6	43	52	15.95
2	318	4	5	28	37	11.64
3	340	2	7	44	53	15.59
4	305	3	4	36	43	14.10
5	332	5	6	38	49	14.76
6	315	2	5	32	39	12.38
7	350	4	8	46	58	16.57
8	298	3	4	32	39	13.09
9	360	6	7	51	64	17.78
10	310	3	5	34	42	13.55
เฉลี่ย	**325.40**	**3.50**	**5.70**	**38.40**	**47.6**	**14.54**

จากตารางที่ 3 แสดงผลการทดลองทั้ง 10 เซสชัน พบว่า ระบบติดตามและระบุตัวตนผู้พูดมีค่า อัตราความผิดพลาดในการแยกแยะผู้พูดมีค่าความคลาดเคลื่อนอยู่ในช่วงร้อยละ 11.64 – 17.78 คิดเป็นค่าเฉลี่ยร้อยละ 14.54 ± 1.85 โดยสะท้อนว่าประสิทธิภาพของระบบมีความสม่ำเสมอในแต่ละรอบการทดสอบ เมื่อพิจารณาองค์ประกอบของความผิดพลาด พบว่า ความคลาดเคลื่อนส่วนใหญ่เกิดจากการระบุผู้พูดผิด ซึ่งมีค่าเฉลี่ย 38.40 วินาทีต่อเซสชัน ขณะที่ความผิดพลาดจากการไม่ตรวจจับเสียงและการตรวจจับผิดมีสัดส่วนค่อนข้างต่ำ (เฉลี่ย 3.50 และ 5.70 วินาที ตามลำดับ) บ่งชี้ว่าระบบสามารถตรวจจับและติดตามผู้พูดได้ต่อเนื่องในสภาพแวดล้อมการประชุมจำลอง แต่ยังมีข้อจำกัดด้านความแม่นยำในการระบุตัวบุคคล ซึ่งส่งผลต่อค่าอัตราความผิดพลาดในการแยกแยะผู้พูดโดยรวม

5.2.2 ผลการทดสอบประสิทธิภาพการประมวลผลภาษาธรรมชาติ

การประเมินความถูกต้องของการถอดความเสียง คณะผู้วิจัยใช้เกณฑ์อัตราคำผิดพลาด โดยดำเนินการทดสอบกับกลุ่มตัวอย่าง 10 คน ให้ทดสอบการพูดชุดประโยคมาตรฐานเฉลี่ยคนละประมาณ 150 คำ รวมจำนวนคำทั้งสิ้น 1,528 คำ ครอบคลุมคำศัพท์ทั่วไปและศัพท์เทคนิค การวัดผลใช้วิธีนับจำนวนคำที่ผิดพลาด ได้แก่ การแทนที่ การตกหล่น และการเติมคำ เทียบกับจำนวนคำทั้งหมดในประโยคต้นฉบับ

ตารางที่ 4 ผลการวิเคราะห์อัตราคำผิดพลาดของการถอดความเสียง

ประเภทประโยค	คำทั้งหมด (คำ)	การแทนที่ (คำ)	การตกหล่น (คำ)	การเติมคำ (คำ)	รวมคำผิด (คำ)	อัตราคำผิดพลาด (ร้อยละ)
สนทนาทั่วไป	450	19	2	4	25	5.56
ศัพท์เทคนิค	474	62	8	5	75	15.82
ซับซ้อน/พูดเร็ว	604	63	42	3	108	17.88
รวม	1,528	144	52	12	208	13.61

จากตารางที่ 4 ระบบมีค่าอัตราคำผิดพลาดเฉลี่ยเท่ากับร้อยละ 13.61 โดยประโยคสนทนาทั่วไปมีอัตราคำผิดพลาดต่ำที่สุด คิดเป็นร้อยละ 5.56 แสดงให้เห็นถึงความสามารถของระบบในการถอดความบทสนทนาทั่วไปได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม ประโยคที่มีความซับซ้อนหรือมีการพูดด้วยความเร็วสูงมี อัตราคำผิดพลาดสูงที่สุด คิดเป็นร้อยละ 17.88 รองลงมาคือประโยคที่มีศัพท์เทคนิค ซึ่งมีอัตราคำผิดพลาดเท่ากับร้อยละ 15.82 สะท้อนให้เห็นว่าความซับซ้อนของภาษาและลักษณะเฉพาะของคำศัพท์ส่งผลต่อประสิทธิภาพการถอดความของระบบ

5.2.3 ผลการประเมินคุณภาพการสรุปผลการประชุม

การประเมินคุณภาพของบทสรุปการประชุมในงานวิจัยนี้ใช้เกณฑ์การประเมินจีอีแวล ซึ่งเป็นกรอบแนวคิดสำหรับประเมินคุณภาพของข้อความที่สร้างโดยแบบจำลองภาษาขนาดใหญ่ โดยอาศัยโมเดลภาษา ChatGPT ในการทำหน้าที่เป็นผู้ตัดสินคุณภาพของผลงาน [11] กระบวนการประเมินประกอบด้วยการออกแบบขั้นตอนการพิจารณาอย่างเป็นลำดับ โดยใช้เทคนิคการกระตุ้นการคิดของปัญญาประดิษฐ์ในลักษณะ กระบวนการให้เหตุผลแบบลำดับขั้น เพื่อให้การพิจารณามีความเป็นระบบและมีเหตุผลรองรับในแต่ละมิติ จากนั้นให้คะแนนในช่วง 1–5 ตามมาตราส่วนลิเคิร์ท [13] โดยกำหนดกำหนดเกณฑ์การแปลผลค่าคะแนนเฉลี่ยในระดับต่าง ๆ ดังนี้

1) ระดับดีที่สุด คะแนนเฉลี่ย 4.51 - 5.00
2) ระดับดีมาก คะแนนเฉลี่ย 3.51 - 4.50
3) ระดับปานกลาง คะแนนเฉลี่ย 2.51 - 3.50

4) ระดับน้อย คะแนนเฉลี่ย 1.51 - 2.50

5) ระดับน้อยที่สุด คะแนนเฉลี่ย 1.00 - 1.50

ตารางที่ 5 ผลการประเมินคุณภาพบทสรุปการประชุมด้วยเกณฑ์จีอีแวล

เกณฑ์การประเมิน	คะแนนเฉลี่ย	S.D.	ระดับคุณภาพ
ความสอดคล้องของเนื้อหา	4.60	0.49	ดีที่สุด
ความเชื่อมโยงของประโยค	4.50	0.50	ดีมาก
ความถูกต้องของไวยากรณ์	4.70	0.46	ดีที่สุด
ความเกี่ยวข้องกับต้นฉบับ	4.40	0.66	ดีมาก
เฉลี่ย	4.55	0.53	ดีที่สุด

จากตารางที่ 5 ผลการประเมินคุณภาพบทสรุปการประชุมด้วยเกณฑ์จีอีแวล แสดงให้เห็นว่าบทสรุปที่ได้มีคุณภาพโดยรวมอยู่ในระดับดีที่สุด (4.55±0.53) เมื่อพิจารณาเป็นรายด้าน พบว่า ด้านความถูกต้องของไวยากรณ์มีระดับคะแนนสูงที่สุด อยู่ในระดับดีที่สุด (4.70±0.46) รองลงมา คือ ด้านความสอดคล้องของเนื้อหา อยู่ในระดับดีที่สุด (4.60±0.49) ด้านความเชื่อมโยงของประโยค อยู่ในระดับดีมาก (4.50±0.50) และด้านความเกี่ยวข้องกับต้นฉบับอยู่ในระดับดีมาก (4.40±0.66) ตามลำดับ สะท้อนถึงประสิทธิภาพของกระบวนการสรุปผลการประชุมที่พัฒนาขึ้นในงานวิจัยนี้สามารถนำไปใช้ประโยชน์ได้จริง

เพื่อพิสูจน์ผลกระทบของการประยุกต์ใช้โมเดลภาษาขนาดใหญ่และเทคนิค RAG ต่อคุณภาพของบทสรุปการประชุม คณะผู้วิจัยได้ออกแบบการทดลองเชิงเปรียบเทียบ โดยกำหนดรูปแบบการสรุปผล 3 วิธี ภายใต้ข้อมูลการประชุมชุดเดียวกัน เพื่อควบคุมตัวแปรด้านข้อมูลนำเข้าให้เหมือนกันทุกประการ และประเมินด้วยเกณฑ์จีอีแวลเดียวกัน ทดสอบ 5 ครั้ง แต่ละครั้งมีจำนวนผู้ประชุม 6-10 คน ได้คะแนนเฉลี่ยดังตารางที่ 6

ตารางที่ 6 ผลการเปรียบเทียบโมเดลภาษาขนาดใหญ่และเทคนิค RAG ต่อคุณภาพของบทสรุปการประชุม

วิธีการ	ความสอดคล้องของเนื้อหา	ความเชื่อมโยงของประโยค	ความถูกต้องของไวยากรณ์	ความเกี่ยวข้องกับต้นฉบับ	คะแนนเฉลี่ย	ระดับคุณภาพ
ไม่ใช้โมเดลภาษา	3.18	3.05	3.82	3.12	3.29	ปานกลาง
ใช้โมเดลภาษาอย่างเดียว	4.08	4.02	4.47	3.87	4.11	ดีมาก
ใช้โมเดลภาษาและ RAG	4.53	4.46	4.61	4.43	4.53	ดีที่สุด

จากตารางที่ 6 พบว่า การใช้โมเดลภาษาขนาดใหญ่สามารถยกระดับคุณภาพบทสรุปได้อย่างมีนัยสำคัญเมื่อเทียบกับวิธีการแบบดั้งเดิมที่ไม่ใช้โมเดลภาษา และเมื่อผสานเทคนิค RAG เข้ากับโมเดลภาษาจะช่วยเสริมความแม่นยำเชิงบริบทและความสอดคล้องกับข้อมูลต้นฉบับได้ดียิ่งขึ้น ส่งผลให้คุณภาพโดยรวมของบทสรุปอยู่ในระดับสูงสุด สะท้อนถึงความสำคัญของการบูรณาการกลไกการสืบค้นข้อมูลกับกระบวนการสร้างข้อความในการพัฒนาระบบสรุปผลการประชุมเชิงอัตโนมัติอย่างมีประสิทธิภาพ

6. สรุปและอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อออกแบบและพัฒนาระบบวิเคราะห์การประชุมสำหรับการติดตามผู้พูดในลักษณะการระบุทิศทางเสียงเพื่อควบคุมการหันกล้องอัตโนมัติ และการสรุปผลการประชุมด้วยโมเดลภาษาขนาดใหญ่ พร้อมทั้งประเมินประสิทธิภาพของระบบในด้านการระบุทิศทางเสียง การถอดความเสียง และคุณภาพของบทสรุป โดยการทดสอบดำเนินการในห้องประชุมจำลองขนาดประมาณ 20–30 ตารางเมตร ระดับเสียงพื้นหลังไม่เกิน 35 dBA มีผู้เข้าร่วมไม่เกิน 10 คน และอยู่ห่างจากอุปกรณ์ประมาณ 1–2 เมตร จำนวนการทดสอบทิศทางเสียงรวม 200 ครั้ง ภายใต้เงื่อนไขดังกล่าว ระบบแสดงประสิทธิภาพที่ดีและมีเสถียรภาพ โดยสามารถสรุปผลได้ดังนี้

1) ผลการออกแบบและพัฒนาระบบต้นแบบพบว่า ระบบสามารถบูรณาการฮาร์ดแวร์และซอฟต์แวร์ได้อย่างมีประสิทธิภาพ ใช้ไมโครโฟนทิศทางร่วมกับกลไกหมุนกล้องแบบ 360 องศา ซอฟต์แวร์พัฒนาเชิงโมดูลบน Raspberry Pi 5 รองรับกระบวนการประชุมครบถ้วน ตั้งแต่บันทึกเสียง ถอดความ ระบุตัวตนผู้พูด จนถึงสรุปผลผ่าน ส่วนต่อประสานกราฟิกผู้ใช้ ซึ่งเอื้อต่อการใช้งานจริงในสภาพแวดล้อมการประชุม ในด้านการทดสอบระบบติดตามและระบุตัวตนผู้พูด พบค่าเฉลี่ยของอัตราความผิดพลาดในการแยกแยะผู้พูด คิดเป็นร้อยละ 14.54 ± 1.85 แสดงถึงความสม่ำเสมอของระบบ ทั้งนี้ ความคลาดเคลื่อนบางส่วนอาจเกิดจากข้อจำกัดด้านประสิทธิภาพการประมวลผลของ Raspberry Pi 5 และข้อจำกัดเชิงกลของเซอร์โวมอเตอร์ ซึ่งอาจตอบสนองไม่ทันต่อการเปลี่ยนตำแหน่งผู้พูดอย่างรวดเร็ว

2) ผลการประเมินประสิทธิภาพของระบบ แสดงให้เห็นว่าระบบมีความสามารถในการระบุทิศทางเสียงในระดับสูง โดยมีค่าความถูกต้องเฉลี่ยทั้งสี่ทิศทาง คิดเป็นร้อยละ 98.50 สะท้อนถึงความแม่นยำของกลไกการตรวจจับและการควบคุมทิศทางกล้อง นอกจากนี้ การประเมินประสิทธิภาพการถอดความเสียงโดยใช้อัตราคำผิดพลาด ยังพบว่าระบบสามารถถอดความบทสนทนาทั่วไปได้อย่างมีประสิทธิภาพสูง

ในขณะที่ประโยคที่มีความซับซ้อนหรือมีศัพท์เทคนิคมีอัตราความผิดพลาดสูงขึ้น เนื่องจากข้อจำกัดของโมเดลรู้จำเสียงที่อาจไม่มีคลังคำศัพท์เฉพาะทาง หรือเกิดการซ้อนทับของหน่วยเสียง เมื่อมีการพูดด้วยความเร็วสูง ทำให้การถอดรหัสเสียงคลาดเคลื่อน ทำให้ระบบ RAG อาจค้นคืนข้อมูลบริบทที่ไม่เกี่ยวข้องมาให้โมเดลภาษา และทำให้โมเดลภาษาสังเคราะห์บทสรุปที่คลาดเคลื่อนจากข้อเท็จจริง แนวทางในการลดผลกระทบนี้ที่ได้ผลจริงคือการใช้พจนานุกรมที่กำหนดเอง (Custom Lexicon) เพื่อระบุศัพท์เฉพาะทางให้กับระบบและการปรับแต่งโมเดลภาษาหรือใช้โมเดลภาษาเพื่อตรวจแก้คำผิดเชิงบริบทก่อนนำไปสรุปผล [14]

3) ผลการประเมินด้านคุณภาพของการสรุปการประชุมด้วยเกณฑ์จีอีแวลแสดงให้ พบว่า บทสรุปที่ระบบสร้างขึ้นมีคุณภาพโดยรวมอยู่ในระดับดีที่สุด โดยมีจุดเด่นด้านความถูกต้องของไวยากรณ์ ความสอดคล้องของเนื้อหา และความเชื่อมโยงของข้อความ แม้ข้อมูลนำเข้าจะมีข้อจำกัดจากความคลาดเคลื่อนในการถอดความเสียงบางส่วน ผลการวิจัยดังกล่าวสะท้อนให้เห็นถึงศักยภาพของการประยุกต์ใช้โมเดลภาษาขนาดใหญ่ร่วมกับระบบวิเคราะห์การประชุม เพื่อสนับสนุนการจัดการความรู้และการสรุปสาระสำคัญของการประชุมได้อย่างมีประสิทธิภาพ และสามารถนำไปต่อยอดใช้งานจริงในบริบทของการประชุมอัจฉริยะในอนาคต

เมื่อพิจารณาผลการทดลองเชิงเปรียบเทียบทั้ง 3 วิธี ภายใต้ข้อมูลการประชุมชุดเดียวกัน พบความแตกต่างอย่างชัดเจน กล่าวคือ วิธีที่ไม่ใช้โมเดลภาษาได้คะแนนเฉลี่ยเพียง 3.29 (ระดับปานกลาง) การใช้โมเดลภาษาอย่างเดียวเพิ่มคะแนนเป็น 4.11 (ระดับดีมาก) และเมื่อผสานโมเดลภาษาขนาดใหญ่เข้ากับเทคนิค RAG คะแนนเฉลี่ยเพิ่มขึ้นเป็น 4.53 (ระดับดีที่สุด) แสดงให้เห็นแนวโน้มเชิงบวกอย่างเป็นลำดับขั้น ผลดังกล่าวสอดคล้องกับงานวิจัยของ Medeiros และ Oliveira [15] ที่ระบุว่า แม้โมเดลภาษาจะมีประสิทธิภาพในการประมวลผลและสร้างข้อความจากข้อมูลเสียง แต่ยังอาจเผชิญข้อจำกัดด้านความรู้ที่ล้าสมัยหรือการสร้างข้อมูลคลาดเคลื่อน (Hallucination) ได้ การผสานเทคนิค

RAG ซึ่งอาศัยกลไกการดึงข้อมูลจากแหล่งภายนอกเข้ามาประกอบการสร้างข้อความ จึงช่วยเพิ่มความแม่นยำ ความทันสมัย และความสอดคล้องของเนื้อหา ทำให้บทสรุปมีคุณภาพและเหมาะสมต่อการใช้งานจริงมากยิ่งขึ้น

7. ข้อเสนอแนะ

จากการดำเนินงานวิจัยและผลการทดลองที่พบ คณะผู้วิจัยมีข้อเสนอแนะเพื่อการพัฒนาต่อยอดระบบให้มีประสิทธิภาพและความรัดกุมยิ่งขึ้นในอนาคต ดังนี้

1) การลดปัญหาคำนอกคลังคำศัพท์ผ่านการใช้พจนานุกรมเฉพาะทางตามโดเมนของการประชุม รวมถึงการปรับแต่งโมเดลให้มีความสอดคล้องกับบริบทเนื้อหา นอกจากนี้ การเพิ่มขั้นตอนการตรวจแก้ข้อความอัตโนมัติภายหลังการถอดความเสียงก่อนเข้าสู่กระบวนการ RAG อาจช่วยยกระดับความถูกต้องของข้อมูลนำเข้าและลดผลกระทบเชิงลูกโซ่ต่อคุณภาพของบทสรุป

2) การใช้งานในสภาพแวดล้อมจริงยังเผชิญข้อจำกัดจากการบดบังใบหน้าและมุมมองของกล้อง การเพิ่มประสิทธิภาพการติดตามและระบุตัวตนผู้พูดในสภาพแวดล้อมที่ซับซ้อน จึงเป็นแนวทางที่มีศักยภาพในการเพิ่มความแม่นยำในการระบุตัวตนผู้พูด และช่วยเสริมการวิเคราะห์การประชุมในสถานการณ์ที่มีความซับซ้อนสูง

3) การขยายการทดสอบไปยังสถานการณ์ที่มีความท้าทายมากขึ้น เช่น สภาพแวดล้อมที่มีเสียงก้องหรือเสียงรบกวนสูง รวมถึงจำนวนผู้เข้าร่วมที่เพิ่มขึ้น เพื่อประเมินความทนทานและความสามารถในการปรับตัวของระบบ ทั้งนี้ควรพิจารณาควบคู่กับประเด็นด้านความเป็นส่วนตัวและความปลอดภัยของข้อมูลอย่างรอบคอบ

8. เอกสารอ้างอิง

[1] P. Kantithammakorn et al., “Using automatic speech recognition to assess Thai speech language fluency in Montreal Cognitive Assessment (MoCA),” *BMC Medical Informatics and Decision Making*, vol. 22, pp. 1–14, 2022.

[2] X. Yang, D. Ji, and Z. Shu, “Design Mechanism of English Translation Calibration System Based on Artificial Intelligence Technology,” in *Proceedings of the 2023 IEEE 3rd International Conference on Social Sciences and Intelligence Management (SSIM),* 2023, pp. 99–103.

[3] Y. Chen, H. Wang, K. Yu, and R. Zhou, “Artificial intelligence methods in natural language processing: A comprehensive review,” *Highlights in Science, Engineering and Technology*, vol. 85, pp. 545–550, 2024.

[4] A. Hassan et al., “Performance evaluation of Raspberry Pi as an IoT edge signal processing device for a real-time flash flood forecasting system,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 841–844, 2022.

[5] J. Klump, J. Hille, M. Guglielmo, and B. Gardner, “Building a RAG system for querying a large corpus of conference abstracts,” in EGU General Assembly Conference Abstracts, Apr. 2025, pp. EGU25-7690.

[6] พิศณุ คูมีชัย, “เทคโนโลยีการรู้จำใบหน้ามนุษย์ด้วยการประมวลผลภาพ,” *วารสารวิชาการโรงเรียนนายเรือด้านวิทยาศาสตร์และเทคโนโลยี*, vol. 5, no. 1, pp. 136-148, 2565.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Boston, MA, USA, 2015, pp. 815–823.

[8] P. Lewis, E. Perez, A. Piktus, A., F. Petroni, V. Karpukhin, V., N. Goyal, and Kiela, D. “Retrieval-augmented generation for knowledge-intensive nlp tasks,” in Proc. 34th Int. Conf. Neural Inf. Process. Syst. (NeurIPS), Vancouver, BC, Canada, 2020, pp. 9459–9474.

[9] T. Thesing, C. Feldmann, and M. Burchardt, M. “Agile versus Waterfall project management: Decision model for selecting the appropriate approach to a project,” *Procedia Computer Science*, vol. 181, pp. 746–756, 2021.

[10] Q. Wang, Y. Huang, G. Zhao, E. Clark, W. Xia, and H. Liao, “Diarizationlm: Speaker diarization post-processing with large language models,” *arXiv preprint arXiv:2401.03506*, 2024.

[11] Y. Liu, G. Dan, C.-H. Jia, B. G. Lewis, W.-t. Yih, and P. Yu, “G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment,” in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore, Dec. 2023, pp. 2511–2522.

[12] อิทธิศักดิ์ ศรีดำ. “ระบบปัญญาประดิษฐ์สำหรับการจัดหมวดหมู่ข้อคิดเห็นและข้อเสนอแนะของประชาชนที่มีต่อโครงการของรัฐ,” *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, vol. 41, no. 3, pp. 134-141, 2565.

[13] ภูมินันท์ บัวงาม, “ระบบแก้ไขปัญหาไอพีทีวีอัตโนมัติโดยใช้วิธีการให้เหตุผลด้วยฐานกรณี,” *วารสารวิทยาศาสตร์และเทคโนโลยีนายเรืออากาศ*, vol. 15, no. 1, pp. 99-111, 2562.

[14] D. Eitan, M. Pirchi, N. Glazer, S. Meital, G. Ayach, G. Krendel, and J. Keshet, “Combining language models for specialized domains: A colorful approach,” *arXiv preprint arXiv:2310.19708*, 2023.

[15] L. S. F. Medeiros and H. Oliveira, “Comparação de Modelos de Embeddings e LLMs para Geração Aumentada por Recuperação em Português,” in *Proceedings of SEMISH 2025*, pp. 429–440, 2025. DOI: 10.5753/semish.2025.9027.