

ผู้ช่วยฝึกการอ่านออกเสียงภาษาอังกฤษโดยใช้ปัญญาประดิษฐ์ผ่านเทคโนโลยี

NVIDIA Riva และ pygame

AN AI-POWERED ENGLISH PRONUNCIATION TRAINING ASSISTANT USING NVIDIA
RIVA AND PYGAME

รวินทร์ ไชยสิทธิพร

Rawin Chaisittiporn

*¹ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏจันทรเกษม*¹ Faculty of Science Chandrakasem Rajabhat University

E-mail ผู้รับผิดชอบบทความ (Corresponding Author): rawin.ch@chandra.ac.th

วันที่รับบทความ 6 มกราคม 2569

วันแก้ไขบทความ

7 เมษายน 2569

วันที่ตอบรับบทความ

30 เมษายน 2569

บทคัดย่อ

การอ่านออกเสียงภาษาอังกฤษได้อย่างถูกต้อง จำเป็นต้องมีการฝึกฝนเพื่อสร้างทักษะ งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเทคโนโลยีด้านปัญญาประดิษฐ์จาก NVIDIA Riva ซึ่งเป็นเทคโนโลยี รู้จำเสียงและสร้างเสียงพูด มาใช้เป็นเครื่องมือสำหรับการติดต่อสื่อสารระหว่างมนุษย์และ คอมพิวเตอร์ โดยได้นำโมเดลภาษาขนาดใหญ่ (LLM) มาใช้เพื่อทำหน้าที่เป็นตัวแทนอัจฉริยะ (Agent) ผ่านการพูดคุย และได้นำ pygame มาใช้เป็นส่วนติดต่อกราฟิก ในรูปแบบภาพเคลื่อนไหวรูปใบหน้า หุ่นยนต์ และได้ใช้ ROS2 เป็นส่วนเชื่อมต่อระหว่างส่วนงานต่างๆ โดยมีเป้าหมายเพื่อการฝึกฝนการ อ่านออกเสียงภาษาอังกฤษ และการสร้างเจตคติที่ดีด้านภาษาอังกฤษ การทำงานของระบบจะลงไป ที่เครื่อง Jetson Orin NX 16GB ซึ่งเป็นเครื่องคอมพิวเตอร์ขนาดเล็ก ประหยัดพลังงานแต่มี ประสิทธิภาพการทำงานสูง ทำให้ระบบสามารถทำงานแบบที่ไม่ต้องใช้เครือข่ายได้ (off-line) ซึ่งเพิ่ม ความเร็วในการทำงานและเพิ่มความปลอดภัยให้ระบบ ต่างจากระบบงานทั่วไป ที่ทำงานบนระบบ คลาวด์ เนื่องจากระบบนั้นจะมีเวลาหน่วงในการทำงานสูง และมีความปลอดภัยต่ำกว่า จากการ ทดสอบการทำงานของระบบ พบว่าสามารถทำงานได้อย่างดี ในเวลาตอบสนองต่ำและมีความแม่นยำ ในการประมวลผลของ NVIDIA Riva ในระดับสูง สรุปได้ว่า เทคนิคของงานวิจัยนี้มีความเหมาะสม สำหรับการพัฒนาเป็นนวัตกรรมทางด้านการสนทนาของตัวแทนอัจฉริยะ ที่เป็นประโยชน์ในอนาคต เป็นอย่างมากในหลายๆด้าน

คำสำคัญ: NVIDIA Riva; โมเดลภาษาขนาดใหญ่; pygame; ROS2

ABSTRACT

Developing correct English pronunciation requires practice in order to build skill. The objective of this research is to apply artificial intelligence technology from NVIDIA Riva, which provides speech recognition and speech synthesis, as a tool for human-computer interaction. A large language model (LLM) is used to function as an intelligent agent that communicates through spoken dialogue, and pygame is employed to create a graphical interface in the form of an animated robot face. ROS2 serves as the middleware connecting the various system components. The system's goal is to support English pronunciation training and to foster positive attitudes toward English learning.

The system is designed to run specifically on the Jetson Orin NX 16GB, a small, energy-efficient yet high-performance computer. This enables the system to operate offline, without relying on a network, which improves processing speed and enhances security—unlike typical cloud-based systems that often suffer from higher latency and lower security. Testing results show that the system performs well, with low response times and high processing accuracy from NVIDIA Riva. In conclusion, the techniques presented in this research are highly suitable for developing innovative conversational intelligent agents, offering significant potential benefits across many future applications.

Keywords: NVIDIA Riva; Large Language Model; pygame; ROS2

1. บทนำ

การนำเทคโนโลยีทางด้านปัญญาประดิษฐ์ มาประยุกต์ใช้กับงานในด้านต่างๆ จะช่วยเพิ่มคุณภาพของผลงานเหล่านั้นมากกว่าเทคนิคแบบเดิม งานวิจัยนี้ได้นำเทคโนโลยีของการรู้จำเสียงอัตโนมัติ (Automatic Speech Recognition) และการสร้างเสียงพูด (Speech Synthesis) ของ NVIDIA Riva [1] มาประยุกต์ใช้ร่วมกับส่วนของโมเดลภาษาขนาดใหญ่ (Large Language Model) โดยมีวัตถุประสงค์เพื่อสร้างสื่อการเรียนรู้ภาษาอังกฤษสำหรับเด็ก แบบสร้างประสบการณ์โดยตรง งานวิจัยได้นำจุดเด่นของเทคโนโลยีด้านปัญญาประดิษฐ์ล่าสุดของ NVIDIA เข้ามาประยุกต์ เพื่อสร้างสรรค์นวัตกรรมด้านการเรียนรู้เชิงสร้างสรรค์

จากการสืบค้นงานวิจัย พบว่าส่วนใหญ่ ไม่ได้กล่าวถึงการใช้ฮาร์ดแวร์ เช่น Jetson ซึ่งส่วนใหญ่ จะใช้ซอฟต์แวร์สำเร็จรูป ซึ่งมีค่าใช้จ่ายในการใช้งาน และไม่สามารถปรับเปลี่ยนเงื่อนไขของการทำงานได้ เพราะไม่ได้พัฒนาโปรแกรมของตนเองขึ้นมา [2] [3] และงานส่วนใหญ่ก็ไม่ได้ใช้รูปแบบของหุ่นยนต์ในการเป็นสื่อการเรียนรู้ หากเป็นงานวิจัยที่ใช้ฮาร์ดแวร์ ก็จะใช้ฮาร์ดแวร์ในระดับที่ประสิทธิภาพยังไม่สูงมาก จึงต้องใช้เทคนิคการทำงานในรูปแบบของการแก้ปัญหาเฉพาะหน้า [4] [5] เช่นการผสานระหว่างส่วนของ edge และ cloud ทำให้ไม่ได้ใช้จุดเด่น

ของระบบ edge อย่างเต็มที่ ได้แก่เรื่องเวลาหน่วงในระดับต่ำ (low latency) และความปลอดภัยจากการทำงานแบบเฉพาะที่ (off-line)

รูปแบบของงานวิจัยนี้เป็นงานวิจัยและพัฒนา โดยผู้วิจัยจึงมีแนวคิดที่จะพัฒนาต้นแบบของหุ่นยนต์สำหรับการสอนการอ่านออกเสียงภาษาอังกฤษ ด้วยเทคโนโลยีล่าสุดที่เหมาะสม เพื่อเพิ่มประสิทธิภาพของการเรียนรู้ภาษาอังกฤษของผู้สนใจทั่วไป ระบบทำงานบนเครื่องคอมพิวเตอร์ที่เหมาะสมกับงานด้านปัญญาประดิษฐ์ ได้แก่ Jetson Orin NX 16 GB ซึ่งใช้ซอฟต์แวร์ Jetpack เวอร์ชัน 6.0 ซึ่งมี GPU แบบ NVIDIA Ampere จำนวน 1024 core ทำให้สามารถรองรับการทำงานได้ตรงตามความต้องการ ในส่วนของซอฟต์แวร์ด้านการรู้จำเสียงอัตโนมัติ จะใช้ NVIDIA Riva ซึ่งทำงานได้ดีกว่าซอฟต์แวร์อื่นๆ ในประเภทเดียวกัน ระบบจะนำเสียงพูดจากไมโครโฟนที่ผ่านการรู้จำเสียงของ NVIDIA Riva จากนั้นส่งไปยังส่วนของโมเดลขนาดใหญ่ ในงานวิจัยใช้โมเดล LLaVa 1.5 7B ซึ่งมีจุดเด่นที่ขนาดเล็กและมีประสิทธิภาพการตอบคำถามได้อย่างถูกต้อง

ในส่วนของการติดต่อกับผู้ใช้จะใช้ pygame ซึ่งเป็นไลบรารีของ Python ที่เน้นงานด้านกราฟิกและเกม โดยนำเสนอเป็นใบหน้าหุ่นยนต์ในแบบภาพเคลื่อนไหว (animation) ที่สามารถแสดงอารมณ์ได้หลายรูปแบบ และจะใช้ ROS2 เป็นส่วนของการติดต่อสื่อสารระหว่างหน่วยต่างๆ ของระบบ โดยจะใช้การส่งข้อมูลผ่านทาง topic ของระบบ ROS2 จุดเด่นสำคัญของงานวิจัยนี้คือการทำงานแบบไม่ต้องติดต่ออินเทอร์เน็ต (off-line) เนื่องจาก Jetson Orin NX 16 GB จะเก็บข้อมูลโมเดลในตัวเองและด้วย GPU จำนวนมาก ทำให้สามารถทำงานได้ด้วยเวลาตอบสนองไม่นานเกินไป และสะดวกต่อการใช้งานในสถานที่จริง ที่อาจจะไม่มีสัญญาณอินเทอร์เน็ต หรือสัญญาณล่าช้า

ในระบบจะมีการทำงาน 2 ส่วน คือส่วนการพูดคุยระหว่างผู้ใช้กับโมเดลภาษาขนาดใหญ่ และส่วนของการเล่นเกม โดยจะเป็นเกมฝึกการพูดภาษาอังกฤษ ซึ่งระบบจะสุ่มประโยคภาษาอังกฤษขึ้นมาครั้งละ 1 ประโยค และรอให้ผู้พูด จากนั้นจะนำคำพูดไปเปรียบเทียบกับหาคำความถูกต้อง และให้คะแนน หรือแจ้งว่าอ่านผิด จากการทดสอบพบว่า ระบบสามารถทำงานในส่วนของการพูดคุยกับโมเดลภาษาขนาดใหญ่ได้ถูกต้องและรวดเร็ว ส่วนการเล่นเกมมีการประเมินถูกต้องเช่นกัน ด้วยเทคโนโลยีเหล่านี้ จะทำให้ผู้ใช้รู้สึกสนุกสนานกับการฝึกพูดภาษาอังกฤษ และเพิ่มทักษะด้านการพูดภาษาอังกฤษได้อย่างมีคุณภาพ

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

มีงานวิจัยที่นำการรู้จำเสียงหรือการสังเคราะห์เสียง มาช่วยในด้านการเรียนรู้ภาษาอังกฤษพอสมควร เช่นงานวิจัยของ สุธฤทัย อรุณศิริโรจน์ [3] ได้นำเทคโนโลยีของการวิเคราะห์เสียง (speech analyzer) มาใช้ในการเรียนการสอน เรื่องการออกเสียงภาษาอังกฤษ พบว่าช่วยในเรื่องการออกเสียงให้มีประสิทธิภาพมากขึ้น ณัฐพล กลุณ สุทธิกุล, ก้องภพ บุญมา, จิราภรณ์ เมืองประทับ, วสันต์ ณ ชัย และ เทพชัย ทรัพย์นิธิ [2] ได้นำเทคโนโลยีของการรู้จำเสียงอัตโนมัติ (ASR) ไปใช้กับแบบฝึกหัดการออกเสียงภาษาอังกฤษ กับผู้เรียนภาษาอังกฤษแบบ EFL (English as a Foreign Language) โดยพบว่ากลุ่มตัวอย่างร้อยละ 60 ได้ผลการออกเสียงภาษาอังกฤษดีขึ้นหลังจากการใช้งานระบบ ซึ่งในงานวิจัยได้ใช้เพียงเทคโนโลยีของการรู้จำเสียงอัตโนมัติเท่านั้น โดยทำงานในรูปแบบเว็บ (web-based)

สำหรับงานวิจัยที่นำโมเดลภาษาขนาดใหญ่ (LLM) มาใช้ด้านการศึกษายังมีไม่มาก เช่นงานวิจัยของ พรชัย จันทร์ถาวร, ปุ่น เทียงบุรณธรรม, ฉวีวรรณ เด่นไพบุลย์, ศุภวัฒน์กร วงศ์ธนสุ และ สมคิด แก้วทิพย์ [6] ได้นำ LLM มาช่วยสังเคราะห์องค์ความรู้จากงานวิจัยเกี่ยวกับ โครงการเมืองแห่งการเรียนรู้ (Learning Cities: LC) ใน

ประเทศไทย โดยได้สกัดประเด็นสำคัญจากงานวิจัย 36 โครงการใน 29 เมือง สุดท้ายนำผลจากโมเดลภาษาขนาดใหญ่และการวิเคราะห์เชิงโครงสร้างมาสังเคราะห์เป็นรูปแบบ เมืองแห่งการเรียนรู้

ในส่วนงานด้านการสร้างเสียงพูด ในประเทศไทยมีงานวิจัยด้านนี้ไม่มาก โดยมีงานวิจัยที่ประเทศไต้หวัน ได้แก่ Yi-Ching Huang และ Lung-Chuan Liao [7] ได้ศึกษาผลของการใช้สื่อดิจิทัลที่ผสานกับระบบแปลงข้อความเป็นเสียงพูด (Text To Speech) เพื่อพัฒนาทักษะการสะกดคำภาษาอังกฤษของนักเรียนระดับชั้นประถมไต้หวัน โดยได้ประเมินความสามารถในการสะกดคำภาษาอังกฤษ และการเรียนรู้ด้วยตนเอง พบว่าผู้เรียนเกิดแรงจูงใจ และสนุกกับการเรียน Tira Nur Fitria [8] ได้นำระบบแปลงข้อความเป็นเสียงพูด ไปใช้สร้างสื่อฟังภาษาอังกฤษสำหรับผู้เรียน โดยใช้แพลตฟอร์ม iSpeech พบว่ามีความสะดวก เป็นเครื่องมือที่ง่าย ประหยัดเวลา และมีประสิทธิภาพ ครูสามารถปรับแต่งเสียง สำเนียง และความเร็วได้ตามวัตถุประสงค์ของบทเรียน ส่งผลให้ผู้เรียนได้รับประสบการณ์ฟังที่เป็นธรรมชาติและยืดหยุ่นมากขึ้น

สำหรับงานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบระหว่างระบบ edge และ cloud ก็มีอยู่พอสมควร เช่นงานวิจัยของ Mohammad Jalili Torkamani และ Israt Zarin [9] ได้ทำงานวิจัยทางด้านการแปลงคำพูดเป็นการกระทำด้วยเทคโนโลยีของการรู้จำเสียงอัตโนมัติ (ASR) และ โมเดลภาษาขนาดใหญ่ (LLM) โดยการประมวลผลจะมีการปรับเปลี่ยนรูปแบบระหว่าง edge และ cloud โดยอัตโนมัติ โดยผู้วิจัยใช้ฮาร์ดแวร์ Jetson Mate Xavier NX ซึ่งมีประสิทธิภาพการทำงานไม่สูงมาก จึงต้องมีการแบ่งภาระงานระหว่าง edge และ cloud เช่นเดียวกับงานวิจัยของ Vishesh Narendra Pamadi และ Pushpa Singh [4] และงานวิจัยของ Mohammed Abdus Salam [5] ก็ได้ทำการเปรียบเทียบการทำงานระหว่างระบบ edge และ cloud เช่นกัน ซึ่งพบว่าการทำงานแบบ edge จะได้เปรียบในเรื่องของเวลาหน่วงซึ่งมีค่าน้อยกว่า แต่ระบบ cloud จะให้ขอบเขตการทำงานที่กว้างกว่าสำหรับผู้วิจัยจึงมีความประสงค์จะพัฒนาระบบการทำงานแบบ edge ที่สมบูรณ์แบบเพื่อการใช้งานได้อย่างแท้จริง ทำให้ระบบมีความปลอดภัยในระดับสูง

3. วิธีการวิจัย

งานวิจัยนี้ผู้วิจัยได้พัฒนาแอปพลิเคชันสำหรับการฝึกการสนทนาภาษาอังกฤษ โดยการทำงานทั้งหมดจะทำงานแบบไม่ต้องการเครือข่ายอินเทอร์เน็ต (off-line) ระบบทำงานในเครื่องคอมพิวเตอร์แบบ Jetson Orin NX หน่วยความจำ 16 GB โดยมีจอมอนิเตอร์ที่มีลำโพงในตัว, คีย์บอร์ด, เมาส์, ไมโครโฟน เป็นอุปกรณ์ประกอบการทำงาน และติดตั้งซอฟต์แวร์ต่างๆที่เกี่ยวข้อง พัฒนาระบบงาน ทดสอบและปรับปรุงการทำงาน

3.1 ติดตั้งระบบการทำงานที่จำเป็น

ซอฟต์แวร์หลักได้แก่ Jetpack 6.0 ซึ่งประกอบไปด้วย Ubuntu 22.04 และซอฟต์แวร์สำคัญด้านปัญญาประดิษฐ์ เช่น CUDA Toolkit, TensorRT เป็นต้น ส่วนของการรู้จำเสียงอัตโนมัติและการสร้างเสียงพูด จะติดตั้งโปรแกรม NVIDIA Riva ในรูปแบบ docker โดยจะต้องลงทะเบียนในส่วนของ NVIDIA GPU Cloud (NGC) เพื่อให้ได้กุญแจการใช้งาน หรือ API Key ซึ่งจะมีสิทธิ์การใช้งานฟรีเป็นเวลา 90 วัน สำหรับการใช้งาน NVIDIA Riva หลังจากติดตั้งแล้ว จะทำการเช็คค่าพารามิเตอร์ต่างๆที่เกี่ยวข้อง เช่นเลขรหัสของลำโพงและไมโครโฟน เป็นต้น

ส่วนของโมเดลภาษาขนาดใหญ่จะทำการติดตั้งโมเดล LLaVA 1.5 7B ผ่าน docker ซึ่งเป็นโมเดลขนาดกลาง ทำให้ใช้พื้นที่หน่วยความจำไม่มากนักไป ซึ่งจากการทดสอบพบว่าโมเดลนี้มีความฉลาดพอสมควร และไม่กินพื้นที่มากนักไป สามารถทำงานใน Jetson Orin NX 16GB ได้เป็นอย่างดี

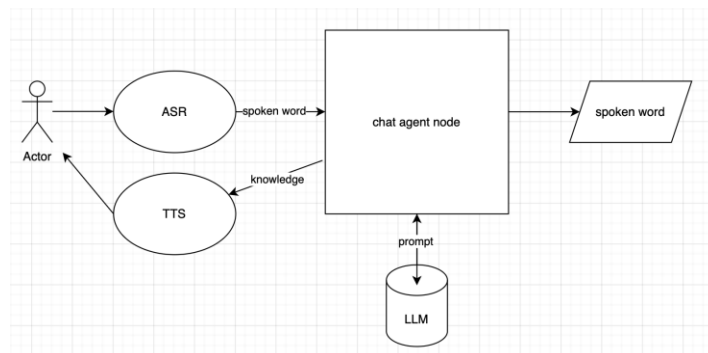
โปรแกรม ROS2 จะติดตั้งเวอร์ชัน Humble ซึ่งทำงานบนระบบปฏิบัติการ Ubuntu 22.04 ซึ่งติดตั้งเวอร์ชัน desktop และสำหรับไลบรารี pygame จะติดตั้งจากอินเทอร์เน็ตเช่นกัน

3.2 พัฒนาโปรแกรมของระบบ

ส่วนโปรแกรมที่ต้องพัฒนาได้แก่ node ของระบบ ROS2 โดยจะมี 2 node ในการทำงานหลัก ได้แก่

1. ส่วนของการรับข้อมูลจาก ASR และส่งข้อมูลไปที่ TTS (chat agent node)

การทำงานของ node นี้ คือการติดต่อกับ Riva ASR และ TTS เพื่อรับข้อมูลเสียงพูดของผู้ใช้ และส่งไปให้ node อื่นเพื่อประมวลผล และรับข้อมูลคำพูดจาก node อื่นเข้ามาเพื่อส่งเสียงพูดออกไปให้ผู้ใช้งานได้ยิน โดยทำงานแบบปฏิสัมพันธ์ (interactive) นอกจากนั้นแล้ว node นี้ยังติดต่อกับส่วนของ โมเดลภาษาขนาดใหญ่เพื่อพูดคุย (chat) และรับ,ส่งข้อมูลระหว่างกัน ภาพที่ 1 แสดงไต่อะแกรมการทำงานของ chat agent node



ภาพที่ 1 ไต่อะแกรมการทำงานของ chat agent node

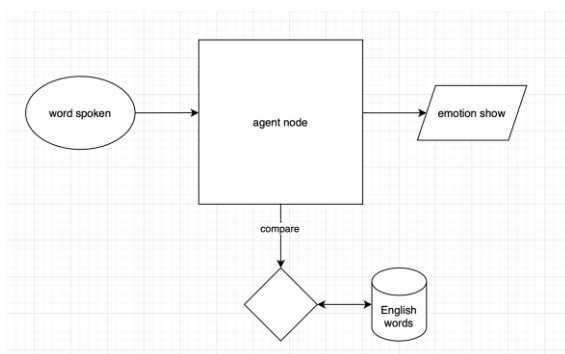
2. ส่วนของการแสดงผลกราฟิกและการประมวลผลของระบบ (agent node)

สำหรับ node นี้ จะเรียกการทำงานของไลบรารี pygame เพื่อแสดงผลภาพเคลื่อนไหวรูปใบหน้าหุ่นยนต์ และข้อความต่างๆ โดยจะรับ ส่งข้อมูลกับ node อื่นๆ และนำข้อมูลไปประมวลผล ก่อนจะแสดงผลภาพเคลื่อนไหวตามอารมณ์ต่างๆ นอกจากนั้น node นี้จะประมวลผลเงื่อนไขต่างๆที่เกี่ยวกับระบบ เช่นการแสดงผลอารมณ์และการตรวจสอบความถูกต้องของการเล่นเกมฝึกออกเสียงภาษาอังกฤษ เมื่อ node ทำงานจะเป็นดังภาพที่ 2



ภาพที่ 2 การทำงานของภาพเคลื่อนไหวใบหน้าหุ่นยนต์ของ agent node

ภาพที่ 3 แสดงไดอะแกรมการทำงานของ agent node ซึ่งส่วนของ node จะอ่านข้อมูลชุดคำพูดภาษาอังกฤษมาเปรียบเทียบกับเสียงการพูดภาษาอังกฤษของผู้ใช้เพื่อแสดงผลการประเมินผลเป็นอารมณ์ของหุ่นยนต์



ภาพที่ 3 ไดอะแกรมการทำงานของ agent node

ภาพที่ 4 แสดงการทำงานขณะเข้าสู่โหมดเล่นเกมอ่านออกเสียงภาษาอังกฤษ จะแสดงใบหน้าของหุ่นยนต์ ซึ่งสามารถแสดงอารมณ์ต่างๆได้ เพื่อเพิ่มประสบการณ์การใช้งานของผู้ใช้



ภาพที่ 4 การทำงานของเกมฝึกอ่านออกเสียงภาษาอังกฤษ

3.3 ทดสอบการทำงานของระบบ

หลังจากได้พัฒนาโปรแกรมที่จำเป็นครบถ้วน ผู้วิจัยได้ทำการทดสอบการทำงานของระบบ โดยมีขั้นตอนการดำเนินการดังนี้

1. เปิดการทำงานของ NVIDIA Riva
2. เปิดการทำงานของ โมเดลภาษาขนาดใหญ่
3. เรียกการทำงานของ agent node
4. เรียกการทำงานของ chat agent node
5. ทดสอบการทำงานทั้งส่วนของการพูดคุยและการเล่นเกม
6. ปรับปรุงแก้ไขเพื่อพัฒนาระบบ

การทดสอบในส่วนของการพูดคุยกับโมเดลภาษาขนาดใหญ่ จะทดสอบใน 2 ด้าน ได้แก่ด้านความถูกต้องและประสิทธิภาพ ด้านความถูกต้องผู้วิจัยใช้การทดสอบโดยเปรียบเทียบระหว่างชุดของคำถามที่เตรียมไว้จำนวน 100 ชุด เทียบกับคำตอบที่ได้จากระบบ และใช้การวัดอัตราความผิดพลาดของการตอบ (answer error rate) ซึ่งการทดสอบนั้น ผู้วิจัยประเมินความถูกต้องของการตอบโดยใช้ข้อมูลจากเว็บไซต์ Wikipedia เป็นข้อมูลอ้างอิง จากนั้นบันทึกผลการทดสอบ โดยสมการของอัตราความผิดพลาดเป็นดังสมการที่ (1)

$$AER = \frac{AE}{N} \quad (1)$$

AER = อัตราความผิดพลาดของการตอบ

AE = จำนวนคำถามที่ตอบผิด

N = จำนวนคำถามทั้งหมด

นอกจากนั้น ผู้วิจัยได้ทดสอบ ความถูกต้องของการพูดของ TTS โดยจะวัดอัตราความผิดพลาดของการพูด (speech error rate) จากการเปรียบเทียบระหว่างคำพูดที่ระบบเขียนออกมาทางจอภาพและเสียงที่พูดออกมา โดยใช้ข้อมูลจากเว็บไซต์ translate.google.com เป็นข้อมูลอ้างอิง ตามสมการที่ (2) ทดสอบเป็นจำนวน 100 ครั้ง

$$SER = \frac{SE}{N} \quad (2)$$

SER = อัตราความผิดพลาดของการพูด

SE = จำนวนการพูดที่ผิดพลาด

N = จำนวนคำพูดทั้งหมด

ในด้านประสิทธิภาพ ผู้วิจัยได้ทดสอบความเร็วด้านการแปลงคำพูดเป็นข้อความของ TTS โดยจะวัดจากเวลาเฉลี่ยตั้งแต่เริ่มพูดเข้าไปผ่านไมโครโฟนจนถึงเวลาที่ระบบแสดงผลคำพูดนั้นออกมา ซึ่งทดสอบด้วยคำพูดจำนวน 100 ชุด

สำหรับการทดสอบในส่วนของการเล่นเกมฝึกพูดภาษาอังกฤษ จะทดสอบโดยการพูดตามประโยคที่เห็น โดยมีการทดสอบทั้งการพูดถูก (positive testing) และกล่าวผิด (negative testing) ด้วยจำนวนการทดสอบอย่างละ 100 คำพูด และทำการวัดค่า EER ดังสมการที่ (3)

$$EER = \frac{EE}{N} \quad (3)$$

EER = อัตราความผิดพลาดของการประเมินผลของระบบ

EE = จำนวนความผิดพลาดของการประเมินผลของระบบ

N = จำนวนคำพูดทั้งหมด

4. ผลการวิจัย

หลังจากการพัฒนาระบบ ผู้วิจัยได้ทำการทดสอบผลการดำเนินการตามที่ได้ออกแบบไว้ โดยได้ผลการทดสอบดังตารางที่ 1 - 3

ตารางที่ 1 ผลการทดสอบความถูกต้องของการพูดคุยกับโมเดลภาษาขนาดใหญ่

ตัวชี้วัด	ผลที่ได้
อัตราการผิดพลาดของการตอบ	0.03
อัตราการผิดพลาดของการพูด	0.01

ตารางที่ 2 การทดสอบประสิทธิภาพของการพูดคุยกับโมเดลภาษาขนาดใหญ่

ตัวชี้วัด	ผลการทดสอบ
ความเร็วเฉลี่ยการพูดของ TTS	0.43 วินาที

สำหรับการทดสอบการทำงานของการเล่นเกมนอกเสียงภาษาอังกฤษ ได้ผลการทดสอบดังตารางที่ 3

ตารางที่ 3 การทดสอบการประมวลผลของเกมพูดภาษาอังกฤษ

ตัวชี้วัด	ผลที่ได้
อัตราความผิดพลาดของการประเมินผลของระบบ	0.02

5. อภิปรายผลและข้อเสนอแนะการวิจัย

จากการทดสอบการทำงานของระบบ พบว่า NVIDIA Riva มีการทำงานทางด้านการรู้จำเสียง และการสังเคราะห์เสียงเป็นอย่างดี โดยมีความแม่นยำในการรู้จำเสียงที่ระดับสูง มีการตอบสนองได้อย่างรวดเร็ว และการใช้ Jetson Orin NX 16GB เป็นอุปกรณ์ในการทำงานมีความเหมาะสม สามารถรองรับการประมวลผลปัญญาประดิษฐ์ในแบบ Edge Technology ได้เป็นอย่างดี ข้อดีของการใช้ Jetson Orin NX นอกจากความเร็วในการตอบสนองแล้ว ยังมีเรื่องของการทำงานในตัวเอง (off-line) ทำให้ลดการพึ่งพาระบบเครือข่ายและเพิ่มความปลอดภัยให้กับข้อมูลในระบบ สอดคล้องกับพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล (PDPA)

การใช้โมเดลภาษาขนาดใหญ่บน Jetson Orin NX ช่วยให้สามารถติดตั้งลงในเครื่องคอมพิวเตอร์เพียงเครื่องเดียวได้ ทำให้เกิดความรวดเร็ว และส่วนของโมเดลภาษาขนาดใหญ่สามารถประยุกต์การทำงานได้หลายรูปแบบ และสามารถเรียนรู้ความรู้เพิ่มเติมในรูปแบบต่อยอดได้ ดังนั้นเทคนิคนี้จะช่วยสร้างนวัตกรรมใหม่ๆที่เป็นประโยชน์ได้อีกมาก และมีความเหมาะสมสำหรับการใช้ในหน่วยงานที่ห่างไกลจากระบบอินเทอร์เน็ต เช่น โรงเรียนในชนบท

การใช้งาน NVIDIA Riva ก็สามารถเรียนรู้แบบต่อยอดได้เช่นกัน ตัวอย่างเช่นการใช้เสียงเฉพาะในบริบทงานบางด้าน หรือภาษาท้องถิ่นแต่ละที่ สำหรับ ROS2 มีข้อดีคือความสามารถด้านการเชื่อมโยงหน่วยโปรแกรมต่างๆเข้าด้วยกัน ทำให้ผู้ออกแบบระบบสามารถสร้างซอฟต์แวร์ในรูปแบบแบ่งส่วน (modularity) ได้ ทำให้สะดวกต่อการพัฒนา สามารถปรับเปลี่ยนโมเดลทางด้านปัญญาประดิษฐ์ในอนาคตได้โดยไม่กระทบกับภาพรวมของระบบ และแก้ปัญหาต่างๆที่เกิดขึ้นได้ง่าย ส่วนการใช้ pygame เพิ่มความน่าสนใจให้กับการใช้งานระบบ ทำให้ผู้ใช้รู้สึกเพลิดเพลินกับการใช้งาน และสร้างปฏิสัมพันธ์กับระบบได้อย่างมีประสิทธิภาพ pygame เพิ่มความสามารถด้านประสบการณ์ของผู้ใช้ (user experience) และสามารถรับข้อมูลต่างๆผ่านทางระบบ ROS2 ซึ่งทำให้การเชื่อมโยงกับงานในส่วนอื่นเป็นไปได้อย่างสะดวก

การพัฒนางานวิจัยต่อยอดในอนาคต สามารถทำได้อย่างหลากหลายรูปแบบตามบริบทงานต่างๆ และส่วนของโมเดลภาษาขนาดใหญ่ก็สามารถเลือกประเภทและขนาดได้ตามความเหมาะสมกับคุณสมบัติ (specification) ของอุปกรณ์ที่ใช้ การปรับส่วนของแอปพลิเคชันและการปฏิสัมพันธ์กับผู้ใช้ก็สามารถปรับเปลี่ยนได้ตามความเหมาะสม และด้วยความสามารถของ NVIDIA Riva จะเพิ่มคุณค่าของนวัตกรรมนั้นเป็นอย่างมาก ดังนั้นองค์ความรู้จากการวิจัยนี้จะเป็นประโยชน์ในเชิงเศรษฐกิจและสังคมอย่างมาก

6. เอกสารอ้างอิง

[1] NVIDIA Corporation. (2025, April 3). NVIDIA Riva user guide (Release 2.19.0).

<https://docs.nvidia.com/deeplearning/riva/user-guide/docs/index.html>

[2] Kritsuthikul, N., Boonma, K., Muangprathub, J., Na Chai, W., & Supnithi, T. (2024).

Improve English pronunciation at word level for Thai EFL learners in southern region using end-to-end automatic speech recognition. *Proceedings of the 32nd International Conference on Computers in Education (ICCE 2024)*. <https://doi.org/10.58459/icce.2024.4917>

- [3] Arunsitot, S. (2017). Implementing a speech analyzer software to enhance English pronunciation competence of Thai students. *วารสารศึกษาศาสตร์ มหาวิทยาลัยบูรพา*, 28(2), 116–129. <https://opac1.lib.buu.ac.th/medias3/edu28n2p116-129.pdf>
- [4] Pamadi, V. N., & Singh, P. (n.d.). Edge AI vs. Cloud AI: A comparative study of performance, latency, and scalability. *International Journal of Research in Modern Engineering & Emerging Technology*. <https://ijrmeet.org/edge-ai-vs-cloud-ai-a-comparative-study-of-performance-latency-and-scalability/>
- [5] Salam, M. A. (2024). Edge-cloud collaboration in real-time AI applications. *Iconic Research and Engineering Journals*, 8(6), 1125–1136. <https://www.irejournals.com/formatedpaper/1709991.pdf>
- [6] Jantaworn, P., Thiengburanathum, P., Denpaiboon, C., Wongthanavas, S., & Keawthip, S. (2024). City to city of learning in Thailand: From research synthesis by large language models. *Asian Journal of Adult Education*. <https://so01.tci-thaijo.org/index.php/AJA/article/download/274059/178740/1114876>
- [7] Huang, Y.-C., & Liao, L.-C. (2015). A study of text-to-speech (TTS) in children's English learning. *Teaching English with Technology*, 15(1), 14–30. <https://files.eric.ed.gov/fulltext/EJ1140575.pdf>
- [8] Fitria, T. N. (2024). Utilizing text-to-speech (TTS) technology in creating listening materials for English language teaching (ELT). *Journal of English Language and Culture*, 15(1), 73–85. <https://journal.ubm.ac.id/index.php/english-language-culture/article/view/4995>
- [9] Torkamani, M. J., & Zarin, I. (2025). *Adaptive edge-cloud inference for speech-to-action systems using ASR and large language models* (arXiv:2512.12769). arXiv. <https://arxiv.org/abs/2512.12769>