

การพัฒนาแอปพลิเคชันต้นแบบรู้จำวัตถุด้วยสมาร์ทโฟน “VocabEyes”
เพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) สำหรับผู้พิการทางสายตา

วารุพินท์ บุตรชั้น

TNI

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

บัณฑิตศึกษา สาขาเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีไทย-ญี่ปุ่น

ปีการศึกษา 2568

DEVELOPMENT OF A SMARTPHONE-BASED OBJECT RECOGNITION
PROTOTYPE APPLICATION “VOCABEYES” FOR TRILINGUAL VOCABULARY LEARNING
(THAI-JAPANESE-ENGLISH) FOR THE VISUALLY IMPAIRED

Warupin Butkan

TNI

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology
Graduate Studies

Thai-Nichi Institute of Technology

Academic Year 2025

หัวข้อวิทยานิพนธ์

การพัฒนาแอปพลิเคชันต้นแบบรู้จำวัตถุด้วยสมาร์ทโฟน
“VocabEyes” เพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-
อังกฤษ) สำหรับผู้พิการทางสายตา

โดย

วารุพินท์ บุตรชั้น

สาขาวิชา

เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ดร.สะพรั่งสิทธิ์ มฤตสาธ

บัณฑิตศึกษา สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้บัณฑิตวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....รองอธิการบดีฝ่ายวิชาการ

(รองศาสตราจารย์ ดร.วรกร ศรีเชวงทรัพย์)

วันที่.....เดือน.....พ.ศ.....

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร.ทิมพันธ์ เจริญพงษ์)

.....กรรมการ

(ดร.ประมุข บุญเสียง)

.....กรรมการ

(ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ดร.สะพรั่งสิทธิ์ มฤตสาธ)

วารุพินท์ บุตรขัน : การพัฒนาแอปพลิเคชันต้นแบบบนสมาร์ตโฟนระบบปฏิบัติการแอนดรอยด์
เพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) สำหรับผู้พิการทางสายตา. อาจารย์ที่ปรึกษา :
ดร.สพรั่งสิทธิ์ มฤตุสาร, 98 หน้า.

งานวิจัยนี้นำเสนอการพัฒนาแอปพลิเคชันต้นแบบบนสมาร์ตโฟนระบบปฏิบัติการแอนดรอยด์
ชื่อ “VocabEyes” เพื่อช่วยเหลือผู้พิการทางสายตาในการเรียนรู้คำศัพท์สามภาษา (ไทย ญี่ปุ่น และ
อังกฤษ) จากสิ่งแวดล้อมรอบตัว โดยมีมุ่งแก้ไขข้อจำกัดของเครื่องมือช่วยเหลือในปัจจุบันที่มักต้องพึ่งพา
อินเทอร์เน็ตและขาดการรองรับการอ่านออกเสียงหลายภาษาอย่างต่อเนื่อง ระบบนี้ได้รับการออกแบบ
ให้ประมวลผลบนอุปกรณ์แบบออฟไลน์ 100% (On-Device Processing) โดยประยุกต์ใช้เทคโนโลยี
คอมพิวเตอร์วิทัศน์ในการจำแนกภาพแบบเติมเฟรมผ่านโมเดล MobileNetV3-Large ที่ผ่านการฝึกฝน
ล่วงหน้าบนชุดข้อมูล ImageNet-21K-P (11,221 คลาส) ซึ่งมีจุดเด่นด้านความหน่วงต่ำและประหยัด
ทรัพยากรฮาร์ดแวร์บนอุปกรณ์พกพา โดยระบบจะแปลงข้อมูลภาพเป็นภาษาต่างๆโดยใช้พจนานุกรมที่
ผู้วิจัยจัดทำขึ้นตามโครงสร้างของ WordNet ซึ่งเชื่อมโยงป้ายกำกับภาพเข้ากับคำศัพท์ภาษาอังกฤษ
10,000 คำ ภาษาไทย 8,500 คำ และภาษาญี่ปุ่น 6,000 คำ โดยประมาณ จากนั้นจึงอ่านออกเสียง
ผลลัพธ์อย่างต่อเนื่องด้วยเทคโนโลยี Text-to-Speech (TTS) แบบหลายภาษา นอกจากนี้ ส่วนต่อประสาน
กับผู้ใช้ยังยึดหลักการออกแบบเพื่อมวลชน (Universal Design) ผ่านการสั่งงานด้วยท่าทางการสัมผัส
แบบเต็มหน้าจอและใช้สีที่มีความเปรียบต่างสูง

การประเมินประสิทธิภาพทางเทคนิคได้ดำเนินการโดยใช้ภาพทดสอบ 600 ภาพ จากวัตถุ
เป้าหมาย 10 ชนิด ใน 6 สถานการณ์จำลองที่ซับซ้อน เช่น ฉากหลังรบกวน การถูกบดบังบางส่วน และ
สภาพแสงน้อย ผลการทดสอบพบว่าระบบมีความแม่นยำเฉลี่ยอันดับแรก (Top-1 Accuracy) สูงถึง
93.83% และความแม่นยำใน 5 อันดับแรก (Top-5 Accuracy) 98.33% สำหรับการประเมินการใช้งาน
เบื้องต้นได้ทำการทดสอบกับกลุ่มตัวอย่างอาสาสมัครสายตาปกติที่จำลองการปิดกั้นการมองเห็นจำนวน
5 คน จากผลการประเมินพบว่าได้คะแนนความสามารถในการใช้งาน (System Usability Scale: SUS)
เฉลี่ยที่ 89.00 ซึ่งอยู่ในระดับดีเยี่ยม (Excellent) และมีคะแนนความพึงพอใจ 4.80 จาก 5.00 อย่างไรก็ตาม
มีข้อจำกัดที่สำคัญ คือผลลัพธ์ได้มาจากการประเมินกับกลุ่มตัวอย่างผู้ที่มีสายตาปกติซึ่งจำลองปิด
กั้นการมองเห็นเท่านั้น ผลการวิจัยนี้จึงเป็นการแสดงศักยภาพของสถาปัตยกรรมออฟไลน์ และเป็นเพียง
การพิสูจน์แนวคิด (Proof-of-Concept) สำหรับเครื่องมือส่งเสริมการเรียนรู้ภาษาต่างประเทศเบื้องต้น
ทั้งนี้ ในอนาคตจำเป็นต้องมีการทดสอบทางคลินิก (Clinical Validation) กับผู้พิการทาง
สายตาจริง เพื่อยืนยันประสิทธิภาพและความสามารถในการใช้งานจริงอย่างเป็นรูปธรรมต่อไป

บัณฑิตศึกษา

สาขาวิชา เทคโนโลยีสารสนเทศ

ปีการศึกษา 2568

ลายมือชื่อนักศึกษา

ลายมือชื่ออาจารย์ที่ปรึกษา

WARUPIN BUTKAN : DEVELOPMENT OF A SMARTPHONE-BASED OBJECT RECOGNITION PROTOTYPE APPLICATION “VOCABEYES” FOR TRILINGUAL VOCABULARY LEARNING (THAI-JAPANESE-ENGLISH) FOR THE VISUALLY IMPAIRED. ADVISOR : DR.SAPRANGSIT MRUETUSATORN, 98 PP.

This research presents the development of an Android-based prototype application named “VocabEyes” to assist the visually impaired in learning trilingual vocabulary (Thai, Japanese, and English) from their surrounding environment. Addressing the limitations of existing assistive tools that often require internet connectivity and lack continuous multilingual support, this system operates entirely offline through on-device processing. The application utilizes computer vision technology for whole-frame image classification, employing a MobileNetV3 model pre-trained on the ImageNet-21K-P dataset (11,221 classes), which was selected for its low latency and high hardware efficiency on mobile devices. To convert visual data into language, the system integrates a custom dictionary based on WordNet semantic mapping, linking image labels to approximately 10,000 English, 8,500 Thai, and 6,000 Japanese words. The results are then continuously read aloud using multilingual Text-to-Speech (TTS) technology. The user interface follows Universal Design principles, incorporating full-screen gesture controls and high-contrast visuals to ensure accessibility.

Technical evaluation was conducted utilizing 600 test images of 10 targeted objects across six distinct scenarios, including complex cases such as background clutter, partial occlusion, and low-light conditions. Based on these 10 objects, the results revealed a high average Top-1 Accuracy of 93.83% and a Top-5 Accuracy of 98.33%. As a notable limitation, the preliminary usability study was conducted with a sample group of only 5 blindfolded normal-vision volunteers to simulate visual impairment. Although the evaluation yielded an average System Usability Scale (SUS) score of 89.00 (Excellent level) and a user satisfaction rating of 4.80 out of 5.00, these results should be interpreted with caution. These findings highlight the potential of the offline architecture and primarily serve for a proof-of-concept as an introductory foreign language learning tool. Comprehensive clinical validation with actual visually impaired users remains essential in future work to definitively confirm its real-world usability.

Graduate Studies

Field of Study Information Technology

Academic Year 2025

Student’s Signature.....

Advisor’s Signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงอย่างสมบูรณ์ได้ด้วยความเมตตาและความกรุณาอย่างสูงจากท่านอาจารย์ที่ปรึกษา ดร.สพรั่งสีทธิ มฤตุสาธร์ ที่ได้เสียสละเวลาอันมีค่าของท่านในการกรุณาให้คำปรึกษา แนะนำแนวทางในการศึกษาวิจัยด้วยความเอาใจใส่เป็นอย่างยิ่ง ผู้วิจัยขอกราบขอบพระคุณในความเมตตาและโอกาสที่ท่านได้มอบให้เสมอมาซึ่งเป็นแรงผลักดันสำคัญยิ่งที่ทำให้งานวิจัยเรื่องการพัฒนาแอปพลิเคชัน VocabEyes นี้บรรลุตามวัตถุประสงค์ที่ตั้งไว้

ผู้วิจัยขอกราบขอบพระคุณท่านประธานกรรมการ และคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ได้แก่ รองศาสตราจารย์ ดร.ทิมพันธุ์ เจริญพงษ์ ดร.ประมุข บุญเสียง ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์ ที่ได้กรุณาให้ข้อเสนอแนะ คำแนะนำ รวมถึงแนวคิดจากหลากหลายมุมมอง อันเป็นประโยชน์อย่างยิ่งในการปรับปรุงเนื้อหาและระเบียบวิธีวิจัยให้มีความสมบูรณ์และถูกต้องตามหลักวิชาการมากยิ่งขึ้น

ความสำเร็จของงานวิจัยชิ้นนี้มาจากความกรุณา การสนับสนุน และความร่วมมืออย่างดียิ่งจากอาจารย์ที่ปรึกษา คณะกรรมการสอบวิทยานิพนธ์ ซึ่งเป็นผู้ทรงคุณวุฒิตามที่กล่าวถึงในข้างต้น รวมถึงผู้เกี่ยวข้องเช่น อาสาสมัครกลุ่มตัวอย่างทั้ง 5 ท่าน ได้แก่ ผู้เชี่ยวชาญด้านต่างๆ ชาวญี่ปุ่น ลาม ภาษาญี่ปุ่น-ไทย และรองผู้อำนวยการโรงเรียนชาวไทย ที่ได้สละเวลาอันมีค่าในการร่วมทดสอบและให้ข้อมูล รวมถึงคำแนะนำที่เป็นประโยชน์อย่างยิ่งเพื่อเป็นข้อเสนอแนะต่อการปรับปรุงให้แอปพลิเคชันมีความสมบูรณ์และตอบโจทย์การใช้งานจริงได้ดียิ่งขึ้นในอนาคต

นอกจากนี้ ผู้วิจัยขอขอบคุณสถาบันเทคโนโลยีไทย-ญี่ปุ่น ที่เปิดโอกาสให้เข้ามาเรียนรู้สิ่งใหม่ อันเป็นประโยชน์ เพื่อเป็นโอกาสในการเพิ่มช่องทางการประกอบอาชีพให้กว้างขวางยิ่งขึ้น รวมถึงขอขอบคุณเพื่อนร่วมการศึกษาทุกท่านที่คอยเป็นกำลังใจและแลกเปลี่ยนความรู้ซึ่งกันและกันตลอดระยะเวลาการศึกษาที่ผ่านมา

ท้ายที่สุดนี้ ผู้วิจัยขอกราบขอบพระคุณบิดา มารดา สมาชิกในครอบครัวทุกท่านที่เป็นกำลังใจสำคัญยิ่ง เป็นผู้สนับสนุนทั้งทางด้านร่างกาย แรงใจ รวมถึงทุนทรัพย์ในการสนับสนุนทุกอย่างก้าวของการศึกษาจนประสบความสำเร็จ คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบให้เป็นกตเวทิต์แก่ผู้มีพระคุณทุกท่าน และหวังเป็นอย่างยิ่งว่าแอปพลิเคชัน VocabEyes จะเป็นประโยชน์ต่อวงการการศึกษาและการพัฒนาสื่อการเรียนรู้ภาษาในอนาคตสืบไป

วารุพินท์ บุตรขัน

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูป.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ขอบเขตของการวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	5
1.5 นิยามศัพท์เฉพาะ.....	5
2 ทฤษฎี หลักการ และงานวิจัยที่เกี่ยวข้อง.....	7
2.1 ภาพรวมของการทบทวนวรรณกรรม.....	7
2.2 แนวคิดพื้นฐานของการรู้จำวัตถุด้วยการจำแนกภาพ.....	7
2.3 พื้นฐานที่เกี่ยวข้องกับการจำแนกภาพ.....	10
2.4 โมเดลสำหรับงานบนอุปกรณ์พกพา.....	12
2.5 ชุดข้อมูลภาพ (Image Dataset).....	13
2.6 แนวคิดการสร้างและจัดการคลังคำศัพท์หลายภาษา.....	13
2.7 Flutter.....	14
2.8 เทคโนโลยี Text-to-Speech (TTS).....	15
2.9 แนวคิดการออกแบบเพื่อมวลชน (Universal Design).....	16
2.10 เมตริกการประเมิน (Evaluation Metrics).....	17

สารบัญ (ต่อ)

บทที่		หน้า
2	2.11 งานวิจัยที่เกี่ยวข้อง.....	20
	2.12 แอปพลิเคชันสำหรับผู้พิการทางสายตา.....	28
3	วิธีการวิจัย.....	30
	3.1 บทนำ.....	30
	3.2 ช่องว่างของงานวิจัย (Research Gap)	30
	3.3 กระบวนการวิจัย.....	31
	3.4 ระยะเวลาการดำเนินงานวิจัย.....	47
4	ผลการวิจัย.....	48
	4.1 บทนำ.....	48
	4.2 ผลการพัฒนาและออกแบบแอปพลิเคชัน	48
	4.3 ผลการประเมินด้านประสิทธิภาพของแอปพลิเคชัน	59
	4.4 ผลการประเมินความสามารถในการใช้งาน (System Usability Scale : SUS).....	60
	4.5 ผลการประเมินความพึงพอใจของผู้ใช้ (Satisfaction Evaluation)	62
	4.6 วิเคราะห์ความสอดคล้องกับข้อกำหนดและคุณสมบัติหลัก.....	65
	4.7 การอภิปรายผลลัพธ์	66
	4.8 สรุปผลการดำเนินงานวิจัยโดยสังเขป.....	66
5	บทสรุป อภิปราย และข้อเสนอแนะ.....	67
	5.1 สรุปผลการวิจัย.....	67
	5.2 การอภิปรายผล	67
	5.3 ปัญหา อุปสรรค และข้อจำกัดในการวิจัย.....	68
	5.4 ข้อเสนอแนะ	70
	บรรณานุกรม	73

สารบัญ (ต่อ)

บทที่	หน้า
ภาคผนวก	78
ภาคผนวก ก. เอกสารนำเสนอสอบป้องกันวิทยานิพนธ์	79
ประวัติย่อผู้วิจัย.....	98



สารบัญตาราง

ตาราง	หน้า
2.1 สรุปชุดข้อมูลสำหรับโมเดลจำแนกภาพ	24
2.2 สรุปความรู้ด้านการออกแบบระบบ.....	26
2.3 สรุปแอปพลิเคชันสำหรับผู้พิการทางสายตา	28
3.1 การเปรียบเทียบระบบช่วยเหลือผู้พิการทางสายตากับงานวิจัยของเรา	30
3.2 สรุปข้อกำหนดและคุณสมบัติหลัก	34
3.3 แบบประเมินความสามารถในการใช้งาน.....	43
3.4 แบบประเมินความพึงพอใจของผู้ใช้	44
3.5 แบบเก็บข้อมูลผู้ตอบแบบสอบถาม	45
3.6 กำหนดการดำเนินงานวิจัย	47
4.1 ผลการประเมินความแม่นยำของระบบ	59
4.2 ผลคะแนน SUS รายบุคคล	60
4.3 ผลการประเมินความพึงพอใจ	62
4.4 สรุปข้อคิดเห็นจากผู้ใช้งาน	64

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญรูป

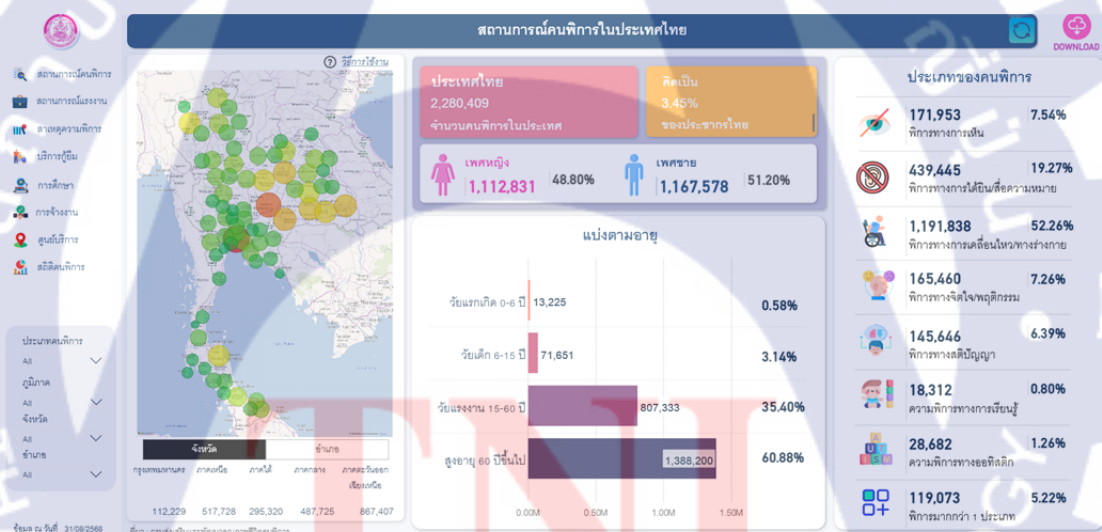
รูป	หน้า
1.1 สถานการณ์ผู้พิการในประเทศไทย.....	1
2.1 แผนภาพแสดงลำดับการตรวจจับวัตถุ	8
2.2 การประยุกต์ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน	12
2.3 แผนภาพโดยรวมของระบบ TTS สองภาษา ไทย-อังกฤษ	15
2.4 ต้นฉบับแบบสอบถาม SUS.....	18
2.5 ภาพรวมสถาปัตยกรรมและส่วนประกอบหลักของ MobileNetV3.....	21
2.6 MobileNet V3 Block.....	21
2.7 เปรียบเทียบ Accuracy ต่อ MAdds ต่อขนาดโมเดล.....	22
2.8 การเปรียบเทียบประสิทธิภาพการประมวลผลศึนิยมบนอุปกรณ์พกพา	22
2.9 Pretrained Models on ImageNet-21-K-P Dataset.....	24
3.1 ภาพรวมกระบวนการวิจัย	31
3.2 ภาพรวมการไหลของระบบ (System Flow).....	36
4.1 รูปตัวอย่างไฟล์ฐานข้อมูลพจนานุกรม 3 ภาษา (TH-JP-EN).....	49
4.2 หน้าจอหลักการถ่ายภาพ.....	50
4.3 ตัวอย่างหน้าจอแสดงผลการทำนายและอ่านออกเสียง 3 ครั้ง.....	51
4.4 รูปภาพตัวอย่างจากแกลเลอรี	52
4.5 หน้าจอการเลือกรูปภาพและทำนายรูปภาพจากแกลเลอรี	52
4.6 หน้าจอถ่ายภาพทดสอบจริงกรณี S1	53
4.7 หน้าจอถ่ายภาพทดสอบจริงกรณี C1.....	54
4.8 หน้าจอถ่ายภาพทดสอบจริงกรณี C2.....	55
4.9 หน้าจอถ่ายภาพทดสอบจริงกรณี C3.....	56
4.10 หน้าจอถ่ายภาพทดสอบจริงกรณี C4.....	57
4.11 หน้าจอถ่ายภาพทดสอบจริงกรณี C5.....	58

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ผู้พิการทางสายตาถือเป็นกลุ่มประชากรที่ประสบกับข้อจำกัดในการดำเนินชีวิตประจำวัน โดยเฉพาะการรับรู้สิ่งต่างๆ รอบตัวซึ่งอาศัยการมองเห็นเป็นหลัก แม้จะสามารถใช้ประสาทสัมผัสอื่น เช่น การฟังและการสัมผัส มาทดแทนได้บางส่วน แต่ก็ยังคงมีข้อจำกัดที่ทำให้การใช้ชีวิตไม่สะดวกสบาย เทียบเท่ากับบุคคลทั่วไป ดังนั้น การพัฒนาเทคโนโลยีเพื่อสนับสนุนการใช้ชีวิตของผู้พิการทางสายตาจึงมีความสำคัญต่อการส่งเสริมโอกาสทางการศึกษา การประกอบอาชีพ และการใช้ชีวิตประจำวันอย่างอิสระ ข้อมูลสถิติจากกรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ [1] ณ วันที่ 31 สิงหาคม พ.ศ.2568 ระบุว่า สถานการณ์ผู้พิการในประเทศไทย มีผู้พิการทางการมองเห็น 171,953 คน หรือคิดเป็น 7.54% ของจำนวนผู้พิการในประเทศไทยทั้งหมด 2,280,409 คน



รูปที่ 1.1 สถานการณ์ผู้พิการในประเทศไทย [1]

องค์การอนามัยโลก (WHO) [2] กล่าวว่า ภาวะการมองเห็น ซึ่งเป็นประสาทสัมผัสที่สำคัญที่สุดของเรา มีบทบาทสำคัญในทุกแง่มุมและทุกช่วงของชีวิต หากขาดการมองเห็นแล้ว เราจะประสบปัญหาในการเรียนรู้ การเดิน การอ่าน การเข้าร่วมกิจกรรมในโรงเรียน และการทำงาน ซึ่งภาวะสายตาผิดปกติเกิดขึ้นเมื่อมีความผิดปกติของดวงตาที่ส่งผลต่อระบบการมองเห็นและการทำงานของสายตาและภาวะ

สายตาผิดปกติมีผลกระทบร้ายแรงต่อบุคคลตลอดช่วงชีวิต การสนับสนุนบุคคลที่ประสบปัญหาทางด้านกายภาพ เช่น การมองเห็น จึงเป็นสิ่งที่ควรคำนึงถึงเช่นกันในฐานะของเพื่อนมนุษย์

ในยุคปัจจุบันเทคโนโลยีสารสนเทศและการสื่อสารมีบทบาทสำคัญต่อการดำรงชีวิตประจำวันของมนุษย์ โดยเฉพาะอย่างยิ่งเทคโนโลยีสมาร์ทโฟนซึ่งได้รับการพัฒนาอย่างต่อเนื่องทั้งในด้านฮาร์ดแวร์และซอฟต์แวร์ ส่งผลให้กลายเป็นอุปกรณ์อเนกประสงค์ที่สามารถตอบสนองความต้องการได้ในหลายมิติ ไม่เพียงเพื่อการสื่อสาร แต่ยังรวมถึงการศึกษา ความบันเทิง การแพทย์ จึงทำให้การใช้ชีวิตสะดวกสบายมากยิ่งขึ้น

ในอีกด้านหนึ่ง ผู้พิการทางสายตาซึ่งเป็นบุคคลที่เผชิญต่อข้อจำกัดในการรับรู้สิ่งแวดล้อมรอบตัว ซึ่งสัมพันธ์โดยตรงกับความปลอดภัยและความเป็นอิสระในการใช้ชีวิต เทคโนโลยีในปัจจุบันนี้ ได้เข้ามามีบทบาทสำคัญ เช่น สมาร์ทโฟนมีกล้องคุณภาพดีและหน่วยประมวลผลบนอุปกรณ์เพียงพอสำหรับทำงานด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) ทำให้มีศักยภาพซึ่งสามารถเป็นดวงตาให้กับผู้ใช้งาน

A. Quatraro and M. Paiano [3] กล่าวว่า การเรียนรู้ภาษาต่างประเทศเป็นสิ่งสำคัญสำหรับทุกคน โดยเฉพาะอย่างยิ่งหากภาษาหลักของคุณไม่ได้ใช้กันอย่างแพร่หลาย และเมื่อพิจารณาจากความต้องการทักษะการสื่อสารในสังคมปัจจุบัน รวมถึงทางอินเทอร์เน็ต การเรียนรู้ภาษาเป็นสิ่งสำคัญยิ่งขึ้นสำหรับผู้พิการทางสายตา เพื่อช่วยลดช่องว่างที่เกิดขึ้นจากความบกพร่องทางการมองเห็น และการมีทักษะในภาษาต่างประเทศ โดยเฉพาะภาษาอังกฤษ จึงกลายเป็นเสมือนสะพานที่ช่วยให้ผู้พิการทางสายตาสามารถเข้าถึงวัฒนธรรมที่หลากหลายและโอกาสในด้านการติดต่อทางสังคมได้

ในยุคที่เทคโนโลยีดิจิทัลและปัญญาประดิษฐ์เข้ามามีบทบาทสำคัญในการเรียนรู้ภาษา แอปพลิเคชันบนอุปกรณ์พกพาได้กลายเป็นเครื่องมือในการพัฒนาทักษะภาษาอังกฤษและภาษาต่างประเทศสำหรับผู้เรียนทั่วไป ในกรณีของผู้พิการทางสายตาจะสามารถเรียนรู้คำศัพท์จากสิ่งแวดล้อมรอบตัวผ่านเทคโนโลยีรู้จำวัตถุ (Object Recognition) ร่วมกับเสียงสังเคราะห์ (Text-to-Speech : TTS) แบบออฟไลน์ ซึ่งจะถูกพัฒนาขึ้นในฐานะแอปพลิเคชันต้นแบบในงานวิจัยนี้

จากการที่ผู้วิจัยทำงานในฐานะล่ามภาษาญี่ปุ่นมากกว่า 15 ปี จึงมีแนวคิดในการบูรณาการองค์ความรู้ด้านภาษาเข้ากับเทคโนโลยีสารสนเทศโดยมุ่งเน้นในการพัฒนาแอปพลิเคชันต้นแบบเพื่อช่วยให้ผู้พิการทางสายตาสามารถใช้ในการเรียนรู้คำภาษาต่างประเทศ โดยการใช้เทคโนโลยีที่มีอยู่แต่เดิม เช่น Flutter, MobileNet, เทคโนโลยีแปลงข้อความเป็นเสียง (Text-to-Speech : TTS) รวมถึงการประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ (Generative AI) และเนื่องจากภาษาหลักในการสื่อสารสากลคือภาษาอังกฤษ จึงได้ขยายขอบเขตของงานโดยเพิ่มภาษาอังกฤษเข้ามาร่วมด้วย เพื่อเป็นประโยชน์สำหรับการเรียนรู้ในวงกว้าง โดยงานวิจัยนี้จะพัฒนาเป็นแอปพลิเคชันต้นแบบรู้จำวัตถุด้วยสมาร์ทโฟนเพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) เพื่อเป็นก้าวแรกในการเรียนรู้ภาษาต่างประเทศสำหรับผู้พิการทางสายตา

อีกทั้ง จากการที่รัฐบาลญี่ปุ่นยินดีให้การสนับสนุนโครงการที่ส่งเสริมและพัฒนาความมั่นคงของมนุษย์โดยผ่านความร่วมมือทางด้านเศรษฐกิจแบบให้เปล่าเพื่อพื้นฐานและความมั่นคงของมนุษย์ (The Grant Assistance for Grassroots Human Security Projects : GGP) โดยมุ่งหวังให้คุณภาพการศึกษาสำหรับผู้พิการทางสายตาดีขึ้น ด้วยเหตุนี้ การพัฒนาแอปพลิเคชันนี้จึงเป็นสิ่งที่สอดคล้องต่อการสนับสนุนแนวคิดดังกล่าว [4]

แอปพลิเคชันในงานวิจัยนี้ใช้ชื่อว่า “VocabEyes” สำคัญคือ การออกแบบประสบการณ์ใช้งานที่เรียบง่าย สามารถใช้งานด้วยมือเดียว ทำงานแบบออฟไลน์เพื่อความเร็วและความเป็นส่วนตัว ซึ่งสามารถใช้งานในสมาร์ตโฟนเครื่องทดลองที่กำหนด โดยงานวิจัยนี้จึงสะท้อนเป้าหมายในด้านการศึกษา (Education) และการเข้าถึง (Accessibility) สำหรับผู้พิการทางสายตา

1.2 วัตถุประสงค์ของการวิจัย

การพัฒนาแอปพลิเคชันต้นแบบบนสมาร์ตโฟนระบบปฏิบัติการแอนดรอยด์เครื่องทดลองที่กำหนดสำหรับการรู้จำวัตถุและอ่านออกเสียงคำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) เพื่อเป็นประโยชน์ด้านการศึกษาภาษาต่างประเทศเบื้องต้นสำหรับผู้พิการทางสายตา

1.3 ขอบเขตของการวิจัย

1.3.1 ขอบเขตด้านเนื้อหา (Content Scope)

1. ข้อมูลและคำศัพท์ จัดทำชุดคำศัพท์ตามรายการชุดข้อมูล ImageNet ที่อยู่ในโมเดลจำแนกภาพสำเร็จรูปโดยเป็นสิ่งของและสิ่งมีชีวิตทั่วไปในชีวิตประจำวัน เช่น เครื่องใช้ไฟฟ้า เฟอร์นิเจอร์ และสัตว์เลี้ยง และอ้างอิงคำแปลคำศัพท์จาก WordNet ในแต่ละภาษา โดยมีคลังคำศัพท์ประมาณ 6,000-10,000 คำ ประกอบด้วยภาษาไทย 8,500 คำ ภาษาญี่ปุ่น 6,000 คำ ภาษาอังกฤษ 10,000 คำ โดยประมาณ

2. การนำเสนอผลลัพธ์ ข้อมูลที่ได้จากการตรวจจับจะถูกแสดงผลในรูปแบบเสียงเพื่อระบุวัตถุ โดยมีการออกเสียงภาษาไทย ภาษาญี่ปุ่น และภาษาอังกฤษตามลำดับเพื่อช่วยให้ผู้พิการทางสายตาสามารถเรียนรู้คำศัพท์จากสิ่งรอบตัว และ/หรือมีการแสดงข้อความที่หน้าจอสำหรับผู้ใช้งานปกติ

3. การออกแบบเพื่อการเข้าถึง (Accessibility) แอปพลิเคชันถูกออกแบบให้เรียบง่ายเหมาะสมกับการใช้งานของผู้พิการทางสายตา ลดความซับซ้อนในส่วนติดต่อผู้ใช้ให้มากที่สุด

4. ข้อจำกัด การวิจัยนี้ใช้งาน MobileNet V3 ซึ่งเป็นโมเดลที่ฝึกสำเร็จรูปบนชุดข้อมูล Image Net-21K-P (11,221 Classes) เพื่อนำมาใช้ในแอปพลิเคชันรู้จำวัตถุซึ่งมีคำศัพท์ในคลังประมาณ 6,000-10,000 คำ (ภาษาไทย 8,500 คำ ภาษาญี่ปุ่น 6,000 คำ ภาษาอังกฤษ 10,000 คำ โดยประมาณ)

ไม่ครอบคลุมการแปลความหมายเชิงประโยคหรือเชิงบริบท โดยเป็นการอ่านออกเสียงคำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) เท่านั้น

1.3.2 ขอบเขตด้านผู้ใช้ (User Scope)

1. กลุ่มผู้ใช้งานที่เป็นผู้พิการทางสายตา ทั้งคนตาบอด (Blind) และมองเห็นเลือนราง (Low Vision) ซึ่งสามารถใช้งานสมาร์ทโฟนเบื้องต้นได้
2. บุคคลทั่วไปที่ต้องการเรียนรู้คำศัพท์ภาษาต่างประเทศเบื้องต้น โดยอาจเป็นได้ทั้งชาวไทย ชาวญี่ปุ่น หรือชาวต่างชาติที่สามารถใช้ภาษาอังกฤษในการสื่อสาร

1.3.3 ขอบเขตด้านเทคโนโลยี (Technology Scope)

1. ใช้กล้องสมาร์ทโฟนระบบ Android ในการถ่ายภาพนิ่ง
2. ใช้โมเดลสำเร็จรูป สำหรับการรู้จำและระบุสิ่งของในภาพเป็นหลัก
3. ใช้ระบบ Text-to-Speech (TTS) ที่รองรับการอ่านออกเสียงคำศัพท์ทั้งภาษาไทย ภาษาญี่ปุ่น และภาษาอังกฤษ ซึ่งการอ่านออกเสียงอาจไม่ตรงตามสำเนียงที่ถูกต้องบางส่วน
4. ไม่ครอบคลุมการประมวลผลแบบ Real-time Video Detection อย่างต่อเนื่อง และจำกัดเฉพาะ การถ่ายภาพนิ่ง (Image Capture) หรือการเลือกภาพนิ่งจากแกลเลอรีภาพถ่าย และตรวจจับวัตถุจากภาพนั้น

1.3.4 ขอบเขตด้านการประเมินผล (Evaluation Scope)

1. ประเมินความถูกต้องแม่นยำ (Accuracy) ในการรู้จำวัตถุของแอปพลิเคชันต้นแบบ “VocabEyes” จากการทดสอบใช้งานจริง
2. ประเมินความสามารถในการใช้งานและความพึงพอใจต่อแอปพลิเคชันต้นแบบ โดยใช้อาสาสมัครสายตาปกติที่จำลองปิดกั้นการมองเห็น (Blindfolded Normal-Vision Volunteers) ตั้งแต่ 5 คนขึ้นไป ผ่านการสัมภาษณ์หรือแบบสอบถาม

กล่าวโดยสรุปคือเป็นการพัฒนาแอปบนสมาร์ทโฟนในระบบปฏิบัติการ Android เพื่อจำแนกวัตถุจากภาพนิ่งด้วยโมเดลสำเร็จรูป MobileNetV3 และรายงานผลด้วยเสียงสามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) สำหรับผู้พิการทางสายตาเป็นหลัก โดยใช้ชุดคำศัพท์ในชีวิตประจำวัน 6,000-10,000 คำ (ภาษาไทย 8,500 คำ ภาษาญี่ปุ่น 6,000 คำ ภาษาอังกฤษ 10,000 คำ โดยประมาณ) ตามรายการในชุดข้อมูลของ ImageNet 21K-P (11,221 คลาส) และออกแบบแอปพลิเคชันอย่างเรียบง่าย เพื่อให้เหมาะสมสำหรับผู้พิการทางสายตา พร้อมทั้งประเมินผ่านการใช้งานจริง รวมถึงประเมินความพึงพอใจของผู้ใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ผู้พิการทางสายตา หรือบุคคลทั่วไปได้รับแอปพลิเคชันต้นแบบที่สามารถเรียนรู้ คำศัพท์ ภาษาไทย-ญี่ปุ่น-อังกฤษในชีวิตประจำวันเบื้องต้นจากสิ่งแวดล้อมรอบตัว
2. สามารถนำแอปพลิเคชันต้นแบบ ไปต่อยอดเพื่อพัฒนาความแม่นยำของระบบให้ดียิ่งขึ้นต่อไปในอนาคต
3. สามารถนำชุดพจนานุกรม EN TH JA (ภาษาอังกฤษ ภาษาไทย ภาษาญี่ปุ่น) จำนวนประมาณ 6,000-10,000 คำ ตามรายการในชุดข้อมูล ImageNet 21K-P (11,221 คลาส) ซึ่งเป็นการจับคู่กับรายการ คำศัพท์ของ WordNet (ภาษาไทย 8,500 คำ ภาษาญี่ปุ่น 6,000 คำ ภาษาอังกฤษ 10,000 คำ โดยประมาณ) ไปใช้ซ้ำ หรือขยายผลต่อยอดในอนาคตได้

1.5 นิยามศัพท์เฉพาะ

1.5.1 **ผู้พิการทางสายตา (The Visually Impaired)** คือ ผู้ที่มีข้อจำกัดรุนแรงด้านการมองเห็น จนส่งผลกระทบต่อชีวิตประจำวัน การสูญเสียการมองเห็นเป็นภาวะที่การมองเห็นนั้นไม่สามารถแก้ไขได้ด้วยแว่นตา เลนส์สัมผัส หรือการรักษาทางการแพทย์จนกลับมามีการมองเห็นปกติได้ จากประกาศกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ [5] เรื่องประเภทและหลักเกณฑ์ความพิการ พ.ศ.2552 ระบุถึงหลักเกณฑ์กำหนดความพิการทางการเห็น ได้แก่

1) **ตาบอด** หมายถึง การที่บุคคลมีข้อจำกัดในการปฏิบัติกิจกรรมในชีวิตประจำวัน หรือการเข้าไปมีส่วนร่วมในกิจกรรมทางสังคม ซึ่งเป็นผลมาจากการมีความบกพร่องในการเห็น เมื่อตรวจวัดการเห็นของสายตาข้างที่ตีกวเมื่อใช้แว่นสายตาธรรมดาแล้ว อยู่ในระดับต่ำกว่า 3 ส่วน 60 เมตร (3/60) หรือ 20 ส่วน 400 ฟุต (20/400) ลงมาจนกระทั่งมองไม่เห็นแม้แต่แสงสว่าง หรือมีลานสายตาแคบกว่า 10 องศา

2) **ตาเห็นเลือนราง** หมายถึง การที่บุคคลมีข้อจำกัดในการปฏิบัติกิจกรรมในชีวิตประจำวัน หรือการเข้าไปมีส่วนร่วมในกิจกรรมทางสังคม ซึ่งเป็นผลมาจากการมีความบกพร่องในการเห็น เมื่อตรวจวัดการเห็นของสายตาข้างที่ตีกว เมื่อใช้แว่นสายตาธรรมดาแล้ว อยู่ในระดับตั้งแต่ 3 ส่วน 60 เมตร (3/60) หรือ 20 ส่วน 400 ฟุต (20/400) ไปจนถึงต่ำกว่า 6 ส่วน 18 เมตร (6/18) หรือ 20 ส่วน 70 ฟุต (20/70) หรือมีลานสายตาแคบกว่า 30 องศา

1.5.2 **Accessibility (การเข้าถึงได้)** สำหรับผู้พิการทางสายตา หมายถึง การออกแบบและพัฒนาสิ่งต่างๆ ไม่ว่าจะเป็น สภาพแวดล้อม เทคโนโลยี แอปพลิเคชัน เว็บไซต์ หรือสื่อสารสนเทศ ให้ผู้พิการทางสายตาสามารถใช้งานได้สะดวกเท่าเทียมกับคนทั่วไป

1.5.3 ปัญญาประดิษฐ์เชิงสร้างสรรค์ (Generative Artificial Intelligence: Gen AI)

คือ ปัญญาประดิษฐ์ประเภทหนึ่งที่ใช้อัลกอริทึมในการเรียนรู้ชุดข้อมูลต่างๆ เพื่อสร้างผลลัพธ์ใหม่ๆ ที่คล้ายคลึงกันแต่แตกต่างจากข้อมูลต้นฉบับ เช่น รูปภาพ เพลง ข้อความ หรือสื่อในรูปแบบอื่นๆ เป็นเทคโนโลยีที่มีการพัฒนาอย่างรวดเร็วซึ่งกำลังเป็นสิ่งที่ปฏิวัติวงการอุตสาหกรรมสร้างสรรค์ โดย Generative AI ได้กลายเป็นเครื่องมือสำหรับนักสร้างสรรค์และนักออกแบบในแวดวงอุตสาหกรรมต่างๆ ซึ่งช่วยพัฒนาแนวคิดและสร้างสรรค์สิ่งใหม่ๆ ด้วยกระบวนการสร้างสรรค์ที่เป็นไปอย่างอัตโนมัติ จึงช่วยประหยัดเวลาและเกิดประสิทธิภาพในการทำงานมากขึ้น สิ่งสำคัญคือผู้ใช้จะต้องมีความเข้าใจในขั้นตอนการทำงานของเทคโนโลยีและการนำเทคโนโลยีนี้ไปใช้ในกระบวนการออกแบบได้อย่างเหมาะสมและเกิดประโยชน์สูงสุด [6]



บทที่ 2

ทฤษฎี หลักการ และงานวิจัยที่เกี่ยวข้อง

ในบทนี้ผู้วิจัยได้ทำการรวบรวมองค์ความรู้ ทฤษฎี หลักการพื้นฐาน และงานวิจัยที่เกี่ยวข้อง เพื่อนำมาใช้เป็นข้อมูล และแนวทางสำหรับการทำวิจัย โดยมีสาระสำคัญดังนี้

2.1 ภาพรวมของการทบทวนวรรณกรรม

นำเสนอการทบทวนวรรณกรรมและองค์ความรู้ที่เกี่ยวข้องกับแอปพลิเคชันที่อาศัยการจำแนกภาพจากกล้องบนอุปกรณ์พกพา และการแปลงผลการทำนายให้เป็นข้อมูลคำศัพท์เพื่อสื่อสารแก่ผู้ใช้งานด้วยเสียง โดยมีเป้าหมายเพื่อวางรากฐานเชิงทฤษฎีสำหรับการออกแบบและพัฒนาแอปพลิเคชันในบทถัดไป เนื้อหาในบทนี้จึงมุ่งอธิบาย แนวคิด หลักการ วิธีการ และงานวิจัยของผู้อื่นที่เป็นกรอบอ้างอิงทั้งในมิติของแบบจำลองการจำแนกภาพที่เหมาะสมกับอุปกรณ์พกพา ทรัพยากรคำศัพท์เชิงความหมายแบบ WordNet สำหรับคลังคำศัพท์หลายภาษา ตลอดจนเทคโนโลยีการสังเคราะห์เสียง และการประเมินผลของแอปพลิเคชัน รวมถึงความสามารถในการใช้งาน

2.2 แนวคิดพื้นฐานของการรู้จำวัตถุด้วยการจำแนกภาพ

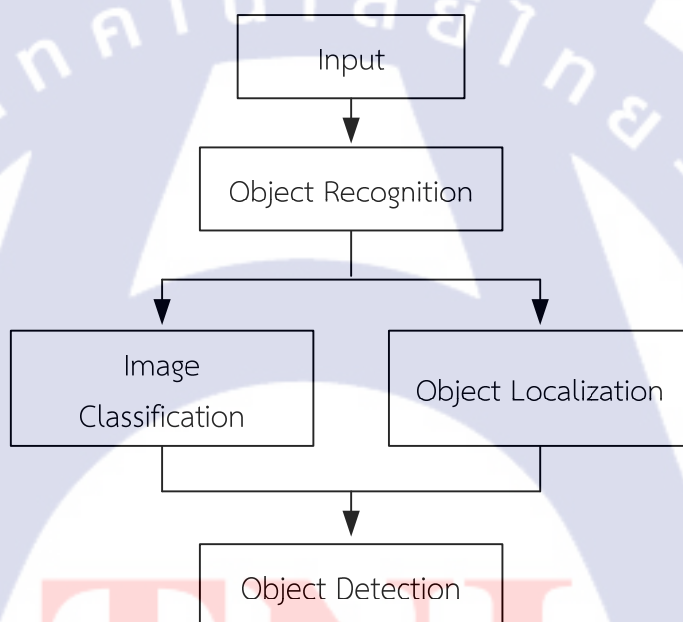
แนวคิดพื้นฐานของการรู้จำวัตถุจากภาพในมุมมองของการจำแนกภาพ (Image Classification) เป็นแนวทางสำคัญในงานคอมพิวเตอร์วิทัศน์ที่มุ่งทำนายว่า วัตถุหลักในภาพ จัดอยู่ในกลุ่มประเภทใด (Class) พร้อมค่าความเชื่อมั่น (Confidence/Probability) เพื่อนำผลลัพธ์ไปใช้งานประยุกต์ต่างๆ บนระบบจริง เช่น บนอุปกรณ์พกพาที่ต้องคำนึงถึงข้อจำกัดด้านเวลาในการประมวลผลและทรัพยากรของเครื่อง

2.2.1 การตรวจจับวัตถุ (Object Detection)

กระบวนการหาตำแหน่ง (Localization) และการจำแนกประเภท (Classification) ของสิ่งที่อยู่ในภาพ โดยทั่วไปจะใช้โมเดลเชิงลึก (Deep Learning) ที่เรียนรู้จากชุดข้อมูลขนาดใหญ่ เช่น ImageNet หรือ COCO Dataset

IBM [7] กล่าวถึง การตรวจจับวัตถุ (Object Detection) ไว้ว่า เป็นเทคนิคที่ใช้เครือข่ายประสาทเทียม (Neural networks) เพื่อระบุตำแหน่งและจำแนกประเภทของวัตถุในภาพ งานคอมพิวเตอร์วิทัศน์ (Computer Vision) นี้มีการใช้งานอย่างหลากหลาย ตั้งแต่การถ่ายภาพทางการแพทย์ไปจนถึงรถยนต์ไร้คนขับ

การตรวจจับวัตถุเป็นงานทางคอมพิวเตอร์วิทัศน์ (Computer Vision) ที่มีเป้าหมายเพื่อระบุตำแหน่งของวัตถุในภาพดิจิทัลเป็นตัวอย่างหนึ่งของปัญญาประดิษฐ์ ที่ประกอบด้วยการฝึกคอมพิวเตอร์ให้มองเห็นเหมือนมนุษย์ โดยเฉพาะอย่างยิ่งการรู้จำและจำแนกประเภทของวัตถุตามกลุ่มความหมาย (Semantic Categories) การระบุตำแหน่งของวัตถุ (Object Localization) เป็นเทคนิคที่ใช้ระบุตำแหน่งของวัตถุเฉพาะในภาพโดยการกำหนดขอบเขตของวัตถุด้วยกรอบสี่เหลี่ยม (Bounding Box) ส่วนการจำแนกประเภทของวัตถุ (Object classification) เป็นเทคนิคที่ใช้ระบุว่าวัตถุที่ตรวจพบอยู่ในกลุ่มประเภทใด งานตรวจจับวัตถุจึงเป็นการผสมผสานระหว่างการระบุตำแหน่งและการจำแนกประเภทของวัตถุ เพื่อตรวจสอบตำแหน่งและประเภทของวัตถุในภาพเดียวหรือหลายภาพพร้อมกัน โดยสามารถสรุปลำดับการตรวจจับวัตถุโดยสังเขปได้ดังรูปที่ 2.1



รูปที่ 2.1 แผนภาพแสดงลำดับการตรวจจับวัตถุ

2.2.2 การรู้จำวัตถุ (Object Recognition)

กระบวนการที่เกี่ยวข้องกับการระบุและจัดหมวดหมู่วัตถุในภาพหรือวิดีโอ โดยการใช้คุณลักษณะทางสายตา เช่น รูปร่าง สี และพื้นผิว เพื่อจำแนกวัตถุเป็นประเภทต่างๆ เว็บไซต์ของ Sciencedirect [8] กล่าวถึงเรื่องนี้ว่า การรู้จำวัตถุ (Object Recognition) คือ กระบวนการระบุวัตถุโดยการให้ป้ายกำกับหรือชื่อเรียกตามข้อมูลทางสายตา กระบวนการนี้เกี่ยวข้องกับการจัดประเภทวัตถุให้เป็นกลุ่มย่อยหรือกลุ่มหลัก และยังสามารถขยายไปถึงการรู้จำการกระทำหรือกิจกรรมต่างๆ โดยใช้แบบจำลองการเรียนรู้หลากหลายรูปแบบในงานคอมพิวเตอร์วิทัศน์ (Computer Vision)

2.2.3 การจำแนกภาพ (Image Classification)

การจำแนกภาพเป็นการกำหนดป้ายกำกับ (Label) ให้กับภาพทั้งภาพ เช่น แมว สุนัข โดยโมเดลที่นิยม เช่น Convolutional Neural Network (CNN) หรือ [9] รุ่นย่อยที่มีประสิทธิภาพ อย่างเช่น MobileNets สำหรับแอปพลิเคชันที่เกี่ยวข้องกับระบบวิทัศน์ซึ่งถูกสร้างขึ้นโดยเน้นที่ประสิทธิภาพและความเร็วในการประมวลผลสำหรับใช้งานในสภาพแวดล้อมที่มีทรัพยากรจำกัด เช่น โทรศัพท์มือถือ

2.2.4 การรู้จำวัตถุด้วยการจำแนกภาพ

การจำแนกภาพ (Image Classification) คือกระบวนการที่ระบบรับภาพอินพุต แล้วทำนายป้ายกำกับ (Label) หรือคลาส (Class) ที่สอดคล้องกับเนื้อหาในภาพ โดยทั่วไปแบบจำลองจะให้ผลลัพธ์เป็นคะแนนของแต่ละคลาส (Class Score) หรือค่าความน่าจะเป็นหลังผ่านฟังก์ชันเชิงสถิติ (เช่น Softmax) เพื่อสะท้อนความเชื่อมั่นของการทำนายในแต่ละคลาส ผลลัพธ์ดังกล่าวมักถูกนำเสนอในรูปแบบ Top-1 (คลาสอันดับ 1) หรือ Top-k (คลาสอันดับต้นๆ หลายอันดับ) เพื่อรองรับกรณีที่ภาพมีความกำกวม วัตถุมีลักษณะใกล้เคียงกัน หรือหากหลังรับทราจจนทำให้การทำนายอันดับแรกเพียงอย่างเดียวอาจไม่เพียงพอ

ในเชิงการใช้งานจริง การรู้จำวัตถุด้วยการจำแนกภาพมักตั้งอยู่บนสมมติฐานว่า ภาพหนึ่งภาพมีวัตถุเด่นหลัก (Dominant Object) หรืออย่างน้อยมีวัตถุสำคัญที่ควรเป็นคำตอบหลักของระบบ แม้ในความเป็นจริงภาพอาจประกอบด้วยวัตถุหลายชนิด ระบบจำแนกภาพจึงมักใช้แนวทางสรุปผลเป็นคลาสเดียวหรือหลายคลาสอันดับต้นๆ ตามระดับความเชื่อมั่น การใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network : CNN) ที่ฝึกบนชุดข้อมูลขนาดใหญ่สามารถยกระดับความแม่นยำของการจำแนกภาพได้อย่างมีนัยสำคัญ โดยหนึ่งในชุดข้อมูลมาตรฐานที่มีบทบาทสูงต่อการพัฒนาโมเดลจำแนกภาพคือ ImageNet

นอกจากนี้ ปัจจัยจากโลกจริง เช่น ความสว่างไม่เพียงพอ ภาพสั่นไหว (Motion Blur) หลุดโฟกัส (Out-of-Focus) มุมมองที่แปลกไปจากข้อมูลฝึกสอน หรือมีการบดบังบางส่วน (Occlusion) ล้วนส่งผลให้ความเชื่อมั่นลดลงและเพิ่มโอกาสการทำนายผิดพลาด ดังนั้น การทำความเข้าใจความหมายของคะแนนและการรายงานผลแบบ Top-k จึงมีความสำคัญต่อการออกแบบวิธีนำเสนอผลและการตัดสินใจของระบบ

คำสำคัญที่ใช้ในบริบทนี้ประกอบด้วย (1) Class/Label หมายถึงหมวดหมู่ผลลัพธ์ที่แบบจำลองทำนาย (2) Confidence/Probability หมายถึงค่าที่สะท้อนความเชื่อมั่นของแบบจำลองต่อคลาสนั้นๆ และ (3) Top-k หมายถึงการเลือกผลลัพธ์อันดับต้นๆ จำนวน k รายการเพื่อเพิ่มความยืดหยุ่นเมื่อผลลัพธ์อันดับแรกมีความไม่แน่นอน [10]

2.2.5 กระบวนการทำงานทั่วไปของระบบรู้จำวัตถุจากกล้อง (Conceptual Pipeline)

ระบบรู้จำวัตถุจากกล้องในงานประยุกต์มักประกอบด้วยขั้นตอนหลักที่เชื่อมต่อกันเป็นลำดับ (Pipeline) เพื่อแปลงภาพจากโลกจริง ไปสู่ผลลัพธ์ที่พร้อมสื่อสารกับผู้ใช้ สรุปแนวคิดเป็น 5 ขั้นตอนสำคัญดังนี้ [11]

1. การรับภาพ (Image Acquisition) เริ่มจากการรับภาพจากกล้องหรือแหล่งภาพ โดยคุณภาพของภาพที่ได้ (แสง ความคมชัด การสั่น) มีผลต่อความแม่นยำของขั้นตอนถัดไปอย่างมาก
2. การเตรียมข้อมูลภาพ (Pre-processing) เป็นการปรับรูปแบบภาพให้สอดคล้องกับอินพุตที่แบบจำลองต้องการ เช่น การปรับขนาด (Resize) การครอบ/เติมขอบ (Crop/Pad) การแปลงค่าสีหรือการทำ Normalization เพื่อกระจายของค่าพิกเซลอยู่ในช่วงที่เหมาะสมกับแบบจำลอง การเตรียมข้อมูลที่เหมาะสมช่วยลดปัญหาความต่างของโดเมนระหว่างภาพใช้งานจริงกับภาพที่ใช้ฝึกสอน
3. การทำนายของแบบจำลอง (Model Inference) เป็นขั้นตอนที่แบบจำลองคำนวณคะแนนของแต่ละคลาสจากภาพอินพุตโดยอาศัยแนวคิดการสกัดคุณลักษณะด้วยโครงข่ายคอนโวลูชันซึ่งเป็นรากฐานของงานจำแนกภาพสมัยใหม่
4. การประมวลผลหลังการทำนาย (Post-processing) รวมถึงการแปลงคะแนนเป็นความน่าจะเป็น การคัดเลือกผลลัพธ์แบบ Top-1 หรือ Top-k การใช้เกณฑ์ความเชื่อมั่นขั้นต่ำ (Threshold) ในกรณีต้องการคัดกรองผลลัพธ์ที่ไม่มั่นใจและการจัดรูปแบบคำตอบให้อยู่ในรูปแบบที่เหมาะสมต่อการสื่อสารหรือการนำไปใช้ต่อ
5. การนำเสนอผล (Result Presentation) เป็นขั้นตอนการส่งมอบผลลัพธ์ให้ผู้ใช้ในรูปแบบที่เข้าใจง่าย เช่น แสดงข้อความ สัญลักษณ์ หรือการสื่อสารด้วยเสียง โดยหลักการสำคัญคือการออกแบบให้การนำเสนอผลสอดคล้องกับความไม่แน่นอนของแบบจำลอง เช่น การรายงานผลลัพธ์อันดับต้นๆ เป็นต้น

2.3 พื้นฐานที่เกี่ยวข้องกับการจำแนกภาพ

2.3.1 การเรียนรู้เชิงลึก (Deep Learning)

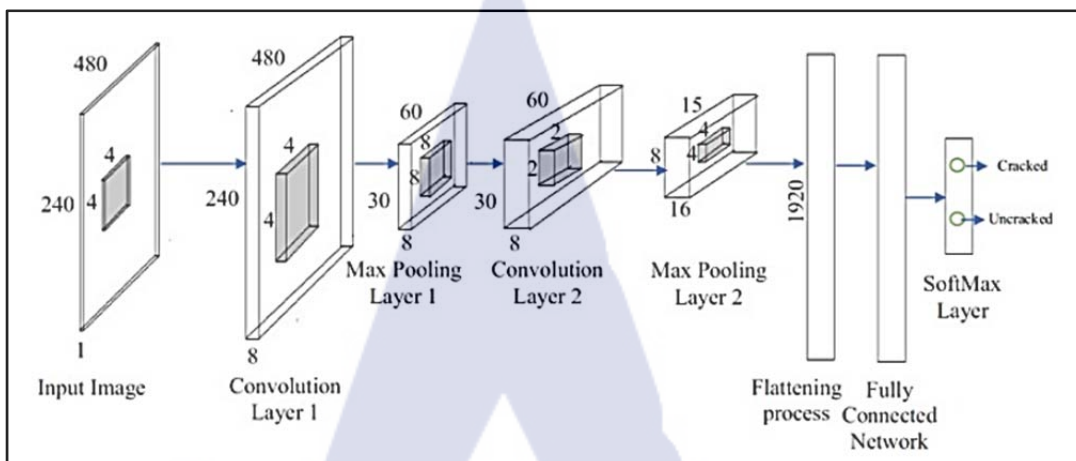
เป็นวิธีการเรียนรู้ข้อมูลแบบอัตโนมัติที่ได้รับแรงบันดาลใจจากการเลียนแบบการทำงานของโครงข่ายประสาทของสิ่งมีชีวิต (Neurons) โดยเป็นการนำระบบโครงข่ายประสาทเทียม (Artificial Neural Network) มาเชื่อมต่อซ้อนทับกันจำนวนหลายชั้น (Layer) เพื่อทำการเรียนรู้จากข้อมูลตัวอย่าง ซึ่งข้อมูลดังกล่าวจะถูกนำไปใช้ประโยชน์ในการตรวจจำรูปแบบ (Pattern) หรือการจำแนกข้อมูล (Classification) ได้อย่างมีประสิทธิภาพ

ในอดีตโครงข่ายประสาทเทียมแบบดั้งเดิมมักจะมีจำนวนชั้นซ่อน (Hidden Layer) เพียงไม่กี่ชั้น เนื่องจากข้อจำกัดด้านปริมาณข้อมูลสำหรับฝึกสอน (Training Data) รวมถึงขีดความสามารถของคอมพิวเตอร์ที่ยังไม่สูงพอต่อการประมวลผลที่ซับซ้อน อย่างไรก็ตาม เมื่อเทคโนโลยีได้รับการพัฒนาจนคอมพิวเตอร์มีประสิทธิภาพสูงขึ้นในปัจจุบัน จึงส่งผลให้การสร้างชั้นโครงข่ายทำได้ง่าย และมีจำนวนชั้นที่มากขึ้น เพื่อให้ระบบสามารถคิดและประมวลผลสิ่งที่ซับซ้อนได้ใกล้เคียงกับสมองของมนุษย์ผ่านการส่งต่อข้อมูลและประมวลผลร่วมกันระหว่างชั้น ซึ่งความซับซ้อนที่เพิ่มขึ้นตามจำนวนชั้นของโครงข่ายที่ซ้อนทับกันนี้เอง จึงเป็นที่มาของคำว่าการเรียนรู้เชิงลึก [12]

2.3.2 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network : CNN)

โครงข่ายประสาทเทียมชนิดหนึ่งที่ถูกออกแบบมาโดยเฉพาะสำหรับประมวลผลข้อมูลที่มีโครงสร้างคล้ายตาราง เช่น ข้อมูลภาพ โดย CNN จะใช้ชั้นต่างๆ ในการดึงคุณลักษณะ (Feature Extraction) และใช้วิธีการแบบลำดับชั้น (Hierarchical Approach) เพื่อจับรูปแบบที่ซับซ้อนมากขึ้นจากข้อมูลนำเข้าอย่างต่อเนื่อง

J. Singh and S. Singh [13] ระบุว่า โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) เป็นประเภทหนึ่งของโครงข่ายประสาทเทียมที่สามารถสร้างการเชื่อมโยงที่มีความหมายเมื่อได้รับข้อมูลภาพเป็นอินพุต โครงข่ายนี้มีลักษณะคล้ายกับสมองมนุษย์ตรงที่มีหลายชั้น ซึ่งเลียนแบบการเชื่อมต่อของระบบประสาท ฟังก์ชันกระตุ้น (Activation function) เช่น ฟังก์ชันซิกมอยด์ (Sigmoid function) จะเชื่อมสัญญาณจากนิวรอนของชั้นก่อนหน้าไปยังสัญญาณคำตอบที่ส่งต่อไปยังนิวรอนของชั้นถัดไป ความสัมพันธ์ระหว่างอินพุตและเอาต์พุตเป็นการแมปแบบไม่เชิงเส้นและมีความยืดหยุ่น หากมีการใช้ฟิลเตอร์อย่างเหมาะสม CNN สามารถเรียนรู้การพึ่งพาทางพื้นที่และเวลาได้ ตัวดำเนินการคอนโวลูชัน จะทำงานโดยการวางฟิลเตอร์ที่เกี่ยวข้องลงบนภาพ และนำค่าที่ตรงกันระหว่างฟิลเตอร์กับค่าของภาพมาคูณกันแล้วนำมาบวกกัน ฟิลเตอร์เหล่านี้สามารถวางลงบนภาพได้โดยตรง CNN สามารถเรียนรู้และจำแนกภาพได้ ตัวอย่างการประยุกต์ใช้ CNN ในการตรวจจบบรอยร้าวของกระจกแสดงให้เห็นในรูปที่ 2.2



รูปที่ 2.2 การประยุกต์ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน [13]

2.4 โมเดลสำหรับงานบนอุปกรณ์พกพา

MobileNet (โมบายเน็ต) สถาปัตยกรรมของ Convolutional Neural Network (CNN) ชุดหนึ่ง ที่ออกแบบโดย Google สำหรับการใช้งานในงานประมวลผลภาพ เช่น Image Classification, Object Detection บนอุปกรณ์เคลื่อนที่หรืออุปกรณ์ฝังตัว ที่มีข้อจำกัดด้านทรัพยากร เช่น หน่วยประมวลผล, พลังงาน และหน่วยความจำ MobileNet ใช้เทคนิค Depthwise Separable Convolutions แทนการใช้ Convolution แบบปกติ เพื่อควบคุมจำนวนพารามิเตอร์และการคำนวณให้ต่ำลงอย่างมากซึ่งช่วยให้โมเดลมีขนาดเล็กและทำงานได้รวดเร็ว

A. G. Howard and M. Zhu [14] ระบุว่า MobileNet คือตระกูลของโมเดลคอมพิวเตอร์วิทัศน์ที่เน้นการใช้งานบนอุปกรณ์พกพาสำหรับ TensorFlow ซึ่งเป็นไลบรารีซอฟต์แวร์แบบโอเพนซอร์ส (Open Source Library) โดยออกแบบมาเพื่อเพิ่มประสิทธิภาพความแม่นยำสูงสุด ในขณะที่คำนึงถึงข้อจำกัดของทรัพยากรสำหรับการใช้งานบนอุปกรณ์หรือระบบฝังตัวซึ่งเป็นโมเดลขนาดเล็ก มีความหน่วงต่ำ และใช้พลังงานต่ำ โดยสามารถปรับแต่งได้ให้เหมาะสมกับข้อจำกัดของทรัพยากรในแต่ละกรณีใช้งาน โดยโมเดลเหล่านี้สามารถนำไปต่อยอดใช้สำหรับงานจำแนกภาพ การตรวจจับ การฝังคุณลักษณะ และการแบ่งส่วนภาพ ได้คล้ายกับโมเดลขนาดใหญ่ที่ได้รับความนิยมอื่นๆ เช่น Inception เป็นต้น

2.5 ชุดข้อมูลภาพ (Image Dataset)

2.5.1 ImageNet (อิมเมจเน็ต)

Imagent Official Website [15] ระบุว่า ฐานข้อมูลภาพที่ถูกจัดระเบียบตามระบบตามลำดับชั้นของ WordNet ซึ่งปัจจุบันมีเฉพาะคำนาม โดยแต่ละโหนดในลำดับชั้นจะถูกแทนด้วยภาพหลายร้อยหรือหลายพันภาพ ทำให้ โครงการนี้มีบทบาทสำคัญในการพัฒนาวิจัยด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) และการเรียนรู้แบบเชิงลึก (Deep Learning) ซึ่งข้อมูลดังกล่าวเปิดให้ใช้งานฟรีสำหรับนักวิจัยเพื่อวัตถุประสงค์ที่ไม่ใช่เชิงพาณิชย์

J. Deng et al. [16] ระบุว่า ImageNet ถูกสร้างขึ้นโดยอาศัยโครงสร้างลำดับชั้นที่จัดเตรียมโดย WordNet ในขั้นตอนการดำเนินงานจนแล้วเสร็จ ImageNet มีเป้าหมายที่จะประกอบด้วยภาพที่มีการติดป้ายกำกับอย่างถูกต้องและความละเอียดเต็มจำนวนประมาณ 50 ล้านภาพ (500-1,000 ภาพต่อหนึ่ง synset) ณ ขณะที่เขียนบทความฉบับนี้ ImageNet ประกอบด้วย 12 ชั้นทรี โดยการวิเคราะห์ส่วนใหญ่จะอ้างอิงจากชั้นทรีของสัตว์เลี้ยงลูกด้วยนมและยานพาหนะ

สำหรับ ImageNet มีเป้าหมายในการนำเสนอชุดข้อมูลภาพที่ครอบคลุมและหลากหลายมากที่สุด ปัจจุบัน 12 ชั้นทรีดังกล่าวมี 5247 Synsets ประกอบด้วยภาพที่มีการใส่คำอธิบายอย่างถูกต้องรวมทั้งสิ้น 3.2 ล้านภาพ

2.6 แนวคิดการสร้างและจัดการคลังคำศัพท์หลายภาษา

2.6.1 WordNet (เครือข่ายคำ)

ธนัท ลีลาน้อย [17] ได้กล่าวถึงในงานวิจัยไว้ว่า Wordnet เป็นข้อมูลคลังคำศัพท์ประเภทหนึ่งที่มีการจัดหมวดหมู่ของคำไว้เป็นกลุ่มคำไวพจน์ (Synonym Set) ซึ่งจะมีการอธิบายความหมาย รายละเอียด คำจำกัดความ และความสัมพันธ์ทางความหมายระหว่างกลุ่มคำไวพจน์ต่างๆ

เครือข่ายคำที่สร้างขึ้นเป็นครั้งแรก เป็นเครือข่ายคำภาษาอังกฤษคือ เครือข่ายคำพริ้นสตัน (Princeton WordNet) ถูกสร้างขึ้นในปี ค.ศ. 1985 โดยหน่วยปฏิบัติการวิจัยทางภาษาศาสตร์แห่งมหาวิทยาลัยพริ้นสตัน ประเทศสหรัฐอเมริกา

เครือข่ายคำมีความสำคัญต่องานทางด้านการประมวลผลภาษาธรรมชาติ เนื่องจากประโยคทุกประโยคร้อยเรียงขึ้นจากคำหลายคำ ฉะนั้นระบบต่างๆที่ต้องการประมวลผลภาษาธรรมชาติให้เหมือนอย่างที่มีมนุษย์ทำได้ จึงต้องรู้ความหมายและข้อมูลสำคัญเกี่ยวกับคำต่างๆ ที่มาร้อยเรียงกันเป็นประโยคเหล่านั้น โดยปกติแล้วข้อมูลดังกล่าวจะอยู่ในพจนานุกรม แต่พจนานุกรมสร้างมาเพื่อให้มนุษย์ใช้งานได้เท่านั้น คอมพิวเตอร์ไม่สามารถเข้าใจได้ เครือข่ายคำจึงถูกพัฒนาขึ้นโดยการผนวกกันของข้อมูลคลังคำศัพท์ (Lexicographic Information) แบบเดิมกับการสร้างรูปแบบทางคณิตศาสตร์ที่คอมพิวเตอร์สามารถประมวลผลได้

Open English Wordnet ซึ่งถูกพัฒนาต่อยอดมาจาก Princeton WordNet ได้ปล่อยเวอร์ชัน 2025 Edition (Released 31st December 2025) มีจำนวนคำ (Words) 161,875 ซินเซต (Synsets) 120,564 และความสัมพันธ์ (Relations) 419,226 [18]

มหาวิทยาลัยพรินสตัน [19] ระบุถึงโครงสร้างของ WordNet ไว้ว่า ความสัมพันธ์หลักระหว่างคำใน WordNet คือ ความเป็นคำพ้องความหมาย (Synonymy) เช่น ระหว่างคำว่า Shut กับ Close หรือ Car กับ Automobile คำพ้องความหมายซึ่งเป็นคำที่สื่อถึงแนวคิดเดียวกันและสามารถใช้แทนกันได้หลายบริบท จะถูกรวมไว้เป็นชุดที่ไม่เรียงลำดับ เรียกว่า ซินเซต (Synset) โดยซินเซตทั้ง 117,000 ชุด ใน WordNet จะเชื่อมโยงถึงกันด้วย ความสัมพันธ์เชิงแนวคิด (Conceptual Relations) จำนวนไม่มาก นอกจากนี้ ซินเซตแต่ละชุดยังมี คำจำกัดความสั้นๆ (Gloss) โดยส่วนมากจะมีประโยคตัวอย่างหนึ่งประโยคหรือมากกว่า เพื่อแสดงการใช้งานของคำในซินเซตนั้น คำรูปเดียวกันที่มีความหมายแตกต่างกันหลายความหมาย จะถูกแทนด้วย ซินเซตที่แตกต่างกันตามจำนวนความหมาย ดังนั้น คู่ระหว่างรูปคำ-ความหมาย (Form-Meaning Pair) แต่ละคู่ใน WordNet จึงมีลักษณะเฉพาะและไม่ซ้ำกัน

2.7 Flutter

Framework สำหรับการพัฒนาแอปพลิเคชันแบบข้ามแพลตฟอร์มซึ่งพัฒนาโดยบริษัท Google ซึ่งใช้ภาษา Dart ช่วยให้นักพัฒนาสามารถสร้างแอปพลิเคชันสำหรับ ระบบปฏิบัติการหลายแพลตฟอร์มจากฐานโค้ดเดียวกันได้ จึงสามารถลดเวลาและค่าใช้จ่ายในการพัฒนาแอปพลิเคชัน

เว็บไซต์ทางการของ Flutter [20] ระบุว่า Flutter คือ UI toolkit แบบ Cross-Platform ที่ถูกออกแบบมาเพื่อให้สามารถใช้โค้ดชุดเดียวกันในการพัฒนาแอปพลิเคชันสำหรับระบบปฏิบัติการที่หลากหลาย เช่น iOS, Android, เว็บ และเดสก์ท็อป พร้อมทั้งยังสามารถเชื่อมต่อกับบริการของแพลตฟอร์มนั้นๆ ได้โดยตรง

เป้าหมายของ Flutter คือการทำให้ นักพัฒนาสามารถสร้างแอปพลิเคชันประสิทธิภาพสูงที่ให้ความรู้สึกเป็นธรรมชาติและกลมกลืนไปกับแต่ละแพลตฟอร์มโดยยอมรับความแตกต่างของแพลตฟอร์มนั้นๆ แต่ยังคงใช้โค้ดร่วมกันให้ได้มากที่สุด

ในระหว่างการพัฒนา แอปพลิเคชันของ Flutter จะทำงานบน Virtual Machine (VM) ซึ่งมีฟีเจอร์เด่นคือ Stateful Hot Reload ที่ทำให้เห็นผลการเปลี่ยนแปลงของโค้ดได้ทันทีโดยไม่ต้องคอมไพล์ใหม่ทั้งหมด สำหรับเวอร์ชันเว็บ Flutter จะรองรับ Hot Restart และ Hot Reload ซึ่งอาจต้องเปิดใช้งานผ่าน flag

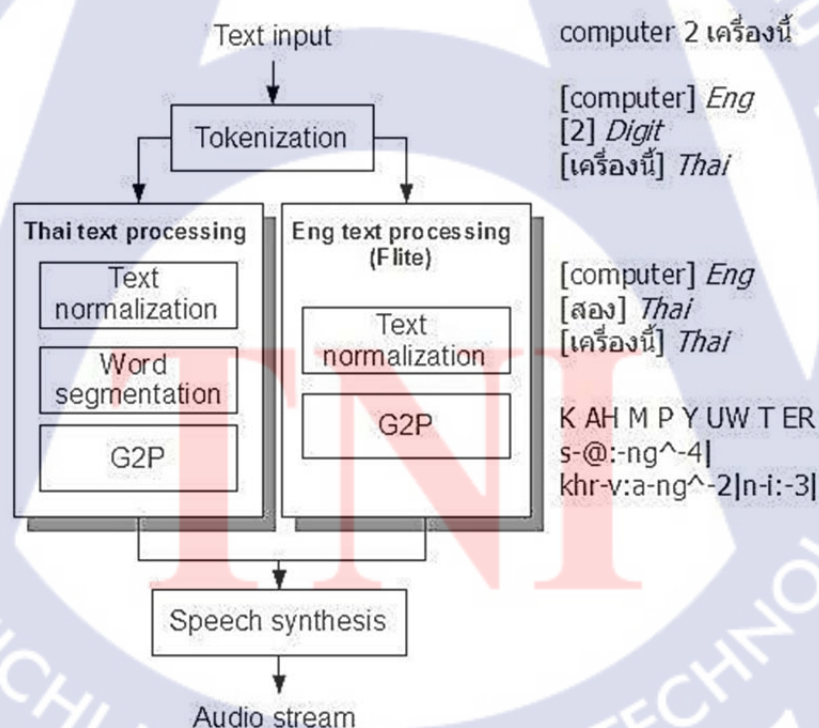
สำหรับเวอร์ชันที่นำไปใช้งานจริง แอปพลิเคชันของ Flutter จะถูกคอมไพล์โดยตรงเป็น Machine Code ไม่ว่าจะเป็นชุดคำสั่ง Intel x64 หรือ ARM หรือแปลงเป็น JavaScript ในกรณีที่สร้างสำหรับเว็บ ตัวเฟรมเวิร์กเป็นโอเพนซอร์ส ภายใต้ใบอนุญาตแบบ BSD (Berkeley Software Distribution) ซึ่งเป็น

ใบอนุญาตซอฟต์แวร์โอเพนซอร์สประเภทให้อิสระสูง (Permissive) ที่มีความยืดหยุ่นสูง และมี Ecosystem ของแพ็คเกจจากนักพัฒนาภายนอกที่แข็งแกร่ง ซึ่งช่วยเสริมความสามารถหลักของไลบรารีได้เป็นอย่างดี

2.8 เทคโนโลยี Text-to-Speech (TTS)

การแปลงข้อความเป็นเสียงพูด หรือในภาษาไทยคือ เทคโนโลยีแปลงข้อความเป็นคำพูดซึ่งเป็นเทคโนโลยีที่เปลี่ยนข้อความดิจิทัลให้กลายเป็นเสียงพูดที่เป็นธรรมชาติ เหมือนมนุษย์โดยอัตโนมัติ [21] เทคโนโลยีนี้ใช้หลักการของ AI และ Machine Learning เพื่อวิเคราะห์ข้อความและสร้างเสียงสังเคราะห์ที่มีความใกล้เคียงกับเสียงคนจริงๆ ซึ่งต่างจากเสียงสังเคราะห์แบบเก่าที่ฟังดูเป็นเสียงหุ่นยนต์

การสังเคราะห์เสียงจากข้อความ ปัจจุบันมีหลาย library เช่น Google TTS, Festival, Microsoft Speech API ซึ่งสามารถกำหนดภาษาและเสียงได้ให้ใกล้เคียงกับมนุษย์ได้ โดย S. Saychum et al. [22] ได้กล่าวถึงระบบ TTS ที่เสนอ พร้อมทั้งอธิบายถึงรายละเอียดขององค์ประกอบบางส่วนที่จำเป็นต้องได้รับการปรับปรุง (Optimize) เพื่อให้สามารถใช้งานออนไลน์ TTS ดังกล่าวบนอุปกรณ์ Android ได้ตามรูปภาพในด้านล่างนี้



รูปที่ 2.3 แผนภาพโดยรวมของระบบ TTS สองภาษา ไทย-อังกฤษ [22]

2.9 แนวคิดการออกแบบเพื่อมวลชน (Universal Design)

วรภรณ มามี [23] กล่าวถึงในงานวิจัยว่า การออกแบบเพื่อมวลชน หรือ Universal Design เป็นแนวคิดหลักที่มุ่งเน้นการศึกษากระบวนการออกแบบสิ่งแวดล้อม สถานที่ และผลิตภัณฑ์ต่างๆ ให้เป็นสากล เพื่อรองรับการใช้งานของมนุษย์ทุกกลุ่มในสังคมอย่างเท่าเทียมโดยปราศจากการแบ่งแยก ซึ่งมีจุดมุ่งหมายสำคัญคือการสร้างสรรค์สิ่ง que ทุกคนสามารถเข้าถึงและใช้ประโยชน์ได้ทันทีโดยไม่ต้องมีการดัดแปลงโครงสร้างหรือออกแบบอุปกรณ์เสริมพิเศษเพื่อตอบสนองความต้องการของบุคคลกลุ่มใดกลุ่มหนึ่งเป็นการเฉพาะ และแนวคิดนี้จึงถูกนำมาประยุกต์ใช้อย่างกว้างขวางในกลุ่มงานด้านการพัฒนาคุณภาพชีวิตผู้สูงอายุ คนพิการ และผู้ด้อยโอกาสที่มีข้อจำกัดในการเข้าถึงสภาพแวดล้อมและเครื่องใช้ทั่วไปในชีวิตประจำวัน โดยหัวใจสำคัญของการออกแบบเพื่อมวลชนเริ่มต้นจากการวิเคราะห์ความหลากหลายของมนุษย์ ทั้งในด้านสรีระและทักษะการรับรู้ ไม่ว่าจะเป็นผู้สูงอายุ ผู้ป่วย สตรีมีครรภ์ เด็กเล็กที่เดินทางด้วยรถเข็น หรือผู้พิการประเภทต่างๆ ทั้งทางด้านร่างกาย การเคลื่อนไหว ทางสติปัญญา และทางประสาทสัมผัส เช่น ผู้พิการทางสายตาและผู้พิการทางการได้ยิน รวมถึงผู้ที่มีข้อจำกัดด้านการสื่อสารหรือการอ่านออกเขียนได้

หลักการของการออกแบบเพื่อมวลชน

1. ความเสมอภาค ใช้งานได้กับทุกคนในสังคมอย่างเท่าเทียมกัน ไม่มีการแบ่งแยกและเลือกปฏิบัติ
2. ความยืดหยุ่น ใช้งานได้กับผู้ที่ถนัดทั้งซ้ายและขวา หรือสามารถปรับสภาพความสูงต่ำให้เหมาะสมกับความสูงของผู้ใช้งานแต่ละคนได้
3. ความเรียบง่ายและเข้าใจได้ดี มีการใช้ภาพหรือคำอธิบายที่ง่ายสำหรับคนทุกประเภท ไม่ว่าจะมีความรู้ระดับใด หรือมีข้อจำกัดด้านภาษาและการอ่าน โดยอาจใช้รูปภาพเป็นสัญลักษณ์สากลเพื่อการสื่อสารที่เข้าใจง่าย
4. มีข้อมูลพอเพียง มีข้อมูลที่ง่ายและเพียงพอสำหรับประกอบการใช้งาน
5. ทนทานต่อการใช้งานที่ผิดพลาด มีระบบป้องกันอันตรายหากมีการใช้งานผิดพลาด ตัวอุปกรณ์ต้องมีความทนทาน ไม่เสียหายได้โดยง่าย
6. พุน่ร่างกาย เน้นความสะดวกและไม่ต้องออกแรงมาก
7. ขนาดและสถานที่ที่เหมาะสม สามารถใช้งานในเชิงปฏิบัติได้จริง โดยออกแบบเผื่อสำหรับผู้ที่มีร่างกายใหญ่โตและผู้ que เคลื่อนไหวร่างกายได้ยาก

2.10 เมตริกการประเมิน (Evaluation Metrics)

2.10.1 ด้านประสิทธิภาพของแอปพลิเคชัน

(1) Top-1 Accuracy

อัตราส่วนของการทำนายผลลัพธ์ที่ถูกต้อง เมื่อพิจารณาเฉพาะคลาสที่โมเดลให้ค่าความน่าจะเป็นสูงสุดเพียงคลาสเดียว ในการจำแนกประเภทแบบหลายคลาส (Multi-Class Classification) โมเดลจะให้ค่าความน่าจะเป็นออกมาสำหรับทุกๆ คลาสที่เป็นไปได้ Top-1 Accuracy จะนับว่าการทำนายนั้น ถูกในกรณีดังนี้

1. คลาสที่โมเดลทำนายว่ามีความน่าจะเป็นสูงสุด
2. เป็นคลาสที่ ถูกต้องจริง (Ground Truth)

สามารถคำนวณได้จากสมการที่ 2.1

$$\text{Top-1 Accuracy} = \frac{\text{จำนวนภาพที่ทำนายถูกต้องในอันดับที่ 1}}{\text{จำนวนภาพทดสอบทั้งหมด}} \quad (2.1)$$

(2) Top-k Accuracy

อัตราส่วนของการทำผลลัพธ์ที่ถูกต้องปรากฏอยู่ภายในผลการทำนาย k อันดับแรกของโมเดล เพื่อประเมินความแม่นยำของโมเดล

$$\text{Top-k Accuracy} = \frac{\text{จำนวนภาพที่มีคำตอบถูกต้องอยู่ใน k อันดับแรก}}{\text{จำนวนภาพทดสอบทั้งหมด}} \quad (2.2)$$

2.10.2 ด้านความสามารถในการใช้งานแอปพลิเคชัน

1. System Usability Scale (SUS)

J. Brooke [24] ระบุว่า SUS เป็นการประเมินความสามารถในการใช้งาน (Usability) แบบรวดเร็วและต้นทุนต่ำสำหรับใช้ในการวัดการรับรู้ถึงการใช้การได้ของระบบอย่างกว้างๆ ประกอบด้วยคำถาม 10 ข้อ ให้คะแนนแบบไลเคิร์ต 5 ระดับ สลับข้อคำถามเชิงบวกกับคำถามเชิงลบ เพื่อลดอคติในการตอบและให้คะแนนเป็นตัวเลขเดียวเพื่อแทนภาพรวมความน่าใช้ของระบบ การรับรู้ความสามารถในการใช้งานโดยรวม คือ การวัดผลจากมุมมองและความรู้สึกของผู้ใช้ว่าผลิตภัณฑ์ระบบ หรือบริการนั้นว่าใช้งานง่ายและน่าพึงพอใจเพียงใดซึ่งเป็นการประเมินภาพรวมทั้งหมด ไม่ใช่เพียงแค่ฟังก์ชันใดฟังก์ชันหนึ่งเท่านั้น โดยตัวอย่างของต้นฉบับแบบสอบถามดังรูปที่ 2.4

อีกทั้ง จากงานวิจัยของ เกริกเกียรติ สมศรี [25] ที่กล่าวถึง System Usability Scale (SUS) สามารถสรุปได้ดังนี้

System Usability Scale (SUS) เป็นแบบสอบถามมาตรฐานสากลสำหรับวัดความสามารถในการใช้งาน (Usability) ของระบบ เชื่อถือได้และใช้กันแพร่หลาย พัฒนาโดย John Brooke ในปี ค.ศ. 1996 จุดเด่นคือความเรียบง่าย ยึดหยุ่นต่อระบบหลากหลายประเภท และสามารถประเมินประสบการณ์ผู้ใช้ในภาพรวมโดยใช้เวลาไม่มาก แบบสอบถามมี 10 ข้อ ใช้รูปแบบมาตราลิเกิร์ต (Likert Scale) 5 ระดับ และคำนวณคะแนนรวมอยู่ในช่วง 0-100 โดยคะแนนนี้เป็นสเกลเฉพาะของ SUS ไม่ใช่เปอร์เซ็นต์

Usability ตามมาตรฐาน ISO 9241-11 คือระดับที่ผู้ใช้สามารถใช้ผลิตภัณฑ์เพื่อ บรรลุเป้าหมายอย่างมีประสิทธิภาพ (Effectiveness) มีประสิทธิภาพ (Efficiency) และมีความพึงพอใจ (Satisfaction) ภายใต้บริบทการใช้งานที่กำหนด แนวคิดนี้เน้นว่าการวัด Usability ต้องพิจารณาปัจจัย ผู้ใช้ ลักษณะงาน และสิ่งแวดล้อมร่วมกัน SUS ออกแบบมาให้สะท้อนมิติเหล่านี้ในภาพรวม รวมถึง Learnability และ Confidence ผ่านแบบสอบถาม 10 ข้อ บนมาตรวัดเดียวกัน

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	☐	☐	☐	☐	☐
	1	2	3	4	5
2. I found the system unnecessarily complex	☐	☐	☐	☐	☐
	1	2	3	4	5
3. I thought the system was easy to use	☐	☐	☐	☐	☐
	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	☐	☐	☐	☐	☐
	1	2	3	4	5
5. I found the various functions in this system were well integrated	☐	☐	☐	☐	☐
	1	2	3	4	5
6. I thought there was too much inconsistency in this system	☐	☐	☐	☐	☐
	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	☐	☐	☐	☐	☐
	1	2	3	4	5
8. I found the system very cumbersome to use	☐	☐	☐	☐	☐
	1	2	3	4	5
9. I felt very confident using the system	☐	☐	☐	☐	☐
	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	☐	☐	☐	☐	☐
	1	2	3	4	5

รูปที่ 2.4 ต้นฉบับแบบสอบถาม SUS

แบบสอบถาม SUS มีความน่าเชื่อถือสูง โดยเฉพาะความเชื่อมั่นภายใน (Internal Consistency) ซึ่งมีค่า Cronbach's Alpha เฉลี่ยมากกว่า 0.85 อย่างต่อเนื่องในหลายบริบท ถือว่าอยู่ในเกณฑ์ดีมาก และยังมีความเสถียรของการวัดแม้ใช้กับกลุ่มตัวอย่างขนาดจำกัด ด้านความเที่ยงตรง (Validity) SUS แสดงความตรงตามสภาพสร้าง (Construct Validity) เมื่อเทียบกับเครื่องมืออื่นที่ซับซ้อนกว่า และมีความสัมพันธ์อย่างมีนัยสำคัญกับตัวชี้วัดที่เกี่ยวข้อง เช่น ความพึงพอใจ ระยะเวลาในการปฏิบัติงาน และความตั้งใจใช้งานซ้ำ

จากรายละเอียดข้างต้น ทำให้ SUS เหมาะสมอย่างยิ่งในการประเมินระบบดิจิทัลที่มีปฏิสัมพันธ์กับผู้ใช้โดยตรง เนื่องจากสามารถสะท้อนระดับความเป็นมิตรต่อผู้ใช้ ความง่ายในการโต้ตอบ ความสามารถในการเรียนรู้และความพึงพอใจโดยรวม โดยผลลัพธ์ที่ได้จาก SUS ยังเอื้อต่อการเปรียบเทียบกับมาตรฐานอุตสาหกรรม หรือติดตามแนวโน้มเพื่อวางแผนปรับปรุงในรอบถัดไปอย่างเป็นระบบ ทั้งยังเป็นเครื่องมือสื่อสารภายในทีมพัฒนาที่ตั้งอยู่บนข้อมูลจริงของผู้ใช้ ช่วยยกระดับคุณภาพระบบให้สอดคล้องกับความต้องการของกลุ่มเป้าหมาย

2. การคำนวณคะแนน SUS

แบบประเมิน SUS จะให้ผลลัพธ์เป็นคะแนนตัวเลขชุดเดียว ซึ่งเป็นค่าดัชนีรวมที่สะท้อนถึงความสามารถในการใช้งานของระบบในภาพรวม ทั้งนี้ คะแนนในแต่ละข้อเพียงอย่างเดียวไม่สามารถนำมาตีความหมายแยกกันได้อย่างมีนัยสำคัญ โดยมีขั้นตอนการคำนวณคะแนนดังนี้

2.1 การหาคะแนนดิบรายข้อ (Score Contribution) คะแนนในแต่ละข้อจะมีค่าอยู่ระหว่าง 0 ถึง 4 คะแนน โดยคำนวณตามประเภทของข้อคำถาม ดังนี้

- สำหรับข้อคี่ (ข้อ 1, 3, 5, 7 และ 9) นำระดับคะแนนที่ผู้ใช้เลือกลบด้วย 1
- สำหรับข้อคู่ (ข้อ 2, 4, 6, 8 และ 10) นำ 5 ลบด้วยระดับคะแนนที่ผู้ใช้เลือก

2.2 การหาคะแนนรวม นำผลรวมคะแนนดิบจากทั้ง 10 ข้อ มาคูณด้วย 2.5 เพื่อให้ได้ค่าคะแนนรวมของ SUS (Overall Value)

2.3 ช่วงคะแนน คะแนนรวมของ SUS จะมีค่าอยู่ในช่วง 0 ถึง 100 คะแนน

3. การแปลผลคะแนน SUS

มีการกำหนดค่าเฉลี่ยอ้างอิงของคะแนน SUS ไว้ที่ 68 คะแนน ซึ่งใช้เป็นเกณฑ์มาตรฐานทั่วไป คะแนนตั้งแต่ 68 คะแนนขึ้นไป บ่งชี้ถึงระดับ Usability ที่ดีหรือสูงกว่าค่าเฉลี่ย ขณะที่คะแนนต่ำกว่า 68 มักสะท้อนปัญหาเชิงการใช้งาน โดยแปลงคะแนนเป็นระดับความพึงพอใจ

นอกจากนี้ พัชราวดี เดชะ และคณะ [26] ได้กล่าวถึงการแปลความหมายของคะแนน SUS ไว้ ตามรายละเอียดดังนี้

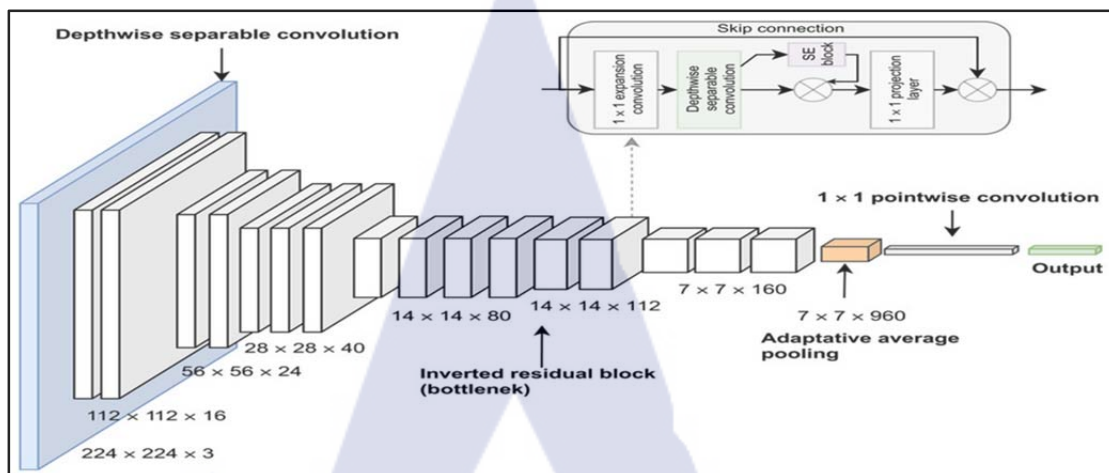
- คะแนน SUS ที่มากกว่า 80.3 หมายถึง การใช้งานอยู่ในระดับดีมาก (Excellent)
- คะแนน SUS อยู่ระหว่าง 68-80.3 หมายถึง การใช้งาน อยู่ในระดับดี (Good)
- คะแนน SUS เท่ากับ 68 หมายถึง การใช้งานอยู่ในระดับปานกลาง (Okay)
- คะแนน SUS อยู่ระหว่าง 51-68 หมายถึง การใช้งานอยู่ในระดับน้อย (Awful)
- คะแนน SUS น้อยกว่า 51 หมายถึง การใช้งานอยู่ในระดับน้อยที่สุด (Poor)

2.11 งานวิจัยที่เกี่ยวข้อง

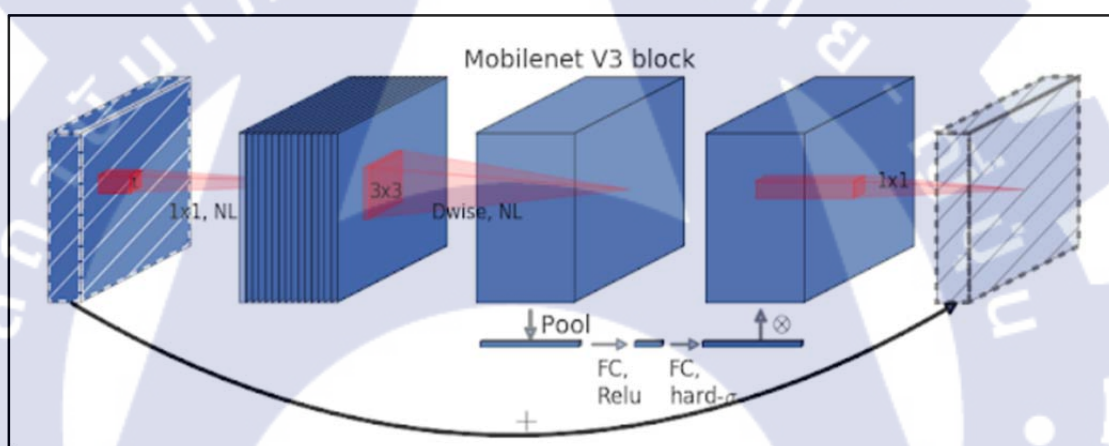
จากการรวบรวมข้อมูลที่เกี่ยวข้องกับการพัฒนาแอปพลิเคชันเพื่อสนับสนุนการเรียนรู้คำศัพท์หลายภาษาสำหรับผู้พิการทางสายตาเป็นหลัก สามารถแยกเป็นหมวดหมู่โดยสังเขปดังนี้

2.11.1 โมเดลสำหรับการจำแนกภาพ

MobileNet V3 โครงข่ายประสาทเทียมสำหรับงานจำแนกภาพบนอุปกรณ์พกพา (On-device Image Classification) ที่พัฒนาขึ้นเพื่อให้ได้สมดุลระหว่าง ความแม่นยำ และเวลาในการประมวลผลบน CPU ของโทรศัพท์มือถือ โดย A. Howard et al.[27] นำเสนอแนวคิดการสร้าง Mobile NetV3 ผ่านการผสมผสานเทคนิคการค้นหาโครงสร้างที่เสริมกัน (Complementary Search Techniques) กับการออกแบบสถาปัตยกรรมรูปแบบใหม่ กล่าวคือ ใช้ Hardware-Aware Network Architecture Search (NAS) เพื่อค้นหาโครงสร้างเครือข่ายที่เหมาะสมกับฮาร์ดแวร์จริง ร่วมกับ NetAdapt เพื่อปรับจำนวนช่องสัญญาณ ฟิลเตอร์ในแต่ละเลเยอร์ให้สอดคล้องกับข้อจำกัดด้านความหน่วง (Latency) และปรับปรุงต่อยอดแนวทางเชิงสถาปัตยกรรม เช่น การออกแบบเลเยอร์ที่มีต้นทุนสูงให้มีประสิทธิภาพมากขึ้น การใช้กลไก Squeeze-and-Excitation (SE) และการเลือกฟังก์ชันกระตุ้นที่เหมาะสมต่อการนำไปใช้งานจริงบนมือถือ เช่น Hard-Sigmoid, h-swish ซึ่งลดภาระการคำนวณเมื่อเทียบกับรูปแบบเดิม ผลการทดลองบนชุดข้อมูล ImageNet รายงานว่า MobileNetV3-Large ให้ความแม่นยำสูงกว่า MobileNetV2 สำหรับการจำแนกภาพ โดยมีค่า Top-1 accuracy เพิ่มขึ้นประมาณ 3.2% พร้อมทั้งลดความหน่วง (Latency) ได้ประมาณ 20% เมื่อเทียบกับ MobileNetV2 ขณะที่ MobileNetV3-Small ให้ความแม่นยำเพิ่มขึ้นประมาณ 6.6% เมื่อเทียบกับ MobileNetV2 ที่มีความหน่วง (Latency) ใกล้เคียงกัน

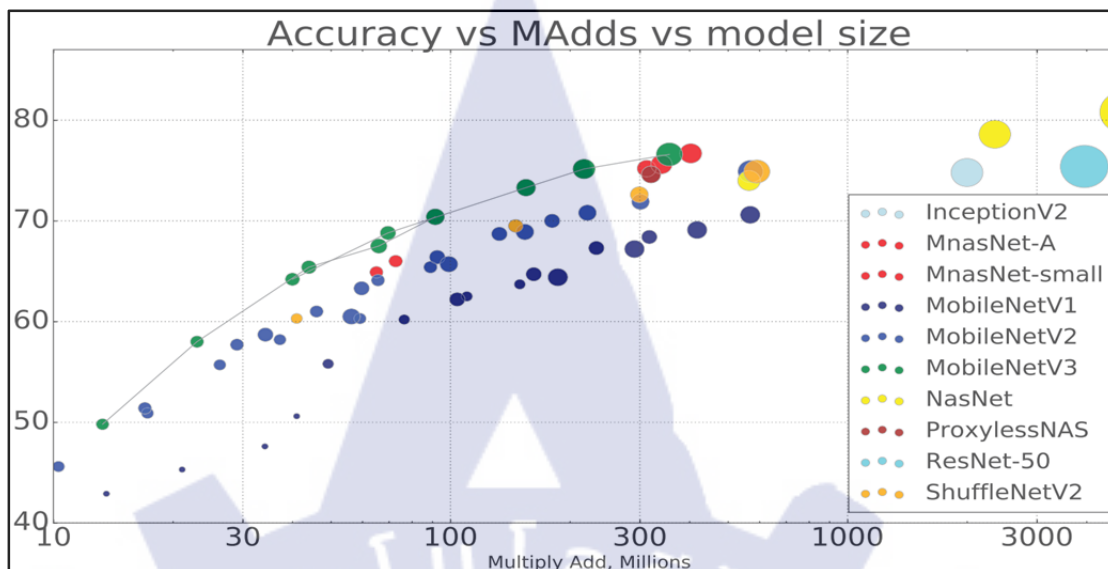


รูปที่ 2.5 ภาพรวมสถาปัตยกรรมและส่วนประกอบหลักของ MobileNetV3 [28]



รูปที่ 2.6 MobileNet V3 Block [27]

MobileNetV3 โดยเฉพาะรุ่น Large เหมาะสมต่อการประยุกต์ใช้เป็นโมเดลหลักสำหรับงานจำแนกภาพบนอุปกรณ์พกพาที่ต้องการทั้งความแม่นยำและความรวดเร็ว ซึ่งสอดคล้องกับงานวิจัยนี้ที่เลือกใช้ MobileNetV3-Large ที่ผ่านการฝึกฝนด้วย ชุดข้อมูล ImageNet เพื่อรองรับการจำแนกภาพบนอุปกรณ์เคลื่อนที่ โดยภาพรวมของ MobileNetV3 มีรายละเอียดดังรูปที่ 2.5 และรูปที่ 2.6 อีกทั้ง เมื่อเปรียบเทียบ MobileNet V3 กับโมเดลจำแนกภาพอื่นๆ แล้ว พบว่ามีประสิทธิภาพเหนือกว่าโมเดลชั้นนำ (State of the art) เช่น MnasNet, ProxylessNas และ MobileNetV2 ดังรูปที่ 2.7



รูปที่ 2.7 เปรียบเทียบ Accuracy ต่อ MAdds ต่อขนาดโมเดล [27]

ตามรายละเอียดที่แสดงใน รูปที่ 2.8 ที่ระดับด้านล่าง จะเห็นได้ว่า MobileNetV3-Large แสดงให้เห็นถึงประสิทธิภาพด้านฮาร์ดแวร์ที่เหนือกว่าอย่างชัดเจน โดยเฉพาะอย่างยิ่ง เมื่อเปรียบเทียบกับ MnasNet-A1 แม้ว่าทั้งคู่จะมีความแม่นยำ (Top-1 Accuracy) เท่ากันที่ 75.2% แต่ MobileNet V3-Large กลับใช้ทรัพยากรในการคำนวณน้อยกว่าอย่างมาก (219 MAdds เทียบกับ 315 MAdds) และมีค่าความหน่วงในการประมวลผล (Inference Latency) บนอุปกรณ์พกพาที่ต่ำกว่าอย่างเห็นได้ชัด (เช่น 51 ms เทียบกับ 71 ms บนเครื่อง Pixel-1) [27]

Network	Top-1	MAdds	Params	P-1	P-2	P-3
V3-Large 1.0	75.2	219	5.4M	51	61	44
V3-Large 0.75	73.3	155	4.0M	39	46	40
MnasNet-A1	75.2	315	3.9M	71	86	61
Proxyless[5]	74.6	320	4.0M	72	84	60
V2 1.0	72.0	300	3.4M	64	76	56
V3-Small 1.0	67.4	66	2.9M	15.8	19.4	14.4
V3-Small 0.75	65.4	44	2.4M	12.8	15.6	11.7
Mnas-small [43]	64.9	65.1	1.9M	20.3	24.2	17.2
V2 0.35	60.8	59.2	1.6M	16.6	19.6	13.9

Table 3. Floating point performance on the Pixel family of phones (P-n denotes a Pixel-n phone). All latencies are in ms and are measured using a single large core with a batch size of one. Top-1 accuracy is on ImageNet.

รูปที่ 2.8 การเปรียบเทียบประสิทธิภาพการประมวลผลตมบนอุปกรณ์พกพา [27]

2.11.2 ชุดข้อมูลสำหรับโมเดลจำแนกภาพ

T. Ridnik et al. [29] ได้นำเสนองานวิจัยเรื่อง ImageNet-21K Pretraining for the Masses ซึ่งมุ่งเน้นการแก้ปัญหาและเพิ่มประสิทธิภาพในการเทรนโมเดลล่วงหน้า (Pretraining) โดยใช้ชุดข้อมูล ImageNet-21K ที่มีความหลากหลายและขนาดใหญ่กว่าชุดข้อมูลมาตรฐาน ImageNet-1K อย่างมาก โดยจะเรียกชุดข้อมูลนี้ว่า ImageNet-21K-P ซึ่งมีจำนวนคลาสอยู่ 11,221 Classes

งานวิจัยนี้ชี้ให้เห็นถึงข้อจำกัดของ ImageNet-1K ซึ่งเป็นเพียงสับเซตที่มี 1.2 ล้านภาพและ 1,000 คลาส ในขณะที่ ImageNet-21K มีภาพกว่า 14 ล้านภาพและกว่า 21,000 คลาส

จากเอกสารงานวิจัยระบุว่า ในขั้นตอนการเตรียมข้อมูล (Preprocessing) เพื่อสร้างชุดข้อมูล ImageNet-21K-P ผู้วิจัยได้ทำการตัดคลาสที่มีข้อมูลน้อยออก ซึ่งเป็นตัดคลาสที่มีจำนวนภาพตัวอย่าง น้อยกว่า 500 ภาพ ออกจากชุดข้อมูล เนื่องจากคลาสเหล่านี้มีจำนวนตัวอย่างน้อยเกินไป ซึ่งไม่เพียงพอให้โมเดลเรียนรู้ได้อย่างมีประสิทธิภาพ และอาจเป็นตัวลวงที่ทำให้คุณภาพการเทรนโดยรวมลดลง

ผลลัพธ์จากการตัดคลาสดังกล่าวออกทำให้จำนวนคลาสดลดลงจากเดิม 21,841 คลาส เหลือเพียง 11,221 คลาส ซึ่งลดลงไปประมาณครึ่งหนึ่ง อย่างไรก็ตาม การตัดคลาสเหล่านี้ออกไปส่งผลให้จำนวนภาพรวมในชุดข้อมูลหายไปเพียง 13% เท่านั้น ซึ่งถือว่าคุ้มค่าแลกกับคุณภาพข้อมูลที่ดีขึ้น

อย่างไรก็ตาม อุปสรรคสำคัญในการใช้งาน ImageNet-21K คือความซับซ้อนของป้ายกำกับ (Labels) ที่ไม่ได้แยกขาดจากกัน (Mutually Exclusive) และมีความสัมพันธ์เชิงโครงสร้างความหมายตามหลัก WordNet เช่น ภาพหนึ่งอาจถูกระบุว่าเป็นเก้าอี้ หรือเฟอร์นิเจอร์ ก็ได้ ซึ่งสร้างความยากลำบากในการปรับแต่งกระบวนการเรียนรู้ของโมเดลให้มีประสิทธิภาพ

เพื่อให้สามารถใช้ประโยชน์จากชุดข้อมูลขนาดใหญ่ได้อย่างเต็มที่ ผู้วิจัยได้พัฒนาเทคนิคใหม่ที่เรียกว่า Semantic Softmax ซึ่งใช้ประโยชน์จากโครงสร้างลำดับชั้นของ WordNet (Hierarchical Structure) โดยแบ่ง Softmax Layer ออกตามระดับชั้นของความหมายแทนที่จะใช้ Softmax เดียวรวมทั้งหมด วิธีนี้ช่วยให้โมเดลสามารถเรียนรู้ความสัมพันธ์ของวัตถุได้ถูกต้องแม่นยำกว่าการเทรนแบบ Single-Label หรือ Multi-Label แบบดั้งเดิม นอกจากนี้ยังมีการนำเทคนิค Semantic Knowledge Distillation มาประยุกต์ใช้เพื่อถ่ายทอดความรู้และเติมเต็มข้อมูลป้ายกำกับที่อาจขาดหายไปในการตัดข้อมูลต้นฉบับ ทำให้การเรียนรู้ของโมเดลมีความสมบูรณ์ยิ่งขึ้น

ผลการทดลองแสดงให้เห็นว่า วิธีการ Pretraining แบบ Semantic Softmax บนชุดข้อมูล ImageNet-21K ช่วยเพิ่มประสิทธิภาพให้กับโมเดลสถาปัตยกรรมต่างๆ ได้อย่างมีนัยสำคัญเมื่อนำไปใช้กับงานเฉพาะทาง (Downstream Tasks) โดยให้ผลลัพธ์ที่ดีกว่าการ Pretraining ด้วย ImageNet-1K และดีกว่าเทคนิค Multi-Label ทั่วไปที่มีก่ประสบปัญหาความไม่สมดุลของข้อมูล

Backbone	ImageNet-21K-P semantic top-1 Accuracy [%]	ImageNet-1K top-1 Accuracy [%]	Maximal batch size	Maximal training speed (img/sec)	Maximal inference speed (img/sec)
MobilenetV3_large_100	73.1	78.0	488	1210	5980
OFA_flops_595m_s	75.0	81.0	288	500	3240
ResNet50	75.6	82.0	320	720	2760
TResNet-M	76.4	83.1	520	670	2970
TResNet-L (V2)	76.7	83.9	240	300	1460
ViT-B-16	77.6	84.4	160	340	1140

รูปที่ 2.9 Pretrained Models on ImageNet-21-K-P Dataset [29]

MobileNetV3 มีความแม่นยำเพิ่มขึ้นเป็น 78.0% จากเดิม 75.2% ซึ่งยืนยันได้ว่าการใช้ชุดข้อมูลขนาดใหญ่ร่วมกับเทคนิคการเทรนที่เหมาะสม สามารถยกระดับขีดความสามารถของโมเดลขนาดเล็กให้มีประสิทธิภาพสูงขึ้นได้โดยไม่กระทบต่อความเร็วในการประมวลผลดังรูปที่ 2.9

ประเด็นสำคัญที่ค้นพบในงานวิจัยนี้คือ ประโยชน์ของการ Pretraining ด้วยชุดข้อมูลขนาดใหญ่ไม่ได้จำกัดอยู่แค่โมเดลที่มีขนาดใหญ่เท่านั้น แต่ยังส่งผลต่ออย่างยิ่งต่อ โมเดลขนาดเล็กที่เน้นการใช้งานบนอุปกรณ์พกพา เช่น MobileNetV3 และ OFA-595 ผลการทดลองระบุว่าโมเดลขนาดเล็กเหล่านี้เมื่อผ่านกระบวนการเทรนด้วยวิธีดังกล่าว สามารถทำลายสถิติความแม่นยำเดิมบน ImageNet-1K ได้ดังสรุปในตารางที่ 2.1

ตารางที่ 2.1 สรุปชุดข้อมูลสำหรับโมเดลจำแนกภาพ

ชื่องานวิจัย	วัตถุประสงค์	วิธีการใช้	ผลลัพธ์
ImageNet-21K Pretraining for the Masses [29]	เพื่อพัฒนาแนวทางการ Pretraining โมเดล Deep Learning บนชุดข้อมูล ImageNet-21K ให้มี ประสิทธิภาพ เข้าถึงได้ง่าย และแก้ไขปัญหาความซับซ้อน ของโครงสร้างป้ายกำกับ เพื่อเพิ่มประสิทธิภาพการ ถ่ายโอนความรู้ไปยังงาน Downstream	ปรับปรุงและเตรียมชุดข้อมูล ImageNet-21K ด้วย กระบวนการ Preprocessing จนได้ ImageNet-21K-P (ตัดคลาส ที่มีข้อมูลน้อยกว่า 500 ภาพ, สร้าง Validation Set, Resize ภาพ)	วิธี Semantic Softmax ให้ผลดีกว่า วิธี Single-Label และ Multi-Label Pretraining อย่าง สม่ำเสมอในหลายงาน Downstream เช่น ImageNet-1K CIFAR- 100 MS-COCO และ Kinetics รวมถึงช่วย

ตารางที่ 2.1 สรุปชุดข้อมูลสำหรับโมเดลจำแนกภาพ (ต่อ)

ชื่องานวิจัย	วัตถุประสงค์	วิธีการใช้	ผลลัพธ์
		• ใช้โครงสร้างลำดับชั้นของ WordNet เพื่อแปลง Single-Label ให้เป็น Semantic Multi-Label • เสนอวิธีฝึกโมเดลใหม่ชื่อ Semantic Softmax โดยใช้ Softmax แยกตามลำดับชั้นเชิงความหมาย แทนการใช้ Single Softmax หรือ Multi-label Sigmoid • ผสาน Semantic Softmax กับ Semantic Knowledge Distillation เพื่อเพิ่มประสิทธิภาพ	เพิ่มประสิทธิภาพของโมเดลขนาดเล็กสำหรับอุปกรณ์พกพา เช่น MobileNetV3

2.11.3 ความรู้ด้านการออกแบบระบบ

งานวิจัยของ H. M Fadhil et al. [30] เรื่อง “Smartphone-Based Object Recognition Application for Educational Engagement” มีวัตถุประสงค์เพื่อพัฒนาแอปพลิเคชันบนสมาร์ทโฟนที่สามารถรู้จำวัตถุผ่านกล้องและใช้เป็นเครื่องมือทางการศึกษาเชิงโต้ตอบสำหรับผู้เรียน โดยเน้นการผสมผสานเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) และการเรียนรู้เชิงลึก (Deep Learning) เพื่อให้ผู้ใช้ โดยเฉพาะเด็ก ได้ฝึกการจำแนกวัตถุและเรียนรู้ชื่อสิ่งของต่างๆ ในลักษณะของการเล่นและการเรียนรู้ควบคู่กัน

แอปพลิเคชันดังกล่าวใช้แบบจำลอง MobileNetV2 ซึ่งเป็นโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ที่ได้รับการออกแบบมาเฉพาะสำหรับการทำงานบนอุปกรณ์พกพาที่มีทรัพยากรจำกัด ด้วยโครงสร้างแบบ Depthwise Separable Convolution และ Inverted Residuals ช่วยลดจำนวนพารามิเตอร์และเวลาในการประมวลผลได้อย่างมาก ขณะเดียวกัน ยังคงความแม่นยำสูงในการจำแนกภาพ (Top-1 Accuracy ประมาณ 94.7%) การพัฒนาแอปพลิเคชันนี้ ดำเนินการโดยการเก็บข้อมูลภาพจากผู้ใช้และจัดทำชุดข้อมูลลักษณะคล้าย ImageNet เพื่อใช้ฝึกโมเดล จากนั้นนำโมเดลที่ผ่านการฝึกมาเชื่อมต่อกับแอปพลิเคชันผ่านกล้องสมาร์ทโฟน เพื่อให้ระบบสามารถตรวจจับและจำแนกวัตถุได้แบบเรียลไทม์

ผลการทดสอบแสดงให้เห็นว่าโมเดล MobileNetV2 มีประสิทธิภาพดีกว่าโมเดลอื่น เช่น ShuffleNet และ NasNet ทั้งในด้านความเร็วและความถูกต้อง โดยใช้ทรัพยากรของสมาร์ทโฟนเพียงเล็กน้อย ซึ่งเหมาะสมอย่างยิ่งต่อการใช้งานในสภาพแวดล้อมจริง นอกจากนี้ ผู้วิจัยยังได้เสนอแนวทางต่อยอดในอนาคต เช่น การปรับให้ผู้ใช้สามารถเทรนโมเดลเพิ่มเติมได้ด้วยตนเอง การเชื่อมต่อกับระบบประมวลผลแบบคลาวด์ และการเพิ่มช่องทางการรับรู้แบบหลายรูปแบบ (Multimodal) เช่น เสียงหรือข้อความ เพื่อยกระดับประสบการณ์การเรียนรู้

ตารางที่ 2.2 สรุปความรู้ด้านการออกแบบระบบ

ชื่องานวิจัย	วัตถุประสงค์	วิธีการใช้	ผลลัพธ์
Smartphone-Based Object Recognition Application for Educational Engagement [28]	พัฒนาแอปพลิเคชันบนสมาร์ตโฟนที่ใช้เทคโนโลยี Deep Learning สำหรับการรู้จำวัตถุและส่งเสริมการเรียนรู้เชิงโต้ตอบ	ใช้แบบจำลอง MobileNetV2 ที่มีโครงสร้างแบบ Depthwise Separable Convolution และ Inverted Residuals ฝึกด้วยข้อมูลภาพที่เก็บจากผู้ใช้งานพัฒนาแอปพลิเคชันด้วย Flutter และใช้ Google ML Kit เพื่อการรู้จำวัตถุแบบเรียลไทม์ผ่านกล้องสมาร์ตโฟน	โมเดล MobileNetV2 ให้ความแม่นยำสูง (Top-1 Accuracy ~94.7%) ใช้พลังงานต่ำ ทำงานได้อย่างมีประสิทธิภาพบนอุปกรณ์พกพา เหมาะสำหรับนำไปใช้ในงานด้านการศึกษาและเรียนรู้แบบมีส่วนร่วม

2.11.4 พื้นฐานการออกแบบสำหรับผู้พิการทางสายตา

อ้างอิงจากงานวิจัยของปรเมศวร์ เบญจวรรณ [31] ที่ระบุถึงแนวทางการออกแบบโมบายแอปพลิเคชันในการเรียนรู้เชิงรุกสำหรับผู้พิการทางสายตา มุ่งเน้นการประเมินโมบายแอปการศึกษาตามแนวทางการออกแบบที่สังเคราะห์ขึ้นสำหรับผู้พิการทางสายตา เพื่อนำไปใช้เป็นเกณฑ์ในการตรวจแอปจริงในตลาดของประเทศไทย ได้มีการสังเคราะห์แนวทางการออกแบบแอปพลิเคชันให้เหมาะสมกับผู้พิการทางสายตานั้นผู้ที่มีสายตาเลือนราง โดยอ้างอิงมาตรฐานสากล เช่น W3C Web Content Accessibility Guidelines (WCAG 2.0) และ BBC Mobile Accessibility Standards and Guidelines ร่วมกับฉันทามติจากผู้เชี่ยวชาญ โดยแบ่งออกเป็น 4 ด้านหลัก ดังนี้

1) ด้านการเน้นผู้เรียนเป็นศูนย์กลาง มุ่งเน้นการปรับแต่งการแสดงผลและการโต้ตอบให้เหมาะสมกับข้อจำกัดทางสายตา

- ความต่างของสี (Contrast) สีของข้อความและพื้นหลังต้องตัดกันอย่างชัดเจน มีค่าอัตราส่วนความต่างสี (Contrast Ratio) อย่างน้อย 3:1 ถึง 4.5:1 คู่สีที่เหมาะสมเรียงตามลำดับคือ ข้อความขาวบนพื้นดำ ข้อความดำบนพื้นขาว ข้อความน้ำเงินบนพื้นเหลือง และข้อความดำบนพื้นชมพูอ่อน

- ขนาดตัวอักษร ผู้ใช้ต้องสามารถปรับขนาดข้อความได้ และควรมีขนาดระหว่าง 30-100 พอยต์ (หรือ 50-100 พอยต์ตามความเห็นผู้เชี่ยวชาญในบางบริบท) โดยไม่สูญเสียเนื้อหาหรือการทำงาน

- การใช้เสียง (Audio Interaction) ต้องมีระบบอ่านคำอธิบายภาพเคลื่อนไหว และกราฟิกด้วยเสียง, การแปลงข้อความเป็นเสียงพูด (Text-to-Speech) เมื่อผู้ใช้ต้องการ, การใช้เสียงอธิบายคำถามและคำตอบในบทเรียน และการสรุปประเด็นการเรียนรู้ด้วยเสียงทั้งก่อนและหลังบทเรียน

- การนำทาง (Navigation) ต้องมีกลไกการนำทางที่เป็นระบบและเหมือนกันในทุกหน้าจอ โดยรองรับการนำทางด้วยเสียง

2) ด้านการเรียนรู้แบบร่วมมือ (Collaborative Learning) สนับสนุนการติดต่อสื่อสารและการทำงานร่วมกันผ่านแอปพลิเคชัน

- การแจ้งเตือน (Notifications) ต้องสามารถรับรู้ได้ทั้งการมองเห็นและการได้ยิน
- การควบคุมเสียง ต้องแยกการควบคุมระดับความดังของเสียงประเภทต่างๆ ได้โดยอิสระ เช่น เสียงดนตรีประกอบ เสียงบรรยาย เสียงผู้บรรยาย และเสียงเอฟเฟกต์ เพื่อไม่ให้รบกวนการฟังเนื้อหาหลัก

- การแลกเปลี่ยนข้อมูล รองรับการแบ่งปันไฟล์เสียงบันทึก (Voice Notes) ระหว่างผู้เรียน การใช้เสียงในการแสดงความคิดเห็นในเว็บบอร์ด (Forum) และฟังก์ชันแปลงเสียงพูดเป็นข้อความ (Speech-to-Text) เพื่อการอธิบาย

3) ด้านการจูงใจผู้เรียนโดยไม่เน้นการวัดประเมินผล (Intrinsic Motivation) การออกแบบเพื่อสร้างแรงจูงใจและลดอุปสรรคในการใช้งาน

- การลดสิ่งรบกวน เนื้อหาต้องไม่มีการสั่นไหวหรือกะพริบมากกว่า 3 ครั้งต่อวินาที และไม่ควรมีการเล่นเสียงอัตโนมัติ (Auto-play) โดยผู้ใช้ไม่ได้สั่งงาน

- จุดสนใจ ส่วนประกอบที่มีปฏิสัมพันธ์ได้ (เช่น ปุ่ม) ต้องมีการเน้นจุดสนใจที่ชัดเจน ทั้งทางสายตาและเสียงอ่าน และต้องไม่ทำให้บริบทรอบข้างเปลี่ยนแปลงเมื่อได้รับจุดสนใจ

- พื้นที่ว่าง ปุ่มหรือส่วนที่กดได้ควรมีพื้นที่ว่างล้อมรอบ เพื่อให้ผู้ที่มีสายตาเลือนรางกดได้ง่ายและแม่นยำขึ้น

- การสรุปประเด็น มีฟังก์ชันสรุปประเด็นสำคัญระหว่างเรียนได้ด้วยการบันทึกเสียง

4) ด้านความรับผิดชอบในการเรียนของตนเอง (Self-Learning Responsibility) ส่งเสริมให้ผู้เรียนควบคุมการเรียนรู้ได้ด้วยตนเอง

- การควบคุมเวลา หากมีการจำกัดเวลา ต้องมีระบบเตือนและสามารถขยายเวลาหรือยกเลิกการจำกัดเวลาได้ (ยกเว้นกรณีสอบแข่งขันที่จำเป็น)

- ทางเลือกในการเรียนรู้ ผู้เรียนควรสามารถเลือกหัวข้อเรียน ข้ามไปเรียนหัวข้ออื่น หรือกลับมาเรียนต่อจากจุดเดิมได้

- การปรับความเร็วเสียง สื่อที่มีเสียงบรรยายต้องสามารถปรับความเร็วของเสียงพูดได้ตามความต้องการของผู้ใช้

- การยืนยันตัวตน รองรับการเข้าสู่ระบบด้วยลายนิ้วมือเพื่อความสะดวกแทนการพิมพ์รหัสผ่าน

นอกจากนี้ มีการเสนอกรอบปฏิบัติที่เชื่อมโยงระหว่างเกณฑ์เชิงออกแบบกับการประเมินเชิงประจักษ์ผ่านตารางประเมินและการวัดคอนทราสต์จากการใช้งานจริงในตลาดด้วย Colour Contrast Analyser (CCA) ซึ่งเป็นโปรแกรมสำหรับวัดความแตกต่างของสีระหว่างพื้นหน้าและพื้นหลังในงานดิจิทัล โดยจะคำนวณค่าสัดส่วนความต่างของความเข้มสีให้โดยอัตโนมัติ เพื่อให้แน่ใจว่าเนื้อหาสามารถอ่านและเข้าถึงได้โดยทุกคน รวมถึงผู้ที่มีความบกพร่องทางการมองเห็น จากผลการประเมินพบช่องว่างด้านการเข้าถึงของแอปพลิเคชันการศึกษาจำนวนมาก การประยุกต์แนวทางที่นำเสนอสามารถนำไปสู่การออกแบบต้นแบบที่เหมาะสมกว่า เช่น สีของข้อความสีขาวบนพื้นหลังสีดำ เพื่อให้มีศักยภาพในการยกระดับการเข้าถึงของผู้พิการทางสายตาในบริบทการใช้งานจริง

2.12 แอปพลิเคชันสำหรับผู้พิการทางสายตา

เสกสรร อามาศย์มนตรี และอัญชลี สุวัฑฒน [32] ระบุว่าสมาร์ทโฟนว่าเป็นเทคโนโลยีอำนวยความสะดวกสำหรับผู้เรียนที่มีความบกพร่องทางการมองเห็น โดยสามารถสรุปได้ดังนี้

ตารางที่ 2.3 สรุปแอปพลิเคชันสำหรับผู้พิการทางสายตา

ชื่อแอป	ผู้พัฒนา	ฟังก์ชันการทำงานของแอปพลิเคชัน
1. Seeing AI	Microsoft Corporation	ถ่ายภาพและอ่านข้อความหรือบรรยายลักษณะ และพูดออกเสียงที่ละภาษา

ตารางที่ 2.3 สรุปแอปพลิเคชันสำหรับผู้พิการทางสายตา (ต่อ)

ชื่อแอป	ผู้พัฒนา	ฟังก์ชันการทำงานของแอปพลิเคชัน
2. Lookout	Google LLC	ใช้ AI ตรวจสอบวัตถุรอบตัว และอ่านข้อความ มีหลายโหมดการใช้งาน พูดออกเสียงทีละภาษา ใช้งานออฟไลน์ได้
3. Pannana	ไทยพีบีเอส	เสียงบรรยายภาพสำหรับผู้พิการทางสายตา ใช้ฟังคำบรรยายประกอบสื่อวิดีโอ ภาพยนตร์
4. Read for the Blind	Thailand Association of The Blind	คลังหนังสือเสียงที่อ่านโดยอาสาสมัครเพื่อผู้พิการทางสายตา
5. Be My Eyes	Be My Eyes (แพลตฟอร์มอาสาสมัคร)	ใช้ระบบวิดีโอคอลเชื่อมต่อผู้พิการทางสายตากับอาสาสมัครแบบออนไลน์

จากตารางที่ 2.3 สามารถสรุปโดยสังเขปว่า Seeing AI (Microsoft) แอปตรวจสอบสิ่งของและอ่านข้อความออกเสียง แต่ทำงานบน Cloud และต้องเชื่อมต่ออินเทอร์เน็ต สำหรับ Lookout (Google) ทำงานใกล้เคียงกัน แต่ยังไม่รองรับการออกเสียงหลายภาษาแบบต่อเนื่องกัน และในส่วนของงานอื่นๆ เน้น Object Detection บนมือถือ แต่ยังขาด Multilingual TTS หรือความสามารถในการทำงานแบบออฟไลน์ หรือยังพึ่งพาการช่วยเหลือโดยอาสาสมัคร

บทที่ 3

วิธีการวิจัย

3.1 บทนำ

งานวิจัยนี้มุ่งพัฒนาแอปพลิเคชันเพื่อผู้พิการทางสายตา โดยเน้นการประมวลผลภาพแบบออฟไลน์เพื่อประสิทธิภาพด้านความเร็วและความเป็นส่วนตัว เลือกใช้สถาปัตยกรรม MobileNetV3 ที่ผ่านการฝึกสอนด้วยชุดข้อมูล ImageNet เนื่องจากมีความสมดุลระหว่างความแม่นยำและการใช้ทรัพยากรบนอุปกรณ์พกพา ระบบจะจำแนกวัตถุและเชื่อมโยงผลลัพธ์เข้ากับฐานข้อมูลคำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) จำนวนประมาณ 6,000-10,000 คำ ซึ่งประกอบด้วยภาษาไทย 8,500 คำ ภาษาญี่ปุ่น 6,000 คำ ภาษาอังกฤษ 10,000 คำ โดยประมาณ พร้อมระบบสังเคราะห์เสียงเพื่อส่งเสริมการเรียนรู้ภาษาต่างประเทศเบื้องต้นสำหรับผู้พิการทางสายตาในฐานะที่เป็นแอปพลิเคชันต้นแบบ

3.2 ช่องว่างของงานวิจัย (Research Gap)

จากการทบทวนและวิเคราะห์แอปพลิเคชันและงานวิจัยที่เกี่ยวข้อง ซึ่งพัฒนาเพื่อช่วยเหลือและส่งเสริมการเรียนรู้สำหรับผู้พิการทางสายตา พบว่าเทคโนโลยีในปัจจุบันยังมีข้อจำกัดในด้านการส่งเสริมการเรียนรู้ภาษาหลายรูปแบบ โดย Seeing AI และ Lookout เน้นการอ่านออกเสียงภาษาหลักเพียงภาษาเดียว ขณะที่ Pannana และ Read for the Blind มุ่งเน้นสื่อบันทึกและคลังหนังสือเสียงซึ่งไม่รองรับการจำแนกวัตถุแบบเรียลไทม์ ส่วน Be My Eyes ยังต้องพึ่งพาอาสาสมัครซึ่งมีข้อจำกัดด้านความเป็นส่วนตัว

สำหรับงานวิจัยของ H. M Fadhil et al. [30] แม้จะใช้เทคโนโลยี MobileNet เพื่อการเรียนรู้เชิงโต้ตอบ แต่ยังคงเน้นการรับรู้เชิงภาพเป็นหลัก VocabEyes จึงได้ต่อยอดแนวคิดดังกล่าวโดยการประยุกต์ใช้ระบบสังเคราะห์เสียง (TTS) สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) เพื่อเปลี่ยนการรู้จำวัตถุเป็นการเรียนรู้เชิงเสียง (Audio-Based Learning) ซึ่งตอบโจทย์ด้านการเข้าถึง (Accessibility) และการส่งเสริมทักษะภาษาสำหรับผู้พิการทางสายตาโดยเฉพาะ โดยสามารถสรุปได้ดังตารางที่ 3.1

ตารางที่ 3.1 การเปรียบเทียบระบบช่วยเหลือผู้พิการทางสายตากับงานวิจัยของเรา

ชื่อระบบ	ฟังก์ชันการทำงาน	ข้อจำกัดหรือจุดที่ต่างจาก VocabEyes
Seeing AI/Lookout	ตรวจจำวัตถุและอ่านข้อความ	รองรับเพียงภาษาเดียวไม่เน้นการเรียนรู้ภาษา
Pannana/Read for the Blind	เสียงบรรยายภาพและหนังสือเสียง	ไม่รองรับการตรวจจำวัตถุแบบเรียลไทม์

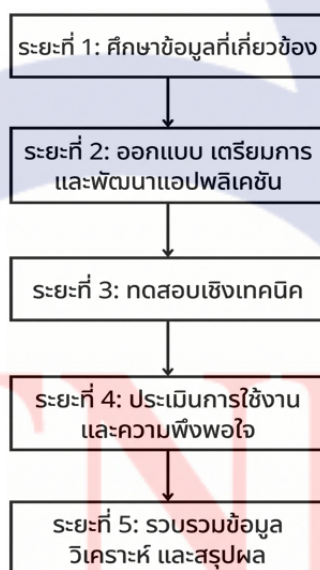
ตารางที่ 3.1 การเปรียบเทียบระบบช่วยเหลือผู้พิการทางสายตากับงานวิจัยของเรา (ต่อ)

ชื่อระบบ	ฟังก์ชันการทำงาน	ข้อจำกัดหรือจุดที่ต่างจาก VocabEyes
Be My Eyes	อาสาสมัครช่วยเหลือผ่านวิดีโอคอล	พึ่งพามนุษย์ และมีความเสี่ยงด้านความเป็นส่วนตัว
AEYE	เรียนรู้ภาษาผ่านการจำแนกวัตถุ	ใช้ TTS สามภาษา และเน้นการเข้าถึงเชิงเสียง

งานวิจัยนี้จึงมุ่งพัฒนา VocabEyes เพื่อปิดช่องว่างดังกล่าว โดยการรวมเทคโนโลยี MobileNetV3 และ TTS เข้าด้วยกัน เพื่อสร้างระบบจำแนกวัตถุที่สามารถอ่านออกเสียงได้ต่อเนื่อง 3 ภาษา (ไทย-ญี่ปุ่น-อังกฤษ) บนอุปกรณ์พกพาแบบออฟไลน์

3.3 กระบวนการวิจัย

งานวิจัยนี้เป็นการวิจัยและพัฒนา (Research & Development, R&D) เชิงทดลอง โดยหลักแบ่งเป็น 5 ระยะหลักดังรูปที่ 3.1



รูปที่ 3.1 ภาพรวมกระบวนการวิจัย

ระยะที่ 1 การศึกษาข้อมูลที่เกี่ยวข้องและกำหนดความต้องการของระบบ (Phase 1) ศึกษาข้อมูลที่เกี่ยวข้อง เตรียมความพร้อมและหาแนวทางในการทำวิจัยรวมถึงกำหนดความต้องการของระบบ

ระยะที่ 2 การออกแบบ เตรียมการ และพัฒนาแอปพลิเคชัน (Phase 2) เช่น กำหนดข้อกำหนดและคุณสมบัติหลักของระบบ (Requirements & Features), สร้างคลังคำศัพท์สามภาษา, สร้าง User

Interface (UI) เพื่อการเข้าถึง, บุรณาการโมเดล TFLite และ TTS ไทย-ญี่ปุ่น-อังกฤษ, ทดลองรัน และ แก้ไขบั๊ก ฯลฯ

ระยะที่ 3 การทดสอบเชิงเทคนิค (Phase 3) (Technical Evaluation) วัดความถูกต้องด้วย Top-1 Accuracy และ Top-k Accuracy

ระยะที่ 4 การประเมินการใช้งานและความพึงพอใจ (Usability & Satisfaction Evaluation) โดยผู้พิการทางสายตา หรืออาสาสมัครเพื่อทดลองใช้งานจริง และทำการประเมินแบบสอบถามความสามารถในการใช้งานและความพึงพอใจ รวมถึงเก็บข้อเสนอเชิงคุณภาพ

ระยะที่ 5 รวบรวมข้อมูล วิเคราะห์และสรุปผล ทำการรวบรวมข้อมูลในด้านต่างๆ นำมาเปรียบเทียบ วิเคราะห์ พร้อมทั้งสรุปผล รวมถึงเสนอแนวทางแก้ไข

ระยะที่ 1 การศึกษาข้อมูลที่เกี่ยวข้อง และกำหนดความต้องการของระบบ

ศึกษาข้อมูลที่เกี่ยวข้องกับพื้นฐานต่างๆ เช่น การจำแนกภาพ การออกแบบ การพัฒนาแอปพลิเคชัน ข้อมูลเกี่ยวกับ Accessibility สำหรับผู้พิการทางสายตา จากงานวิจัยหรือข้อมูลที่เกี่ยวข้อง เพื่อใช้ในการออกแบบและพัฒนาแอปพลิเคชัน และนำรายละเอียดมาเรียบเรียงลงในบทที่ 2 ทฤษฎี หลักการ และงานวิจัยที่เกี่ยวข้อง รวมถึงกำหนดความต้องการของระบบ

3.3.1 กรอบแนวคิดการวิจัย (Conceptual Framework)

กรอบแนวคิดของงานวิจัยนี้ประกอบด้วยองค์ประกอบหลักดังนี้

1. ปัจจัยนำเข้า (Input) คือองค์ประกอบหรือเทคโนโลยีที่นำมาใช้สร้างระบบ

Deep Learning Model การใช้โมเดลสำเร็จรูปเพื่อใช้ในการรู้จำวัตถุที่รวดเร็วและแม่นยำบนสมาร์ตโฟน ผู้วิจัยเลือกใช้การจำแนกภาพแทน Object Detection เพื่อดึงเฉพาะวัตถุที่เด่นชัดที่สุดในภาพมาเป็นคำตอบหลัก ซึ่งช่วยลดความซับซ้อนและป้องกันไม่ให้ผู้พิการทางสายตาได้รับข้อมูลเสียงที่มากเกินไปจนสับสน

- Multilingual Data Mapping การใช้ฐานข้อมูล Dictionary.csv เพื่อเชื่อมโยง Label ที่ได้จากการประมวลผลของโมเดลไปยังคำศัพท์ 3 ภาษา (ไทย-ญี่ปุ่น-อังกฤษ)

- Assistive Technology Features

- ระบบการสังเคราะห์เสียง (Text-to-Speech : TTS)

- การออกแบบ UI/UX แบบ Accessibility (Touch Interaction: Long Press, Single Tap)

2. กระบวนการ (Process) คือขั้นตอนที่แอปพลิเคชันสำหรับการจัดการกับข้อมูล

- แปลงข้อมูลภาพ (Image) เป็นข้อมูลป้ายระบุชื่อ (Label)

- ประมวลผลและลำดับการอ่านออกเสียง (Thai → Japanese → English)

- จัดการขัดจังหวะเสียง (Audio Interrupt Logic) เพื่อความคล่องตัวของผู้ใช้
- 3. ผลลัพธ์ (Output/Outcome) คือสิ่งที่ต้องการวัดผลจากการพัฒนาแอปพลิเคชันนี้
 - ประสิทธิภาพของระบบ (System Performance) ความแม่นยำในการระบุวัตถุ

(Accuracy)

- ความสามารถในการใช้งานและความพึงพอใจของผู้ใช้ (Usability & Satisfaction)

คะแนนจากแบบประเมิน System Usability Scale (SUS) และแบบสอบถามความพึงพอใจต่อการใช้งานแอปพลิเคชัน VocabEyes

กล่าวโดยสรุปคืองานวิจัยนี้เป็นการประยุกต์ใช้เทคโนโลยีที่มีอยู่แล้ว เช่น การรู้จำวัตถุ (Object Recognition) ร่วมกับการบริหารจัดการข้อมูลหลายภาษา (Data Mapping) เพื่อสร้างเครื่องมือช่วยการเรียนรู้คำศัพท์ที่เข้าถึงได้ง่ายและตอบสนองต่อพฤติกรรมการใช้งานของผู้พิการทางสายตา

3.3.2 ข้อกำหนดและคุณสมบัติหลัก (Requirement & Features)

กำหนด Requirements ของแอปพลิเคชันโดยแบ่งเป็นดังนี้

1. Functional Requirements

1.1 Object Recognition สามารถจำแนกวัตถุจากภาพถ่ายหรือรูปถ่ายจากกล้องมือถือได้

2. Vocaburary Output สามารถประมวลผลและแสดงผลคำศัพท์ออกมาได้ อย่างน้อย 1 ภาษาจาก 3 ภาษา (ไทย, ญี่ปุ่น, อังกฤษ) เนื่องจากข้อจำกัดด้านคลังคำศัพท์จาก WordNet

3. Text-to-Speech (TTS) สามารถอ่านข้อความเป็นเสียงคำศัพท์เพื่อสื่อสารกับผู้ใช้ได้

2. Non-Functional Requirements

Offline Capability สามารถทำงานแบบออฟไลน์ได้โดยไม่ต้องพึ่งพาอินเทอร์เน็ตขณะใช้งาน (On-device processing) เพื่อความเป็นส่วนตัวและความเร็ว สามารถประหยัดค่าใช้จ่าย User Features

2.1 Full-Screen Gesture การสั่งงานด้วยการกด หรือกดค้างที่หน้าจอทั่วทั้งบริเวณได้ แทนการหาเพื่อกดปุ่มบนหน้าจอ เพื่อให้สะดวกต่อการใช้งานของผู้พิการทางสายตา

2.2 High Contrast UI สำหรับผู้ที่มีสายตาเลือนรางเพื่อให้มองเห็นได้ชัดเจน

2.3 Tutorial Guide ระบบเสียงแนะนำการใช้งานแอปตั้งแต่เริ่มต้นเปิดโปรแกรม เพื่อให้ผู้พิการทางสายตาสามารถเข้าใจวิธีการใช้งานแอปพลิเคชันได้โดยง่าย

ตารางที่ 3.2 สรุปข้อกำหนดและคุณสมบัติหลัก

ประเภท		หัวข้อ	รายละเอียดข้อกำหนด
Requirements	Functional Requirements	Object Recognition	ระบบต้องสามารถจำแนกวัตถุจากภาพถ่ายหรือภาพที่เลือกจากแกลเลอรีได้
		Vocabulary Output	ระบบต้องสามารถประมวลผลและแสดงผลคำศัพท์ได้อย่างน้อย 1 จาก 3 ภาษา ได้แก่ ภาษาไทย ภาษาอังกฤษ และภาษาญี่ปุ่น
		Text-to-Speech (TTS)	ระบบต้องสามารถอ่านออกเสียงคำศัพท์เป็นเสียงอ่านเพื่อสื่อสารกับผู้ใช้ได้
	Non-Functional Requirements	Offline Capability	ระบบต้องทำงานแบบ On-Device Processing โดยไม่ต้องพึ่งพาอินเทอร์เน็ต เพื่อความเสถียรและความเป็นส่วนตัว ประหยัดค่าใช้จ่าย
User Features		Full-Screen Gesture	การสั่งงานด้วยการกด หรือกดค้างที่หน้าจอทั่วทั้งบริเวณได้แทนการกดปุ่ม
		High Contrast UI	การออกแบบส่วนต่อประสานผู้ใช้ให้มีสีตัดกันชัดเจน เพื่อรองรับกลุ่มผู้ที่มีสายตาเลือนราง (Low Vision)
		Tutorial Guide	มีคำแนะนำการใช้งาน (Tutorial) ตั้งแต่เริ่มเปิดโปรแกรม

3.3.3 ขั้นตอนการวิจัย

1. ศึกษาข้อมูลที่เกี่ยวข้องกับพื้นฐานด้านการจำแนกภาพ เช่น ข้อมูลที่เกี่ยวข้องกับการออกแบบ และพัฒนาแอปพลิเคชัน, และข้อมูลเกี่ยวกับ Accessibility สำหรับผู้พิการทางสายตาจากงานวิจัย หรือข้อมูลที่เกี่ยวข้อง เพื่อใช้ในการออกแบบและพัฒนาแอปพลิเคชัน
2. จัดทำหรือรวบรวมข้อมูลพจนานุกรม อังกฤษ-ไทย-ตามรายการใน MobileNet Pretrained Imagenet จำนวนประมาณ 6,000-10,000 คำ ตามโดยอ้างอิงข้อมูลคำศัพท์จาก WordNet โดยใช้ Google Colab
3. แปลงไฟล์โมเดลสำเร็จรูป MobileNetV3 Large Pretrained ImageNet 21K-P จากไฟล์ในรูปแบบ .pth เป็น .tflite ของ TensorFlow เพื่อนำมาใช้ปรับใช้ในแอปพลิเคชันต้นแบบที่สร้างขึ้นโดยใช้โปรแกรม Flutter
4. พัฒนาแอปพลิเคชัน โดยใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ (Generative AI) เป็นตัวช่วยสร้างโค้ด ทั้งในส่วนติดต่อผู้ใช้ (User Interface) และโปรแกรมเบื้องหลัง เพื่อให้ได้มาซึ่งแอปพลิเคชัน

ตามวัตถุประสงค์ โดยทำการทดลอง และปรับแก้ข้อผิดพลาดที่เกิดขึ้น จนกว่าจะได้ผลลัพธ์เป็นแอปพลิเคชันที่พร้อมใช้งาน

5. Output แอปพลิเคชันที่พัฒนาแล้ว ออกมาในรูปแบบของไฟล์ .apk เพื่อติดตั้งในสมาร์ตโฟนเครื่องต้นแบบที่กำหนด

6. ประเมินความแม่นยำเบื้องต้นของแอปพลิเคชันรู้จำวัตถุที่พัฒนาขึ้นตามเงื่อนไขที่กำหนด รวบรวม และบันทึกข้อมูลเก็บไว้ในไฟล์อิเล็กทรอนิกส์

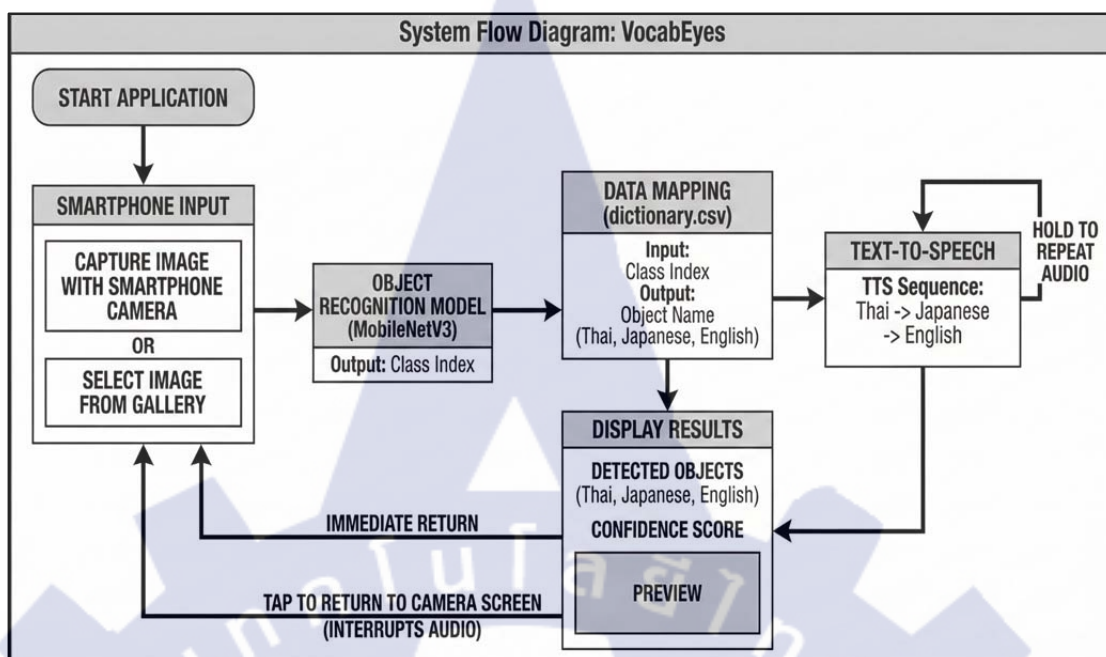
7. ประเมินความสามารถในการใช้งานหรือความพึงพอใจของกลุ่มตัวอย่างเป้าหมาย ผ่านการทำแบบสอบถาม และ/หรือการสัมภาษณ์ รวบรวม และบันทึกข้อมูลเก็บไว้ในไฟล์อิเล็กทรอนิกส์

8. วิเคราะห์ข้อมูล สรุปรวบรวมประเด็นปัญหา พร้อมเสนอแนวทางการแก้ไข

ระยะที่ 2 การออกแบบ เตรียมการ และพัฒนาแอปพลิเคชัน

3.3.4 การออกแบบขั้นตอนการทำงานของแอปพลิเคชัน

1. Smartphone Input ผู้ใช้เลือกใช้งานผ่านการถ่ายภาพสดหรือเลือกจาก Gallery
2. Object Recognition (MobileNetV3) ประมวลผลด้วยโมเดล MobileNetV3 Large เพื่อจำแนกสิ่งของจากภาพ ซึ่งเป็นใช้โมเดลที่ฝึกบนชุดข้อมูล ImageNet 21-K-P และส่งค่าผลลัพธ์ออกมาเป็น Class Index
3. Data Mapping (Dictionary Mapping):
 - ระบบนำ Class Index ไปค้นหาในไฟล์คลังคำศัพท์รูปแบบไฟล์ .csv ที่จัดทำไว้ โดยการจับคู่คำศัพท์ 3 ภาษา ได้แก่ ภาษาไทย ญี่ปุ่น และภาษาอังกฤษด้วย WordNet ID
4. Display & TTS Output
 - แสดงผลลัพธ์บนหน้าจอ (Object Name & Confidence Score)
 - ระบบ TTS อ่านออกเสียงภาษาไทย ภาษาญี่ปุ่น ภาษาอังกฤษตามลำดับ
5. User Interaction
 - Long Press (กดค้าง) เพื่อฟังเสียงอ่านซ้ำ (Repeat Audio)
 - Single Tap (กดอีกครั้ง) เพื่อ Interrupt (หยุดการอ่านทันที) และกลับไปหน้าจอรับภาพ (Camera Screen) เพื่อเริ่มการทำงานใหม่



รูปที่ 3.2 ภาพรวมการไหลของระบบ (System Flow)

3.3.5 การออกแบบเครื่องมือที่ใช้และองค์ประกอบหลักของระบบ

1. ฮาร์ดแวร์

สมาร์ทโฟน ระบบปฏิบัติการ Android ระดับกลางขึ้นไป (RAM $\geq$ 4GB) ที่มีกล้องคุณภาพเหมาะสมสำหรับงานภาพนิ่ง ยี่ห้อ : Samsung รุ่น: Galaxy Note 20

2. ซอฟต์แวร์และไลบรารี

Flutter (ภาษา Dart) สำหรับพัฒนา UI/UX ที่เป็นมิตรต่อการเข้าถึง เช่น ปุ่มขนาดใหญ่ ใช้งานมือเดียวได้สะดวก และรองรับการใช้งานแบบออฟไลน์ตามขอบเขตงาน

โมเดลสำเร็จรูป MobileNetV3 Large (TensorFlow Lite) ที่ฝึกบนชุดข้อมูล ImageNet 21K-P (11,221 Classes) เพื่อการรู้จำระดับ คำหรือคลาส

ไฟล์พจนานุกรม ที่เป็น Label mapping ตามรายการคลาสที่อยู่ใน ImageNet ของโมเดล Mobile Net V3 Large pretrained ImageNet 21-K-P ภาษาไทย-ญี่ปุ่น-อังกฤษ

โมดูล Text-to-Speech (TTS) สำหรับออกเสียงผลลัพธ์เป็นภาษาไทย ภาษาญี่ปุ่น และอังกฤษ ตามลำดับ

3. องค์ประกอบหลักของระบบ

- 1) โมดูลจับภาพนิ่ง (Image Capture)
- 2) โมดูลอนุมาน MobileNetV3 (TFLite Interpreter)

3) โมดูลจับคู่คำศัพท์ (Dictionary Mapping EN-TH-JA $\approx$ 6,000-10,000 รายการ ตามข้อมูลวัตถุใน ชุดข้อมูล Imagenet)

4) โมดูลสังเคราะห์เสียงสามภาษา (TTS Pipeline)

5) โมดูลบันทึกผลทดสอบและเวลาแฝง (Logging)

3.3.6 ประชากรและกลุ่มตัวอย่าง

1. ประชากร ผู้พิการทางสายตาในประเทศไทยที่สามารถใช้งานสมาร์ทโฟนขั้นพื้นฐานได้ ทั้งผู้ที่ตาบอดสนิทและผู้ที่มีสายตาเลือนราง ซึ่งเป็นกลุ่มประชากรเป้าหมายหลักสำหรับการนำแอปพลิเคชันไปใช้งานจริงในอนาคต

2. กลุ่มตัวอย่าง สำหรับการประเมินความสามารถในการใช้งานระยะเริ่มต้น (Preliminary Usability Study) ในงานวิจัยนี้ ใช้วิธีการเลือกกลุ่มตัวอย่างแบบเจาะจง โดยใช้กลุ่มตัวอย่างเป็น อาสาสมัครสายตาปกติที่จำลองการปิดกั้นการมองเห็น (Blindfolded Normal-Vision Volunteers) จำนวน 5 คน เพื่อทดสอบการใช้งานแอปพลิเคชันต้นแบบ

ทั้งนี้ การใช้กลุ่มตัวอย่างจำนวน 5 คน เป็นไปตามหลักการของ Nielsen Norman Group (NN/g) ที่ระบุว่าผู้ทดสอบจำนวน 5 คน ถือเป็นจำนวนที่เพียงพอสำหรับการค้นพบปัญหาด้านความสามารถในการใช้งาน (Usability) เชิงคุณภาพในระยะเริ่มต้น (Preliminary Phase) อย่างมีประสิทธิภาพ [33]

3.3.7 เครื่องมือวิจัย

1. คอมพิวเตอร์โน้ตบุ๊ก สำหรับใช้งานในงานวิจัย ยี่ห้อ : Apple รุ่น : MacBook Air M3 ระบบปฏิบัติการ: MacOS

2. Flutter Application สำหรับพัฒนา/Debug Application

3. Android Studio Application สำหรับพัฒนา/Debug Application

4. สมาร์ทโฟนระบบปฏิบัติการ Android เครื่องที่กำหนดสำหรับทดลองรันโปรแกรมที่สร้างขึ้นรวมถึงการ Debug ยี่ห้อ : Samsung รุ่น: Galaxy Note 20

5. โมดูลสำเร็จรูป MobileNet V3 Pretrained ImageNet 21K-P ใช้จำแนกวัตถุ

6. Speech Recognition and Synthesis from Google : Google TTS ใช้ในการอ่านออกเสียงสามภาษา ไทย ญี่ปุ่น อังกฤษ แบบออฟไลน์

7. Google Colab สำหรับแปลงไฟล์ .pth เป็น .tflite และ Mapping คำศัพท์

8. โปรแกรมที่เกี่ยวข้องในการจัดทำข้อมูล เช่น Word, Excel เป็นต้น

9. Generative AI ใช้ในการช่วยสร้างแอปพลิเคชัน ฯลฯ

10. อินเทอร์เน็ตสำหรับค้นคว้าข้อมูลต่างๆ
11. แบบสอบถาม SUS หรืออื่นๆ เพื่อประเมินความสามารถในการใช้งานและความพึงพอใจต่อการใช้งานแอปพลิเคชันต้นแบบ เป็นต้น

3.3.8 ข้อจำกัดของการวิจัย

1. ใช้โมเดลสำเร็จรูป MobileNetV3 ที่ฝึกฝนแล้วบนชุดข้อมูล Imagenet 21K-P (11,221 คลาส) จึงไม่ครอบคลุมวัตถุเฉพาะทางทั้งหมด
2. ประมวลผลเฉพาะภาพนิ่ง ไม่ใช่ Real-Time Video
3. พจนานุกรม EN-TH-JA ครอบคลุม ประมาณ 6,000-10,000 คำ ตามรายการชุดข้อมูล Imagenet 21K-P (11,221 classes) ในโมเดลสำเร็จรูปที่ได้รับการฝึกมาแล้ว โดยอ้างอิงคำศัพท์จาก WordNet ในแต่ละภาษา ซึ่งประกอบด้วยภาษาไทย 8,500 คำ ภาษาญี่ปุ่น 6,000 คำ ภาษาอังกฤษ 10,000 คำ โดยประมาณ จึงยังไม่ครอบคลุมบริบทหลากหลาย

3.3.9 เงื่อนไขการประเมินความแม่นยำของแอปพลิเคชันต้นแบบ

การประเมินความแม่นยำของแอปพลิเคชันต้นแบบ VocabEyes ในงานวิจัยนี้กำหนดเงื่อนไขการทดสอบ (Test conditions) ให้ใกล้เคียงกับสถานการณ์ใช้งานจริงของผู้ใช้ทั่วไปมากที่สุด ภายใต้ข้อจำกัดด้านเวลาและทรัพยากร โดยมุ่งประเมินความสามารถของระบบในการรู้จำวัตถุจากภาพถ่ายที่ได้จากกล้องโทรศัพท์มือถือ และรายงานผลลัพธ์ในรูปแบบ Top-1 และ Top-k เช่น Top-3 หรือ Top-5 ตามลำดับการจัดอันดับผลทำนายของแบบจำลอง

3.3.9.1 สถานการณ์ที่ใช้ในการประเมิน (Evaluation Scenarios)

การทดสอบถูกออกแบบให้ครอบคลุม 2 กลุ่มสถานการณ์หลัก ได้แก่ Simple Case และ Complex Case เพื่อสะท้อนระดับความยากง่ายของภาพถ่ายที่เกิดขึ้นจริงในการใช้งาน เช่น ภาพที่มีพื้นหลังซับซ้อน วัตถุถูกบัง วัตถุมีขนาดเล็ก หรือสภาพแสงไม่เอื้ออำนวย โดยมีรายละเอียดแต่ละกรณีดังนี้

(1) สถานการณ์ถ่ายภาพแบบเรียบง่าย (Simple Case) ไม่มีสิ่งกีดขวางหรือฉากหลังเรียบ เป็นกรณีพื้นฐานที่วัตถุเป้าหมายอยู่เด่นชัดในภาพ ไม่มีวัตถุอื่นบดบัง และฉากหลังค่อนข้างเรียบหรือมีสิ่งรบกวนน้อย สำหรับใช้เป็นค่ามาตรฐานอ้างอิง (Baseline) เพื่อดูศักยภาพของระบบในสถานะที่เหมาะสมที่สุด โดยลักษณะการถ่ายภาพที่กำหนดมีดังนี้

- วัตถุอยู่กึ่งกลางภาพหรืออยู่ในเฟรมอย่างชัดเจน
- เห็นวัตถุครบทั้งชิ้น (Full Object Visible)

- ฉากหลังเรียบ เช่น ผนัง โต๊ะเรียบ พื้นเรียบ
- แสงปกติ ไม่มีมืดหรือสว่างจ้าจนเกินไป

(2) สถานการณ์ภาพถ่ายที่ซับซ้อน (Complex Case) แบ่งเป็น 5 กรณีย่อย เพื่อจำลองปัจจัยที่ทำให้ระบบรู้จำยากขึ้นในการใช้งานจริง ดังนี้

2.1) วัตถุเดียวมีพื้นหลังรก (Single Object with Background) มีวัตถุเป้าหมายเพียงหนึ่งชิ้น แต่พื้นหลังมีรายละเอียดมาก เช่น มีลวดลาย มีสิ่งของรอบข้าง หรือมีความคล้ายคลึงกับวัตถุ เพื่อตรวจสอบความทนทานของระบบต่อ Background Clutter และความสามารถในการแยกแยะวัตถุจากฉากหลัง โดยลักษณะภาพถ่ายที่กำหนดมีดังนี้

- วัตถุเป้าหมาย 1 ชิ้น แต่พื้นหลังมีสิ่งของ หรือลวดลายชัดเจน
- อาจมีสีใกล้เคียงกับวัตถุ หรือมีรูปทรงคล้ายกันในฉากหลัง

2.2) เห็นวัตถุบางส่วน (Partial Occlusion) วัตถุเป้าหมายถูกบังบางส่วน (เช่น ถูกมือจับบัง ถูกสิ่งของอื่นบัง หรืออยู่ขีดขอบภาพ) ทำให้เห็นไม่ครบทั้งชิ้น

เพื่อประเมินความสามารถของระบบเมื่อข้อมูลภาพของวัตถุไม่สมบูรณ์ ซึ่งเกิดขึ้นจริงบ่อยในการใช้งาน โดยมีลักษณะการภาพถ่ายที่กำหนดดังนี้

- วัตถุถูกบังหรือเห็นบางส่วนประมาณ 70-80% เพื่อความสมจริง และไม่ยากเกินไป
- ยังพอเห็นลักษณะเด่นบางส่วนของวัตถุ เช่น ขอบ รูปร่าง หรือ ส่วนประกอบสำคัญ

2.3) วัตถุหลายอย่างในภาพ (Multiple Objects) ในภาพมีหลายวัตถุอยู่ด้วยกัน ทั้งวัตถุเป้าหมายและวัตถุอื่น ทำให้เกิดการแย่งความสนใจของแบบจำลอง เพื่อประเมินความแม่นยำของระบบในภาพที่มีความหนาแน่นของวัตถุ (Object Clutter) และดูโอกาสที่ระบบจะหายเป็นวัตถุอื่น โดยลักษณะการภาพถ่ายที่กำหนดมีดังนี้

- มีอย่างน้อย 3 วัตถุในเฟรมถ่ายภาพ
- วัตถุเป้าหมายยังปรากฏชัดเจนแต่ไม่จำเป็นต้องเด่นที่สุด

2.4) วัตถุเล็ก พื้นหลังใหญ่ (Small Object in Large Background) วัตถุเป้าหมายมีขนาดเล็กเมื่อเทียบกับพื้นที่ภาพ เช่น อยู่ไกล หรืออยู่มุมใดมุมหนึ่ง ทำให้รายละเอียดวัตถุมีน้อย เพื่อประเมินความสามารถของระบบต่อขนาดวัตถุที่เล็ก (Small Object Challenge) ซึ่งมักทำให้รู้จำผิด โดยลักษณะการภาพถ่ายที่กำหนดมีดังนี้

- วัตถุเป้าหมายกินพื้นที่ภาพน้อย เช่น 30% ของเฟรม
- ยังสามารถมองเห็นรูปร่างของวัตถุได้ แต่รายละเอียดลดลงอย่าง

ชัดเจน

2.5) ไฟสลัวหรือแสงน้อย (Low-Light) ถ่ายในสภาพแสงน้อย เช่น ห้องไฟสลัว ช่วงเย็น หรือพื้นที่ที่มีแสงไม่เพียงพอ ทำให้ภาพมีน้อยสีหรือความคมชัดลดลง เพื่อประเมินความคงทนของระบบต่อคุณภาพภาพที่ลดลงจากปัจจัยแสง โดยลักษณะภาพถ่ายที่กำหนดมีดังนี้

- แสงน้อยกว่าปกติอย่างชัดเจน แต่ยังเห็นวัตถุได้จริง (ไม่มีดจนมองไม่ออก)
- อาจเกิด Noise หรือความเบลอล็กน้อยตามธรรมชาติของกล้อง

3.3.9.2 ตัวอย่างและคำจำกัดความของแต่ละกรณีโดยสังเขป

- S1 หมายถึง Simple Case ไม่มีสิ่งกีดขวาง/ฉากหลังเรียบ
- C1 หมายถึง Complex Case วัตถุเดียวมีพื้นหลังรบกวน
- C2 หมายถึง Complex Case เห็นวัตถุบางส่วน/ถูกบัง
- C3 หมายถึง Complex Case วัตถุหลายอย่างในภาพ
- C4 หมายถึง Complex Case วัตถุเล็ก พื้นหลังใหญ่
- C5 หมายถึง Complex Case ไฟสลัว/แสงน้อย

3.3.10 วัตถุเป้าหมายและจำนวนภาพทดสอบ

เพื่อให้สอดคล้องกับข้อจำกัดด้านเวลา งานวิจัยนี้กำหนดให้ทดสอบด้วย วัตถุเป้าหมายจำนวน 10 วัตถุ ซึ่งเน้นหลักเป็นสิ่งที่พบได้ทั่วไปภายในบ้านหรือในชีวิตประจำวัน เพื่อสะท้อนการใช้งานจริง (Real-world usage) โดยมีวัตถุเป้าหมายการถ่ายภาพ คือ 1. Curtain (ผ้าม่าน) 2. Cat (แมว) 3. Apple (แอปเปิ้ล) 4. Pineapple (สับปะรด) 5. Tree (ต้นไม้) 6. Bicycle (จักรยาน) 7. Electric Fan (พัดลม) 8. Iron (เตารีด) 9. Egg (ไข่) 10. Broom (ไม้กวาด)

วัตถุทั้ง 10 ชนิด ถูกคัดเลือกโดยอ้างอิงจากทฤษฎี Occupational Therapy Practice Framework: Domain and Process (OTPF-4) ซึ่งเป็นกรอบแนวคิดมาตรฐานของ American Occupational Therapy Association (AOTA) และ Haptic Perception ซึ่งครอบคลุมถึงกิจวัตรประจำวัน (Activities of Daily Living : ADLs/ Instrumental Activities of Daily Living : IADLs), การแยกแยะด้วยการสัมผัส (Tactile Discrimination) และการรับรู้ทิศทางและการเคลื่อนไหว (Orientation & Mobility) การวิเคราะห์วัตถุตัวอย่างตามกรอบแนวคิด OTPF-4 และการรับรู้ผ่านการสัมผัส และการรับรู้ทิศทางและการเคลื่อนไหว

1. ด้านกิจกรรมดำเนินชีวิต

วัตถุที่เลือกครอบคลุมกิจกรรมสำคัญตามกรอบ OTPF-4 [34] ซึ่งแบ่งเป็น

- ADLs (กิจวัตรประจำวันพื้นฐาน) เช่น แอปเปิ้ล ไข่ และสับปะรด ซึ่งเกี่ยวข้องกับ การรับประทานและการจัดการอาหาร
- IADLs (กิจวัตรประจำวันต่อเนื่อง) เช่น เตารีด ไม้กวาด และจักรยาน ซึ่งเป็น อุปกรณ์ในการดูแลที่พักอาศัยและการเดินทาง
- Environmental Factors (ปัจจัยด้านสิ่งแวดล้อม) เช่น ผ้า màn และต้นไม้ ซึ่งทำ หน้าที่เป็นปัจจัยแวดล้อมทางกายภาพที่ส่งผลต่อประสิทธิภาพการทำงาน

2. ด้านการรับรู้ผ่านการสัมผัส (Haptic Perception)

วัตถุแต่ละชิ้นมีคุณสมบัติทางกายภาพที่ใช้ทดสอบระบบประสาทสัมผัสตามแนวคิด ของ Lederman และ Klatzky [35] ได้แก่

- พื้นผิวและรูปร่าง (Texture & Shape) การใช้สับปะรดเพื่อทดสอบการรับรู้ ความขรุขระ และไข่เพื่อทดสอบการรับรู้รูปร่างที่เปราะบาง
- อุณหภูมิ (Thermal Property) การใช้เตารีดเพื่อประเมินความสามารถในการ จำแนกความร้อน ซึ่งเป็นทักษะด้านความปลอดภัย
- แรงสั่นสะเทือน (Vibration) การรับรู้การทำงานของพัดลมหรือการเคลื่อนไหว ของแมว

3. ด้านทิศทางและการเคลื่อนไหว (Orientation & Mobility)

วัตถุเหล่านี้ทำหน้าที่เป็นองค์ประกอบสำคัญในการนำทางตามหลัก O&M [36]

ดังนี้

- Landmarks (จุดอ้างอิง) วัตถุที่ติดตั้งคงที่ เช่น ต้นไม้ หรือผ้า màn ช่วยระบุ ตำแหน่งพื้นที่
- Clues (เบาะแส) เสียงและลมจากพัดลม หรือตำแหน่งของจักรยาน ช่วยให้ผู้ใช้ปฏิบัติ กิจกรรมรับรู้ทิศทางและหลบหลีกสิ่งกีดขวางได้อย่างปลอดภัย

นอกจากนี้ มีการกำหนดจำนวนภาพทดสอบโดยเก็บข้อมูลเป็นการถ่ายภาพรวม 10 ครั้งต่อวัตถุ ส่งผลให้มีภาพทดสอบ 100 ภาพต่อกรณี รวมทั้งหมด 600 ภาพ ตามรายละเอียดดังนี้

- Simple Case 10 ภาพ/วัตถุ รวมทั้งหมด 100 ภาพ
- Complex Case (2.1)–(2.5) กรณีละ 10 ภาพ/วัตถุ รวมทั้งหมด 500 ภาพ

3.3.11 สถิติที่ใช้ในการวิเคราะห์ข้อมูลและประเมินผล

เพื่อความเรียบง่ายแต่ครอบคลุมทั้งด้านความถูกต้องของโมเดลและความสามารถในการใช้งาน รวมถึงประสบการณ์ผู้ใช้ จึงได้กำหนดการประเมินผลแบ่งเป็น 2 มิติหลัก ได้แก่ (1) การประเมินเชิงเทคนิคเรื่องความแม่นยำของการจำแนกภาพบนอุปกรณ์ โดยมีเมตริกที่ใช้ได้แก่ Top-1 Accuracy และ Top-k Accuracy (2) การประเมินความสามารถในการใช้งานและความพึงพอใจของผู้ใช้ โดยใช้แบบสอบถาม SUS แบบประเมินความพึงพอใจของผู้ใช้ เป็นต้น

3.3.12 การประเมินความแม่นยำ (Accuracy Evaluation)

(1) Top-1 Accuracy ให้ N คือจำนวนภาพทดสอบ และ Correct คือจำนวนที่ทำนายถูกต้องตามสูตรในรูปด้านล่าง และรายงานเป็นร้อยละ พร้อมจำนวนตัวอย่างที่ใช้ คำนวณได้จากสมการที่ 2.1

(2) Top-k Accuracy สัดส่วนของข้อมูลทดสอบที่คำตอบที่ถูกต้องปรากฏอยู่ภายในผลการทำนาย k อันดับแรกของโมเดล เพื่อประเมินความแม่นยำของโมเดลในการใช้งานจริง เนื่องจากแอปพลิเคชันนำเสนอผลลัพธ์หลายอันดับ คำนวณได้จากสมการที่ 2.2

3.3.13 การประเมินความสามารถและความพึงพอใจของผู้ใช้ (Usability and Satisfaction Evaluation)

การประเมินประสบการณ์ผู้ใช้ (User Experience - UX) โดยเก็บรวบรวมและทำการวิเคราะห์ข้อมูลจากการทำแบบสอบถาม และ/หรือการสัมภาษณ์ เพื่อประเมินความสามารถในการใช้งานและความพึงพอใจ รวมถึงประสบการณ์ของผู้ใช้ที่มีต่อแอปพลิเคชันต้นแบบรู้จำวัตถุด้วยสมาร์ทโฟน “VocabEyes” เพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) และรวบรวมความคิดเห็น หรือข้อเสนอแนะของผู้ใช้งาน เพื่อนำมาพัฒนาปรับปรุงให้มีประสิทธิภาพมากยิ่งขึ้น โดยแบบสอบถามหนึ่งฉบับจะประกอบด้วยส่วนย่อย 3 ส่วน คือ (ก) แบบประเมินความสามารถในการใช้งาน (ข) แบบประเมินความพึงพอใจของผู้ใช้ (ค) ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม ตามรายละเอียดด้านล่าง

(ก) การประเมินความสามารถในการใช้งาน

ใช้แบบสอบถามที่จัดทำขึ้นมาโดยยึดเกณฑ์จาก System Usability Scale (SUS) ดังตารางที่ 3.3 โดยมีทั้งภาษาไทยและภาษาญี่ปุ่นซึ่งปรับข้อความบางส่วนให้เข้ากับการประเมินความสามารถในการใช้งานแอปพลิเคชัน VocabEyes โดยมีการเกณฑ์การคำนวณคะแนนตามหัวข้อ 2.10.2 ด้านความสามารถในการใช้งานแอปพลิเคชัน เรื่อง System Usability Scale (SUS) หัวข้อย่อย วิธีการคำนวณคะแนน SUS โดยสังเขปดังนี้

(1) เกณฑ์การให้คะแนน 1 = ไม่เห็นด้วยอย่างยิ่ง 2 = ไม่เห็นด้วย 3 = ปานกลาง 4 = เห็นด้วย 5 = เห็นด้วยอย่างยิ่ง

(2) วิธีคำนวณคะแนน SUS ข้อเลขคี่ (1,3,5,7,9): คะแนนย่อย = ค่าที่ตอบ - 1 และข้อเลขคู่ (2,4,6,8,10): คะแนนย่อย = 5 - ค่าที่ตอบ รวมคะแนนย่อยทั้ง 10 ข้อ แล้วคูณด้วย 2.5 ได้คะแนน SUS อยู่ระหว่าง 0-100 คะแนน (~68 = ค่ากลางอุตสาหกรรม, >80 ดีมากหรือยอดเยี่ยม)

(3) การแปลผลคะแนน SUS อ้างอิงจากเกณฑ์การแปลความหมายคะแนนตามหัวข้อ 2.10.2 ด้านความสามารถในการใช้งานแอปพลิเคชัน เรื่อง System Usability Scale (SUS) หัวข้อย่อย การแปลผลคะแนน SUS

ตารางที่ 3.3 แบบประเมินความสามารถในการใช้งาน

ข้อ No.	คำถาม 質問	เกณฑ์การให้คะแนน 判断基準				
		1 = ไม่เห็นด้วยอย่างยิ่ง まったくそう思わない 5 = เห็นด้วยอย่างยิ่ง 非常にそう思う				
		1	2	3	4	5
1	ฉันคิดว่าอยากใช้งานแอปนี้บ่อยๆ このアプリを頻繁に使いたいと思います。					
2	ฉันรู้สึกว่าแอปนี้ซับซ้อนเกินความจำเป็น このアプリは必要以上に複雑だと感じます。					
3	ฉันคิดว่าแอปนี้ใช้งานง่าย このアプリは使いやすいと思います。					
4	ฉันคิดว่าจำเป็นต้องมีผู้เชี่ยวชาญคอยช่วยเหลือจึงจะ ใช้แอปนี้ได้ このアプリを使うには、専門家のサポートが 必要だと思います。					
5	ฉันรู้สึกว่าฟังก์ชันต่าง ๆ ในแอปนี้ทำงานเชื่อมโยงกันดี เป็นอย่างดี このアプリの各機能はうまく統合されている と感じました。					
6	ฉันรู้สึกว่าแอปนี้มีความไม่สอดคล้องกันมากเกินไป このアプリには一貫性が欠けている部分が多 いと感じます。					
7	ฉันคิดว่าคนส่วนใหญ่จะเรียนรู้การใช้งานแอปนี้ได้โดย รวดเร็ว 多くの人はこのアプリの使い方をすぐに習得 できると思います。					

ตารางที่ 3.3 แบบประเมินความสามารถในการใช้งาน (ต่อ)

ข้อ No.	คำถาม 質問	เกณฑ์การให้คะแนน 判断基準				
		1 = ไม่เห็นด้วยอย่างยิ่ง まったくそう思わない 5 = เห็นด้วยอย่างยิ่ง 非常に思う				
		1	2	3	4	5
8	ฉันรู้สึกว่าแอปนี้ใช้งานยุ่งยาก และซับซ้อนมาก このアプリは使いにくく、とても複雑だと感じます。					
9	ฉันรู้สึกมั่นใจในการสั่งการทำงานขณะที่ใช้งานแอปนี้ このアプリを使うとき、自信を持って操作できます。					
10	ฉันจำเป็นต้องเรียนรู้หลายอย่างก่อนเริ่มใช้งานแอปนี้ アプリを使い始める前に、多くのことを学ぶ必要があると感じます。					

(ข) แบบประเมินความพึงพอใจของผู้ใช้

เป็นคำถามเสริมซึ่งไม่คำนวณรวมในคะแนน SUS ดังตารางที่ 3.4 มีวัตถุประสงค์เพื่อประเมินความพึงพอใจของผู้ใช้ เก็บข้อมูลและข้อเสนอแนะเพื่อการปรับปรุง

ตารางที่ 3.4 แบบประเมินความพึงพอใจของผู้ใช้

ข้อ No.	คำถาม 質問	เกณฑ์การให้คะแนน 判断基準				
		1 = ไม่เห็นด้วยอย่างยิ่ง まったくそう思わない 5 = เห็นด้วยอย่างยิ่ง 非常に思う				
		1	2	3	4	5
1	เสียง TTS ภาษาไทยฟังชัดเจนเพียงพอสำหรับฉัน 自分にとってタイ語のTTS音声は十分に聞き取りやすいです。					
2	เสียง TTS ภาษาญี่ปุ่นฟังชัดเจนเพียงพอสำหรับฉัน 自分にとって日本語のTTS音声は十分に聞き取りやすいです。					
3	เสียง TTS ภาษาอังกฤษฟังชัดเจนเพียงพอสำหรับฉัน 自分にとって英語のTTS音声は十分に聞き取りやすいです。					

ตารางที่ 3.4 แบบประเมินความพึงพอใจของผู้ใช้ (ต่อ)

ข้อ No.	คำถาม 質問	เกณฑ์การให้คะแนน 判断基準				
		1	2	3	4	5
4	ลำดับการอ่าน (ไทย→ญี่ปุ่น→อังกฤษ) เหมาะสมและช่วยให้ฉันเรียนรู้คำศัพท์ 読み上げの順序（タイ語→日本語→英語）は適切で、語彙学習に役立ちます。					
5	ปุ่มถ่ายภาพ/ถ่ายใหม่ปุ่มเดียว ใช้งานสะดวกสำหรับฉัน 撮影／再撮影を1つのボタンで行えるのは、操作しやすいです。					
6	ขนาดตัวอักษร/ปุ่มใหญ่พอและใช้งานมือเดียวได้สะดวก 文字やボタンのサイズは十分大きく、片手でも操作しやすいです。					
7	ระบบทำงานออฟไลน์ได้รวดเร็วเพียงพอสำหรับฉัน オフラインでも動作が速く、自分にとって十分だと感じます。					
8	ข้อเสนอแนะอื่นๆ (โปรดระบุ) その他のご意見・ご要望（ご自由にお書きください）					

(ค) ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

ใช้ในการเก็บข้อมูลทั่วไปของผู้ตอบแบบสอบถามดังตารางด้านล่าง

ตารางที่ 3.5 แบบเก็บข้อมูลผู้ตอบแบบสอบถาม

ข้อ No.	ข้อมูลผู้ตอบแบบสอบถาม 回答者情報
1	เพศ 性別 ☐ ชาย ☐ หญิง ☐ ไม่ระบุ ☐ 男性 ☐ 女性 ☐ 無回答
2	อายุ 年齢 ☐ ต่ำกว่า 18 ปี ☐ 18-25 ปี ☐ 25-35 ปี ☐ 36-45 ปี ☐ 46-55 ปี ☐ 55 ปีขึ้นไป ☐ 18 歳未満 ☐ 18-25歳 ☐ 25-35歳 ☐ 36-45 歳 ☐ 46-55歳 ☐ 55 歳以上

ตารางที่ 3.5 แบบเก็บข้อมูลผู้ตอบแบบสอบถาม (ต่อ)

ข้อ No.	ข้อมูลผู้ตอบแบบสอบถาม 回答者情報	
3	สถานะการมองเห็น 視覚の状態	☐ ตาบอด ☐ มองเห็นเลือนราง ☐ อาสาสมัครจำลองปิดกั้นการมองเห็น ☐ อื่นๆ ☐ 全盲 ☐ 弱視 (ロービジョン) ☐ 視覚遮断体験ボランティア ☐ その他
4	ประสบการณ์ใช้สมาร์ทโฟน スマートフォンの使用経験	☐ ต่ำ ☐ ปานกลาง ☐ สูง ☐ 低 ☐ 中 ☐ 高
5	สมาร์ทโฟนที่ใช้งานอยู่ 使用しているスマートフォン	ยี่ห้อ メーカー: รุ่น 機種: ระบบปฏิบัติการ OS:
6	การใช้งานระบบเสียง 音声の使用法	☐ ลำโพง ☐ หูฟัง ☐ สปีคเกอร์ ☐ イヤホン
7	ใช้ฟีเจอร์ผู้อ่านหน้าจอ (TalkBack) หรือไม่ 画面読み上げ機能 (TalkBack) の使用	☐ ใช่ ☐ ไม่ใช่ ☐ 使用する ☐ 使用しない

ระยะที่ 3 การทดสอบเชิงเทคนิค

ทำการทดสอบเชิงเทคนิคตามเงื่อนไขที่กำหนดไว้ในข้อ 3.3.9 เงื่อนไขการประเมินความแม่นยำของแอปพลิเคชันต้นแบบ ข้อ 3.3.10 วัตถุประสงค์และจำนวนภาพทดสอบ และวัดความถูกต้องด้วย Top-1 Accuracy และ Top-k Accuracy ตามเงื่อนไขที่กำหนดไว้ในข้อ 3.3.11 สถิติที่ใช้ในการวิเคราะห์ข้อมูลและประเมินผล หัวข้อย่อย 3.3.12 การประเมินความแม่นยำ (Accuracy Evaluation)

ระยะที่ 4 การประเมินการใช้งานและความพึงพอใจ

ทำการประเมินการใช้งานและความพึงพอใจตามเงื่อนไขที่กำหนดไว้ในข้อ 3.3.11 สถิติที่ใช้ในการวิเคราะห์ข้อมูลและประเมินผล หัวข้อย่อย (ก) การประเมินความสามารถในการใช้งาน (ข) แบบประเมินความพึงพอใจของผู้ใช้ (ค) ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม โดยอาสาสมัครกลุ่มตัวอย่างจำนวนตั้งแต่ 5 คนขึ้นไป

ระยะที่ 5 รวบรวมข้อมูล วิเคราะห์ และสรุปผล

ทำการรวบรวมข้อมูลในด้านต่างๆ เช่น การประเมินเชิงเทคนิค การประเมินการใช้งานและความพึงพอใจ จัดเก็บลงในไฟล์อิเล็กทรอนิกส์ และนำมา วิเคราะห์เปรียบเทียบ พร้อมทั้งสรุปผล รวมถึงเสนอแนวทางแก้ไข และจัดทำเอกสารรายงาน

บทที่ 4

ผลการวิจัย

4.1 บทนำ

บทนี้นำเสนอผลการดำเนินงานและผลการวิเคราะห์ข้อมูลงานวิจัยในการพัฒนาแอปพลิเคชัน VocabEyes ซึ่งเป็นแอปพลิเคชันต้นแบบสำหรับการรู้จำวัตถุและอ่านออกเสียงคำศัพท์ 3 ภาษา (ไทย-ญี่ปุ่น-อังกฤษ) ประกอบด้วยการรายงานผลการพัฒนาและออกแบบแอปพลิเคชัน และการประเมินผล โดยออกเป็น 2 ส่วนหลัก ได้แก่

1. การประเมินความแม่นยำ (Accuracy Evaluation)
2. การประเมินความสามารถในการใช้งานและความพึงพอใจของผู้ใช้ (Usability and Satisfaction Evaluation)

พร้อมวิเคราะห์ความสอดคล้องกับ Functional Requirements และ Non-Functional Requirements ที่กำหนดไว้

4.2 ผลการพัฒนาและออกแบบแอปพลิเคชัน

จากการดำเนินการพัฒนาแอปพลิเคชันต้นแบบ VocabEyes ตามขั้นตอนและข้อกำหนดที่ระบุไว้ในบทที่ 3 ผู้วิจัยได้ผลการดำเนินงานดังนี้

4.2.1 การทำงานของระบบ แอปพลิเคชัน VocabEyes

ถูกพัฒนาขึ้นโดยเน้นการประมวลผลบนอุปกรณ์ (On-device Processing) เพื่อรองรับการทำงานแบบไม่พึ่งพาอินเทอร์เน็ต โดยมีกระบวนการทำงานหลักแบ่งออกเป็น 4 ขั้นตอน ดังนี้

1. การรับภาพ (Image Input) แอปพลิเคชันสามารถรับภาพนิ่งจากกล้องถ่ายภาพของสมาร์ทโฟนโดยตรง หรือภาพถ่ายจากแกลเลอรีเพื่อใช้ประมวลผลได้
2. การรู้จำวัตถุ (Object Recognition) โมเดลสามารถทำการวิเคราะห์วัตถุจากภาพถ่ายจากโมเดลสำเร็จรูป MobileNet V3 Pretrained ImageNet 21K-P (11,221 Classes) ผ่านการใช้งานแอปพลิเคชัน VocabEyes ได้
3. การประมวลผลคำศัพท์ ระบบจะดึงคำศัพท์ที่เกี่ยวข้องมาแสดงผลพร้อมกัน 3 ภาษา อย่างน้อยหนึ่งภาษาจากภาษาไทย ญี่ปุ่น และอังกฤษโดยขึ้นอยู่กับคลังคำศัพท์ที่มี
4. การอ่านออกเสียง (Speech Synthesis) ระบบใช้ฟังก์ชัน Text-to-Speech (TTS) จาก Google TTS ที่ได้ทำการดาวน์โหลดข้อมูลลงในสมาร์ทโฟนเครื่องทดสอบเอาไว้แล้ว สำหรับใช้ในการอ่านออกเสียงคำศัพท์ภาษาไทย ญี่ปุ่น และอังกฤษ

4.2.2 ฐานข้อมูลพจนานุกรม 3 ภาษา

ผลลัพธ์การจัดทำไฟล์ .csv เพื่อใช้เป็นฐานข้อมูลพจนานุกรม 3 (TH-JP-EN) ดังรูปที่

4.1

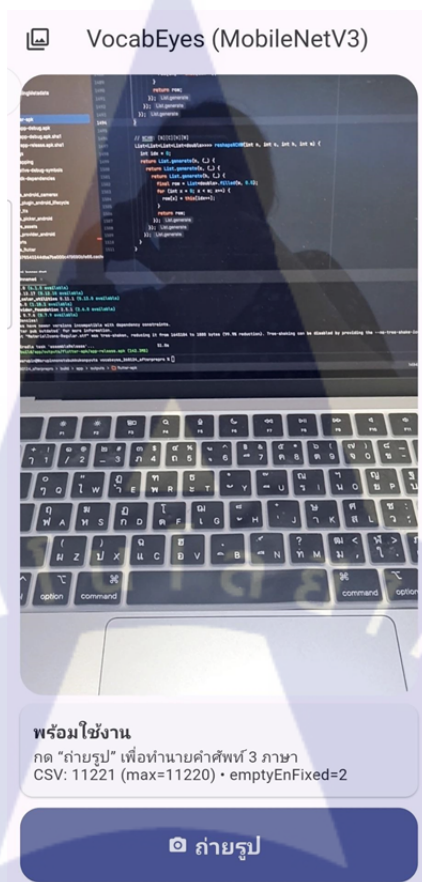
	A	B	C	S	T	U	V	W	X	Y
1	class_index	wind	class_description	wordnet_id_en_omw	wordnet_id_th_omw	wordnet_id_ja_omw	th_words	ja_words	th_primary	ja_primary
2	0	n00005787	benthos	omw-a-00005787-ก	omw-th-00005787-ก	omw-j-00005787-ก	สิ่งมีชีวิตใต้น้ำ	底生動物	สิ่งมีชีวิตใต้น้ำ	底生動物
3	1	n00004844	cell	omw-a-00004844-ก	omw-th-00004844-ก	omw-j-00004844-ก	เซลล์	セル 細胞	เซลล์	セル
4	2	n00007946	person	omw-a-00007946-ก	omw-th-00007946-ก	omw-j-00007946-ก	คน บุคคล ตัวเอก 人 員 者 人 人間 個人	人	คน	人
5	3	n00015388	animal	omw-a-00015388-ก	omw-th-00015388-ก	omw-j-00015388-ก	สัตว์ สัตว์ป่า สัตว์เลี้ยง สัตว์ 4つ脚 4つ足 アニマル 動物	สัตว์	4つ脚	4つ脚
6	4	n00017222	plant	omw-a-00017222-ก	omw-th-00017222-ก	omw-j-00017222-ก	พืชมงคล พืช	植物 草木	พืชมงคล	植物
7	5	n00021265	food	omw-a-00021265-ก	omw-th-00021265-ก	omw-j-00021265-ก	สารอาหาร อาหาร	栄養 栄養分 栄養素	สารอาหาร	栄養
8	6	n00021939	artifact	omw-a-00021939-ก	omw-th-00021939-ก	omw-j-00021939-ก	สิ่งของ สิ่งทอ สิ่งประดิษฐ์	アーチファクト アーティファクト	สิ่งของ	アーチファクト
9	7	n00120010	hop	omw-a-00120010-ก	omw-th-00120010-ก	omw-j-00120010-ก	การกระโดด การกระโดด	ホップ 跳 跳躍	การกระโดด	ホップ
10	8	n00141669	check-in	omw-a-00141669-ก	omw-th-00141669-ก	omw-j-00141669-ก	การเช็คอิน	チェックイン	การเช็คอิน	チェックイン
11	9	n00289000	dressage	omw-a-00289000-ก	omw-th-00289000-ก	omw-j-00289000-ก	การแสดงอันงดงาม	馬術	การแสดงอันงดงาม	馬術
12	10	n00289394	pie	omw-a-00289394-ก	omw-th-00289394-ก	omw-j-00289394-ก	พาย	パイ	พาย	パイ
13	11	n00324978	funambulism	omw-a-00324978-ก	omw-th-00324978-ก	omw-j-00324978-ก	การเดินโซด	綱渡り	การเดินโซด	綱渡り
14	12	n00326094	rock climbing	omw-a-00326094-ก	omw-th-00326094-ก	omw-j-00326094-ก	การปีนเขา	岩登り	การปีนเขา	岩登り
15	13	n00439458	contact sport	omw-a-00439458-ก	omw-th-00439458-ก	omw-j-00439458-ก	กีฬาที่สัมผัสกัน	コンタクトスポーツ	กีฬาที่สัมผัสกัน	コンタクトスポーツ
16	14	n00439661	outdoor sport	omw-a-00439661-ก	omw-th-00439661-ก	omw-j-00439661-ก	กีฬากลางแจ้ง	アウトドアスポーツ	กีฬากลางแจ้ง	アウトドアスポーツ
17	15	n00439802	gymnastics	omw-a-00439802-ก	omw-th-00439802-ก	omw-j-00439802-ก	กีฬายานาติก	гимнастика ジムナスティクス 体操	กีฬายานาติก	гимнастика ジムナスティクス 体操
18	16	n00439875	acrobatics	omw-a-00439875-ก	omw-th-00439875-ก	omw-j-00439875-ก	กีฬายานาติก	アクロバット タンブリング	กีฬายานาติก	アクロバット
19	17	n00439826	track and field	omw-a-00439826-ก	omw-th-00439826-ก	omw-j-00439826-ก	กีฬาระเบียงและลาน	トラック	กีฬาระเบียงและลาน	トラック
20	18	n00440039	track	omw-a-00440039-ก	omw-th-00440039-ก	omw-j-00440039-ก	การกระโดดไกล	跳び	การกระโดดไกล	跳び
21	19	n00440382	broad jump	omw-a-00440382-ก	omw-th-00440382-ก	omw-j-00440382-ก	การกระโดดสูง	高跳び	การกระโดดสูง	高跳び
22	20	n00440509	high jump	omw-a-00440509-ก	omw-th-00440509-ก	omw-j-00440509-ก	กีฬาสกี	スキー	กีฬาสกี	スキー
23	21	n00440747	skiing	omw-a-00440747-ก	omw-th-00440747-ก	omw-j-00440747-ก	การแข่งกีฬากลางไกล	クロコン	การแข่งกีฬากลางไกล	クロコン
24	22	n00440941	cross-country skiing	omw-a-00440941-ก	omw-th-00440941-ก	omw-j-00440941-ก	สกีกระโดด	スキージャンプ	สกีกระโดด	スキージャンプ
25	23	n00441073	ski jumping	omw-a-00441073-ก	omw-th-00441073-ก	omw-j-00441073-ก	กีฬายานา	水泳	กีฬายานา	水泳
26	24	n00441824	water sport	omw-a-00441824-ก	omw-th-00441824-ก	omw-j-00441824-ก	การว่ายน้ำ	水泳	การว่ายน้ำ	水泳
27	25	n00442115	swimming	omw-a-00442115-ก	omw-th-00442115-ก	omw-j-00442115-ก	การว่ายน้ำ	水泳	การว่ายน้ำ	水泳
28	26	n00442437	bathe	omw-a-00442437-ก	omw-th-00442437-ก	omw-j-00442437-ก	การว่ายน้ำ	水泳	การว่ายน้ำ	水泳
29	27	n00442647	do	omw-a-00442647-ก	omw-th-00442647-ก	omw-j-00442647-ก	การแข่งม้า	競馬	การแข่งม้า	競馬
30	28	n00442981	horse	omw-a-00442981-ก	omw-th-00442981-ก	omw-j-00442981-ก	การแข่งม้า	競馬	การแข่งม้า	競馬
31	29	n00443231	floating	omw-a-00443231-ก	omw-th-00443231-ก	omw-j-00443231-ก	การแข่งม้า	競馬	การแข่งม้า	競馬
32	30	n00443692	cliff diving	omw-a-00443692-ก	omw-th-00443692-ก	omw-j-00443692-ก	การแข่งม้า	競馬	การแข่งม้า	競馬
33	31	n00443803	flip	omw-a-00443803-ก	omw-th-00443803-ก	omw-j-00443803-ก	การแข่งม้า	競馬	การแข่งม้า	競馬
34	32	n00444651	skin diving	omw-a-00444651-ก	omw-th-00444651-ก	omw-j-00444651-ก	การดำน้ำ	スキndaイビング	การดำน้ำ	スキndaイビング
35	33	n00444846	scuba diving	omw-a-00444846-ก	omw-th-00444846-ก	omw-j-00444846-ก	การดำน้ำแบบสกน	スキndaイビング	การดำน้ำแบบสกน	スキndaイビング
36	34	n00444937	surfing	omw-a-00444937-ก	omw-th-00444937-ก	omw-j-00444937-ก	การดำน้ำแบบสกน	スキndaイビング	การดำน้ำแบบสกน	スキndaイビング
37	35	n00445055	surfing	omw-a-00445055-ก	omw-th-00445055-ก	omw-j-00445055-ก	การดำน้ำแบบสกน	スキndaイビング	การดำน้ำแบบสกน	スキndaイビング
38	36	n00445226	water-skiing	omw-a-00445226-ก	omw-th-00445226-ก	omw-j-00445226-ก	สกีน้ำ	水スキー	สกีน้ำ	水スキー
39	37	n00445351	rowing	omw-a-00445351-ก	omw-th-00445351-ก	omw-j-00445351-ก	กีฬายานา	水泳	กีฬายานา	水泳
40	38	n00445685	sculling	omw-a-00445685-ก	omw-th-00445685-ก	omw-j-00445685-ก	การแข่งม้า	競馬	การแข่งม้า	競馬
41	39	n00445802	boxing	omw-a-00445802-ก	omw-th-00445802-ก	omw-j-00445802-ก	กีฬายานา	水泳	กีฬายานา	水泳

รูปที่ 4.1 รูปตัวอย่างไฟล์ฐานข้อมูลพจนานุกรม 3 ภาษา (TH-JP-EN)

4.2.3 การออกแบบส่วนติดต่อผู้ใช้

เน้นความเรียบง่ายและไม่ซับซ้อน เพื่อให้ผู้ใช้สามารถเข้าใจการทำงานได้ทันที จึงกำหนดให้แสดงผลหลักเพียงสองหน้าจอ โดยมีองค์ประกอบหลักดังนี้

- 1) หน้าจอหลัก (Main Screen) ออกแบบให้พื้นที่ส่วนใหญ่เป็นช่องมองภาพจากกล้อง และสามารถสั่งการทำงานโดยการกดที่บริเวณทั่วทั้งหน้าจอได้ ดังรูปที่ 4.2



รูปที่ 4.2 หน้าจอหลักการถ่ายภาพ

2) หน้าจอแสดงผลคำศัพท์และอ่านออกเสียง แสดงผลคำศัพท์ 3 ภาษา ไทย ญี่ปุ่น อังกฤษ ในคราวเดียว หรืออย่างน้อยหนึ่งภาษาโดยขึ้นอยู่กับคลังคำศัพท์ที่มี เพื่อให้ผู้ใช้สามารถเปรียบเทียบคำศัพท์ระหว่างภาษาได้สะดวกกรณีผู้ใช้ที่มองเห็นเรือนรางหรือผู้ใช้สายตาปกติ พร้อมการอ่านออกเสียง ดังรูปที่ 4.3



รูปที่ 4.3 ตัวอย่างหน้าจอแสดงผลการทำนายและอ่านออกเสียง 3 ครั้ง

3) เสียงคำแนะนำการใช้งานแอปพลิเคชัน (Audio Tutorial Guide) ทุกครั้งที่เปิดแอปพลิเคชันในครั้งแรก จะมีเสียงแนะนำการใช้งานโดยพูดว่า “พร้อมใช้งาน กรุณากดหนึ่งครั้งที่หน้าจอเพื่อถ่ายภาพ กดค้างเพื่อฟังซ้ำ และกดอีกครั้งเพื่อกลับสู่หน้าจอถ่ายภาพ”

4) ทำนายรูปจากการถ่ายภาพหรือรูปถ่ายจากแกลเลอรี สามารถถ่ายภาพจากสมาร์ตโฟนโดยตรง รวมถึงเลือกภาพถ่ายที่ถ่ายภาพเก็บไว้ หรือบันทึกจากอินเทอร์เน็ต และแหล่งอื่นๆ เพื่อให้โมเดลในแอปพลิเคชัน VocabEyes ทำนายผลได้ โดยตัวอย่างของภาพทดสอบที่บันทึกไว้ซึ่งเลือกมาจากแกลเลอรีดังรูปที่ 4.4 และหน้าจอการเลือกรูปภาพจากแกลเลอรี (รูปภาพซ้าย) และผลการทำนายภาพที่เลือก (รูปภาพขวา) ตามลำดับดังรูปที่ 4.5



รูปที่ 4.4 รูปภาพตัวอย่างจากแกลเลอรี



รูปที่ 4.5 หน้าจอการเลือกรูปภาพและทำนายรูปภาพจากแกลเลอรี

4.2.4 ตัวอย่างการทดสอบถ่ายภาพจริงในแต่ละกรณี ภาพตัวอย่างในกรณีของการถ่ายภาพจริงโดยใช้แอปพลิเคชัน VocabEyes ในการถ่ายภาพวัตถุ “แอปเปิ้ล (Apple)”

สถานการณ์ที่ใช้ในการประเมิน (Evaluation Scenarios)

1) Simple Case สถานการณ์ถ่ายภาพแบบเรียบง่าย (ไม่มีสิ่งกีดขวาง หรือฉากหลังเรียบ) หรือเรียกว่า S1 หมายถึง Simple Case ไม่มีสิ่งกีดขวาง/ฉากหลังเรียบ ดังรูปที่ 4.6



รูปที่ 4.6 หน้าจอถ่ายภาพทดสอบจริงกรณี S1

2) Complex Case สถานการณ์ภาพถ่ายที่ซับซ้อน

Complex case แบ่งเป็น 5 กรณีย่อยเพื่อจำลองปัจจัยที่ทำให้ระบบรู้จำยากขึ้นในการใช้งานจริง ดังนี้

2.1) วัตถุเดียวมีพื้นหลังรบกวน (Single Object with Background) หรือเรียกว่า C1 หมายถึง Complex Case วัตถุเดียวมีพื้นหลังรบกวน ดังรูปที่ 4.7



รูปที่ 4.7 หน้าจอถ่ายภาพทดสอบจริงกรณี C1

2.2) เห็นวัตถุบางส่วน (Partial Occlusion) หรือเรียกว่า C2 หมายถึง Complex Case เห็นวัตถุบางส่วน/ถูกบัง ดังรูปที่ 4.8



รูปที่ 4.8 หน้าจอถ่ายภาพทดสอบจริงกรณี C2

2.3) วัตถุหลายอย่างในภาพ (Multiple objects) หรือเรียกว่า C3 หมายถึง Complex Case วัตถุหลายอย่างในภาพ ดังรูปที่ 4.9



รูปที่ 4.9 หน้าจอถ่ายภาพทดสอบจริงกรณี C3

2.4) วัตถุเล็ก พื้นหลังใหญ่ (Small Object in Large Background) หรือเรียกว่า C4 หมายถึง Complex Case วัตถุเล็ก พื้นหลังใหญ่ ดังรูปที่ 4.10



รูปที่ 4.10 หน้าจอถ่ายภาพทดสอบจริงกรณี C4

2.5) ไฟสลัวหรือแสงน้อย (Low-Light) หรือเรียกว่า C5 หมายถึง Complex Case ไฟสลัวหรือแสงน้อย



รูปที่ 4.11 หน้าจอถ่ายภาพทดสอบจริงกรณี C5

4.3 ผลการประเมินด้านประสิทธิภาพของแอปพลิเคชัน

4.3.1 วัตถุประสงค์

การประเมินความแม่นยำของระบบมีวัตถุประสงค์เพื่อพิจารณาว่าโมเดลรู้จำวัตถุที่นำมาใช้ในแอปพลิเคชันต้นแบบสามารถทำนายวัตถุจากภาพถ่ายได้ถูกต้องเพียงใดหรืออยู่ในระดับที่เหมาะสมต่อการนำไปใช้งานจริงหรือไม่ โดยการประเมินใช้ตัวชี้วัดมาตรฐาน ได้แก่ Top-1 Accuracy และ Top-k Accuracy ($k = 5$) ตามที่กำหนดไว้ในบทที่ 3

4.3.2 ผลการประเมินความแม่นยำ

จากการทดสอบถ่ายภาพ 10 วัตถุ จำนวน 10 ภาพต่อวัตถุ ในแต่ละกรณี รวมทั้งหมด 600 ภาพ ภายใต้เงื่อนไขที่กำหนดไว้ในบทที่ 3 ได้ผลลัพธ์โดยสรุปดังนี้

ตารางที่ 4.1 ผลการประเมินความแม่นยำของระบบ

Scene	จำนวนภาพ	Top-1	Top-5
S1	100	100	100
C1	100	97	100
C2	100	90	100
C3	100	89	90
C4	100	90	100
C5	100	97	100
Overall	600	93.83	98.33

หมายเหตุ : อ้างอิง 3.9.2 ตัวอย่างและคำจำกัดความของแต่ละกรณีโดยสังเขปดังนี้

S1 หมายถึง Simple Case ไม่มีสิ่งกีดขวาง/ฉากหลังเรียบ

C1 หมายถึง Complex Case วัตถุเดียวมีพื้นหลังรบกวน

C2 หมายถึง Complex Case เห็นวัตถุบางส่วน/ถูกบัง

C3 หมายถึง Complex Case วัตถุหลายอย่างในภาพ

C4 หมายถึง Complex Case วัตถุเล็ก พื้นหลังใหญ่

C5 หมายถึง Complex Case ไฟสลัว/แสงน้อย

4.3.3 การวิเคราะห์ผลความแม่นยำ

การวิเคราะห์ผลลัพธ์จากค่าเฉลี่ยของ Top-1 Accuracy รวม 93.83% และค่า Top-5 Accuracy รวม 98.33% แสดงให้เห็นว่าค่า Top-1 ที่สูงกว่า 90% สะท้อนให้เห็นถึงความสามารถของโมเดลในการเลือกคำตอบอันดับแรกได้ถูกต้องในระดับสูง สำหรับค่า Top-5 ที่สูงถึง 98.33% สะท้อนว่าแม้กรณีที่อันดับแรกผิด ระบบยังสามารถเสนอคำตอบที่ถูกต้องอยู่ใน 5 อันดับแรกเกือบทั้งหมด

จากความแตกต่างระหว่าง Top-1 และ Top-5 มีเพียงประมาณ 4.5% สะท้อนว่าความคลาดเคลื่อนส่วนใหญ่เกิดจากการจัดอันดับ ไม่ใช่การจำแนกผิดประเภทอย่างสิ้นเชิง

4.3.4 การเชื่อมโยงกับ Functional Requirement ด้านการรู้จำวัตถุ

ในบทที่ 3 ได้กำหนด Functional Requirement ว่า ระบบต้องสามารถรู้จำวัตถุจากภาพนิ่งและแสดงผลคำศัพท์ได้อย่างถูกต้อง ผลการประเมินที่ได้ (Top-1 = 93.83%) แสดงให้เห็นว่าระบบสามารถปฏิบัติตามข้อกำหนดดังกล่าวได้ในระดับที่น่าพอใจ และเหมาะสมต่อการใช้งานในบริบทการเรียนรู้คำศัพท์เบื้องต้น จึงถือว่า Requirement ด้าน Object Recognition บรรลุผลตามวัตถุประสงค์

4.4 ผลการประเมินความสามารถในการใช้งาน (System Usability Scale : SUS)

4.4.1 วัตถุประสงค์

ตามที่ระบุในบทที่ 3 งานวิจัยใช้แบบประเมิน System Usability Scale (SUS) เพื่อวัดความสามารถในการใช้งานของแอปพลิเคชัน โดยกลุ่มตัวอย่างอาสาสมัครจำนวน 5 คน ทดลองใช้งานและตอบแบบสอบถาม SUS จำนวน 10 ข้อ โดยมีคะแนนเต็ม 100 คะแนน

4.4.2 ผลคะแนน SUS รายบุคคล

จากการคำนวณตามสูตรมาตรฐานของ SUS ได้ผลดังนี้

ตารางที่ 4.2 ผลคะแนน SUS รายบุคคล

ผู้ตอบแบบสอบถาม	SUS score
นาย A ผู้เชี่ยวชาญฝ่ายตรวจสอบคุณภาพ ชาวญี่ปุ่น การศึกษาระดับปริญญาโท	95
นาย A ผู้เชี่ยวชาญฝ่ายตรวจสอบคุณภาพ ชาวญี่ปุ่น การศึกษาระดับปริญญาโท	77.5
นาย C ผู้เชี่ยวชาญฝ่ายระบบสารสนเทศ ชาวญี่ปุ่น การศึกษาระดับปริญญาตรี	100
นาย D ผู้เชี่ยวชาญฝ่ายควบคุมการผลิต ชาวญี่ปุ่น การศึกษาระดับปริญญาโท	100
นาง E รองผู้อำนวยการโรงเรียน ชาวไทย การศึกษาระดับปริญญาโท	72.5

หมายเหตุ : ค่าเฉลี่ย SUS 89 คะแนน และส่วนเบี่ยงเบนมาตรฐาน 13.06

4.4.3 การตีความคะแนน SUS

อ้างอิงจากเกณฑ์การแปลความหมายคะแนนตามหัวข้อ 2.10.2 ด้านความสามารถในการใช้งานแอปพลิเคชัน เรื่อง System Usability Scale (SUS) หัวข้อย่อย การแปลผลคะแนน จากการที่ผลคะแนนเฉลี่ยของแอปพลิเคชันอยู่ที่ 89.0 คะแนน ซึ่งมากกว่าเกณฑ์ 80.3 คะแนน จึงจัดอยู่ในระดับดีมาก (Excellent) แสดงให้เห็นว่า ผู้ใช้ประเมินว่าระบบมีความสามารถในการใช้งานสูง ใช้งานง่าย และไม่ซับซ้อน และส่วนเบี่ยงเบนมาตรฐาน 13.06 แสดงให้เห็นว่าความเห็นของผู้ประเมินมีความสอดคล้องกันในระดับปานกลางถึงค่อนข้างดี ไม่มีค่าที่เบี่ยงเบนผิดปกติอย่างรุนแรง ถือว่าอยู่ในเกณฑ์มาตรฐานที่ยอมรับได้ เนื่องจากการทดสอบกับกลุ่มคนจำนวนน้อย

ทั้งนี้ แม้ว่าแอปพลิเคชันจะได้รับคะแนน SUS เฉลี่ยโดยรวมในระดับดีมาก (89.00 คะแนน) แต่พบว่าค่าเบี่ยงเบนมาตรฐาน (S.D.) ค่อนข้างสูงที่ 13.06 ซึ่งเกิดจากความแตกต่างของคะแนนระหว่างกลุ่มผู้ทดสอบ โดยกลุ่มผู้ทดสอบชาวญี่ปุ่น 3 ท่านให้คะแนนในระดับสูงมาก (95-100 คะแนน) ในขณะที่ผู้ทดสอบชาวไทย 2 ท่านให้คะแนนที่ต่ำกว่า (72.5-77.5 คะแนน) สาเหตุของความแปรปรวนนี้ส่วนหนึ่งอาจเกิดจากปัจจัยภายนอกด้านความพร้อมของผู้ทดสอบในขณะนั้น เช่น ผู้ทดสอบบางรายมีภาระเหนียวล้าจากการทำงานหนักก่อนหน้านี้จะมาเข้าร่วมการทดสอบ ซึ่งได้ทราบจากบทสนทนาทั่วไปก่อนเริ่มทำการทดสอบ จึงอาจส่งผลต่อสมาธิและการตีความข้อคำถามในแบบประเมิน SUS โดยเฉพาะข้อคำถามที่เป็นประโยคเชิงปฏิเสธหรือมีความซับซ้อน ทำให้การตอบคำถามอาจคลาดเคลื่อนไปจากความรู้สึกริจริงขณะใช้งานแอปพลิเคชัน หรือข้อจำกัดด้านเวลา ทำให้ไม่สามารถทำความเข้าใจเนื้อหาที่แท้จริงได้ครบถ้วน นอกจากนี้ อาจเกิดจากความแตกต่างทางวัฒนธรรมและบริบททางภาษา ระหว่างผู้ทดสอบชาวไทยและญี่ปุ่น ซึ่งมีผลต่อพฤติกรรมการตีความส่วนติดต่อผู้ใช้ (UI) รวมถึงการตีความข้อคำถามในแบบประเมิน SUS โดยเฉพาะข้อคำถามที่เป็นประโยคเชิงปฏิเสธที่มีความซับซ้อน ทำให้เกิดความแตกต่างในการให้คะแนน, ความไม่คุ้นเคยของผู้ทดสอบกับการจำลองสถานการณ์ปิดตา ซึ่งอาจส่งผลกระทบต่อความเข้าใจและการให้คะแนนแบบประเมินได้ เป็นต้น

4.4.4 การเชื่อมโยงกับ Requirements

ในบทที่ 3 ได้กำหนดข้อกำหนดด้านการออกแบบส่วนติดต่อผู้ใช้ (User Interface) ว่าต้องใช้งานง่าย ขั้นตอนไม่ซับซ้อน ผู้ใช้สามารถเข้าใจการทำงานได้โดยไม่ต้องเรียนรู้มาก จากคะแนน SUS ที่อยู่ในระดับ Excellent สนับสนุนว่า ระบบสามารถบรรลุข้อกำหนดด้านความสามารถในการใช้งาน (Usability) ได้ตามที่วางแผนไว้

4.5 ผลการประเมินความพึงพอใจของผู้ใช้ (Satisfaction Evaluation)

รายละเอียดของผลการประเมินความพึงพอใจผู้ใช้ในแต่ละส่วนมีดังต่อไปนี้

4.5.1 วัตถุประสงค์

นอกเหนือจากการประเมินความสามารถในการใช้งาน งานวิจัยยังประเมินความพึงพอใจของผู้ใช้ต่อระบบในภาพรวม โดยใช้แบบสอบถามเสริมเฉพาะบริบทซึ่งประกอบด้วยคำถาม 7 ข้อ เป็นแบบ Likert Scale 5 ระดับ คะแนนเต็ม 5 คะแนน

4.5.2 ผลการประเมินค่าเฉลี่ยรายบุคคล

ตารางที่ 4.3 ผลการประเมินความพึงพอใจ

ผู้ตอบแบบสอบถาม	ค่าเฉลี่ย
นาย A ผู้เชี่ยวชาญฝ่ายตรวจสอบคุณภาพ ชาวญี่ปุ่น การศึกษาระดับปริญญาโท	4.71
นาย B ล่ามภาษาญี่ปุ่น-ไทย ชาวไทย การศึกษาระดับปริญญาตรี	4.86
นาย C ผู้เชี่ยวชาญฝ่ายระบบสารสนเทศ ชาวญี่ปุ่น การศึกษาระดับปริญญาตรี	5.00
นาย D ผู้เชี่ยวชาญฝ่ายควบคุมการผลิต ชาวญี่ปุ่น การศึกษาระดับปริญญาโท	4.71
นาง E รองผู้อำนวยการโรงเรียน ชาวไทย การศึกษาระดับปริญญาโท	4.71

จากตารางข้างต้นคำนวณออกมาเป็นค่าเฉลี่ยรวมอยู่ที่ 4.80 จากคะแนนเต็ม 5 คะแนน และส่วนเบี่ยงเบนมาตรฐาน 0.13

4.5.3 การวิเคราะห์ผล

ค่าเฉลี่ย 4.80 จากคะแนนเต็ม 5 คะแนน แสดงให้เห็นว่า ผู้ใช้มีความพึงพอใจต่อระบบในระดับสูงมาก อีกทั้ง ส่วนเบี่ยงเบนมาตรฐาน 0.13 มีค่าต่ำมาก แสดงว่าความเห็นของผู้ใช้มีความสอดคล้องกันสูง ไม่มีความเห็นที่แตกต่างกันอย่างมีนัยสำคัญ จากผลดังกล่าวสะท้อนว่าระบบสามารถตอบสนองความคาดหวังของผู้ใช้ รูปแบบการแสดงผลคำศัพท์สามภาษาได้รับการยอมรับ และฟังก์ชันอ่านออกเสียงตอบโต้มีประสิทธิภาพการใช้งาน

4.5.4 การสรุปและวิเคราะห์ข้อคิดเห็นจากผู้ใช้

จากการเก็บรวบรวมข้อคิดเห็นเชิงคุณภาพจากอาสาสมัครผู้ทดลองใช้งานทั้ง 5 ท่าน เกี่ยวกับความพึงพอใจในการใช้งานแอปพลิเคชัน VocabEyes สามารถจำแนกประเด็นสำคัญได้ดังนี้

1. ด้านการออกแบบส่วนต่อประสานกับผู้ใช้

ผู้ใช้งานส่วนใหญ่มีความเห็นสอดคล้องกันว่าแอปพลิเคชันมี โครงสร้างที่เรียบง่าย และไม่ซับซ้อน (Minimalist Design) ซึ่งเป็นจุดเด่นสำคัญที่ช่วยให้ผู้พิการทางสายตาทั้งกลุ่มที่สายตาเลือนราง (Low Vision) และตาบอดสนิท (Blind) เข้าถึงการใช้งานได้ทันที โดยมีจุดเด่นที่สำคัญ คือ ระบบการสัมผัสหน้าจอส่วนใดก็ได้เพื่อส่งถ่ายภาพ ช่วยลดอุปสรรคในการค้นหาปุ่มคำสั่ง ทำให้การใช้งานมีความคล่องตัวสูง

2. ด้านระบบเสียงและการนำเสนอภาษา

ในส่วนของการอ่านออกเสียงคำศัพท์ 3 ภาษา (ไทย-ญี่ปุ่น-อังกฤษ) ผู้ใช้งานระบุว่า ลำดับการออกเสียงและการเรียบเรียงภาษาทำออกมาได้ดี ฟังเข้าใจง่าย และมีความต่อเนื่อง และแอปพลิเคชัน ไม่เพียงแต่มีประโยชน์ต่อผู้พิการทางสายตาเท่านั้น แต่ยังมีศักยภาพในการเป็นสื่อการเรียนรู้สำหรับเด็กปฐมวัย เนื่องจากผู้ทดสอบในระยะเริ่มต้นชาวญี่ปุ่นได้ให้ข้อเสนอแนะว่า ด้วยส่วนติดต่อผู้ใช้งานที่ถูกออกแบบมาให้มีความเรียบง่าย และความสามารถในการเชื่อมโยงคำศัพท์ทั้งสามภาษา (ไทย ญี่ปุ่น และ อังกฤษ) เข้าด้วยกันได้อย่างราบรื่น แอปพลิเคชันนี้จึงมีศักยภาพที่จะสามารถนำไปแนะนำให้ เป็นเครื่องมือส่งเสริมการเรียนรู้ภาษาผ่านการเล่น (Play-Based Language Learning Tool) สำหรับเด็กปฐมวัยที่มีความอยากรู้อยากเห็นสูงและยังอยู่ในช่วงวัยของการเรียนรู้ภาษาแม่ได้อีกด้วย เช่น สามารถสนุกสนานไปกับการใช้งานในการดูคำทำนายว่าตรงกับสิ่งของที่นำมาทำนายหรือไม่ เป็นต้น

ในประเด็นเรื่องผลกระทบที่อาจเกิดขึ้นจากความแตกต่างของภาษาของระบบ Text-to-Speech (TTS) หรือคุณภาพการแปลจากพจนานุกรมที่มีต่อความพึงพอใจของผู้ใช้นั้น จากการสอบถามผู้เข้าร่วมทดสอบโดยตรงได้รับคำแนะนำที่เป็นประโยชน์อย่างยิ่ง ดังนี้

- คุณภาพเสียง ผู้ทดสอบยืนยันว่าการออกเสียงของ TTS มีความชัดเจนและเข้าใจง่าย
- คุณภาพการนำเสนอภาษา (การแปลพจนานุกรม)

- กลุ่มผู้ทดสอบชาวไทย ไม่ได้ระบุถึงข้อกังวลใดๆ ในส่วนนี้
- กลุ่มผู้ทดสอบชาวญี่ปุ่น ตั้งข้อสังเกตว่า แม้คำศัพท์บางคำที่แปลมาจะอยู่ในหมวดหมู่ ความหมายที่ถูกต้อง แต่อาจไม่ใช่คำที่ใช้กันอย่างเป็นธรรมชาติในบริบทชีวิตประจำวัน โดยข้อจำกัดนี้เกิดขึ้นเนื่องจากระบบปัจจุบันจะดึงคำศัพท์ในลำดับแรก (คำหลัก) ที่เชื่อมโยงกับรหัส WordNet ID มาโดยอัตโนมัติ ซึ่งบางครั้งอาจเป็นคำพ้องความหมายที่พบได้น้อย ยกตัวอย่างเช่น ภาษาอังกฤษคำว่า “Refrigerator” ถูกจับคู่กับคำว่า “アイスボックス” (Aisu bokkusu หรือ Icebox) ตามคำใน WordNet ซึ่งแม้ว่าในเชิงโครงสร้างและอรรถศาสตร์จะต้องตามกลุ่มคำพ้อง (Synonym Set): アイスボックス (Icebox) | 冷蔵庫 (Refrigerator) แต่ในการสนทนาทั่วไปจะฟังดูไม่เป็นธรรมชาติ ซึ่งคำที่เหมาะสมกว่าควรจะเป็น “冷蔵庫” (Reizouko) ในทำนองเดียวกัน คำว่า “Fan” ถูกจับคู่กับคำว่า “ブロウ” (Blower) เนื่องจากเป็นคำที่ปรากฏเป็นอันดับแรกในกลุ่มคำพ้อง: ブロウ

(Blower) | ブロワー (Blower) | 扇風機 (Fan) | 電気扇 (Electric fan) ในขณะที่คำว่า “扇風機” (Senpuuki) เป็นคำที่เหมาะสมกว่าสำหรับใช้ในชีวิตประจำวัน

อย่างไรก็ตาม แม้จะมีการตั้งข้อสังเกตถึงความคลาดเคลื่อนทางภาษาเหล่านี้ กลุ่มผู้ทดสอบชาวญี่ปุ่นกลับให้คะแนนความพึงพอใจ (SUS scores) สูงมากเป็นพิเศษ (95–100 คะแนน) ในขณะที่กลุ่มผู้ทดสอบชาวไทยซึ่งไม่พบปัญหาด้านการแปลเลย กลับให้คะแนนที่ต่ำกว่า (72.5–77.5 คะแนน)

ความสัมพันธ์ในเชิงผกผันนี้บ่งชี้ว่า จากกลุ่มตัวอย่างขนาดเล็กในการทดสอบเบื้องต้นนี้ ($n=5$) คุณภาพการแปลอาจไม่ใช่ปัจจัยหลักที่ส่งผลต่อความพึงพอใจโดยรวม หรือเป็นสาเหตุของความต่างของคะแนน

ข้อกังวลเพียงอย่างเดียวที่ผู้เข้าร่วมทดสอบทั้งสองกลุ่มมีร่วมกันคือ การจำแนกประเภทวัตถุผิดพลาดในบางครั้ง เมื่อเจอกับวัตถุที่มีความซับซ้อนหรืออยู่นอกเหนือขอบเขตข้อมูล อย่างไรก็ตาม หลังจากผู้วิจัยได้อธิบายว่าความแม่นยำในการจำแนกประเภทของโมเดลนั้นถูกประเมินแยกเป็นคนละดัชนีชี้วัดกับ “ความง่ายในการใช้งานแอปพลิเคชัน” ผู้ทดสอบจึงเข้าใจขอบเขตดังกล่าวและไม่ถือว่าการจำแนกผิดพลาดนั้นเป็นอุปสรรคต่อการใช้งาน

ดังนั้น จึงมีการตั้งสมมติฐานว่าความแตกต่างของคะแนน SUS อาจได้รับอิทธิพลจากปัจจัยภายนอกอื่นๆ ตามที่กล่าวมาข้างต้น มากกว่าที่จะเป็นผลมาจากองค์ประกอบด้านภาษาหรือเสียงของระบบ

3. ข้อเสนอแนะเพื่อการพัฒนาในอนาคต

จากการวิเคราะห์คำแนะนำ ผู้ทดลองใช้แอปพลิเคชันได้เสนอประเด็นที่น่าสนใจเพื่อเพิ่มประสิทธิภาพของแอปพลิเคชัน ดังนี้

- 1) ควรมีการแจ้งเตือนระดับความน่าเชื่อถือของรูปภาพ (Image Confidence Score) ผ่านระบบเสียง เพื่อให้ผู้พิการทางสายตามั่นใจในความถูกต้องของวัตถุที่ระบบตรวจจับได้
- 2) ผู้ใช้ต้องการฟังก์ชันในการปรับความเร็วของเสียงอ่าน (Voice Speed) และความสามารถในการเลือกฟังเฉพาะภาษาใดภาษาหนึ่งตามความต้องการส่วนบุคคล

ตารางที่ 4.4 สรุปข้อคิดเห็นจากผู้ใช้งาน

ประเด็นการวิเคราะห์	สรุปข้อคิดเห็นจากผู้ใช้งาน
ความง่ายในการใช้งาน	อินเตอร์เฟซเรียบง่าย เข้าใจหลักการทำงานได้ง่าย
ระบบการสั่งงาน	การสัมผัสหน้าจอส่วนใดก็ได้เพื่อถ่ายภาพ ช่วยอำนวยความสะดวกให้ผู้พิการทางสายตา

ตารางที่ 4.4 สรุปข้อคิดเห็นจากผู้ใช้งาน (ต่อ)

ประเด็นการวิเคราะห์	สรุปข้อคิดเห็นจากผู้ใช้งาน
คุณภาพเสียง	เสียงอ่านชัดเจน ลำดับภาษา (ไทย-ญี่ปุ่น-อังกฤษ) เข้าใจง่าย
โอกาสในการต่อยอด	สามารถนำไปใช้เป็นสื่อการสอนภาษาสำหรับเด็ก หรือผู้เริ่มเรียนภาษาได้
สิ่งที่ควรพัฒนา	การเพิ่มเสียงพูดแจ้งระดับความมั่นใจของการทำนาย การปรับความเร็ว และการเลือกภาษา

จากข้อคิดเห็นที่ได้รับสะท้อนให้เห็นว่า แอปพลิเคชัน VocabEyes ในงานวิจัยนี้บรรลุเป้าหมายในการออกแบบเพื่อมวลชน (Universal Design) ที่เน้นความเรียบง่ายในการใช้งาน แต่ควรเพิ่มฟีเจอร์ด้านการควบคุมให้มากขึ้น เพื่อยกระดับประสบการณ์ผู้ใช้งานให้ตอบโจทย์เฉพาะบุคคลมากยิ่งขึ้น

4.6 วิเคราะห์ความสอดคล้องกับข้อกำหนดและคุณสมบัติหลัก

จากบทที่ 3 ได้มีการกำหนดข้อกำหนดและคุณสมบัติหลัก (Requirements & Features) ของแอปพลิเคชัน โดยสังเขปดังนี้

1. การรู้จำวัตถุจากภาพนิ่ง
2. การประมวลผลและแสดงคำศัพท์
3. การอ่านออกเสียงคำศัพท์
4. การทำงานแบบไม่พึ่งพาอินเทอร์เน็ต

เมื่อพิจารณาผลการประเมินทั้งหมด สามารถวิเคราะห์ได้ดังนี้

4.6.1 ด้านการรู้จำวัตถุ ค่า Accuracy Top-1 Accuracy รวม 93.83% และค่า Top-5 Accuracy รวม 98.33% แสดงให้เห็นว่าระบบสามารถทำหน้าที่หลักได้อย่างมีประสิทธิภาพ

4.6.2 ด้านการประมวลผลและแสดงคำศัพท์ จากการทดสอบการใช้งาน ไม่พบกรณีที่ระบบไม่แสดงผลคำศัพท์ จึงถือว่าฟังก์ชันนี้ทำงานสอดคล้องกับที่ออกแบบไว้

4.6.3 ด้านการอ่านออกเสียง ผู้ใช้ให้คะแนนความพึงพอใจในระดับสูงมาก โดยมีค่าเฉลี่ย 4.80 จากคะแนนเต็ม 5.00 คะแนน สะท้อนว่าฟังก์ชัน Text-to-Speech (TTS) สามารถสนับสนุนประสบการณ์การเรียนรู้ได้จริง

4.6.4 ด้านการใช้งานแอปพลิเคชันโดยรวม คะแนน SUS ที่อยู่ในระดับ Excellent โดยมีค่าเฉลี่ยอยู่ที่ 89 คะแนน จากคะแนนเต็ม 100 คะแนน ยืนยันว่าแนวทางการออกแบบอินเทอร์เฟซสามารถนำไปสู่ผลลัพธ์ที่ดีในทางปฏิบัติ

4.7 การอภิปรายผลลัพธ์

เมื่อพิจารณาผลลัพธ์ทั้งหมดร่วมกัน พบว่าระบบมีความแม่นยำค่อนข้างสูง ผู้ใช้ประเมินว่าใช้งานง่าย และความพึงพอใจอยู่ในระดับสูงมาก จึงสามารถสรุปได้ว่า แอปพลิเคชัน VocabEyes สามารถบรรลุวัตถุประสงค์ของงานวิจัยทั้งในมิติทางเทคนิคและมิติการใช้งาน

4.8 สรุปผลการดำเนินงานวิจัยโดยสังเขป

ผลการดำเนินงานแสดงให้เห็นว่า

- ค่า Top-1 Accuracy = 93.83%
- ค่า Top-5 Accuracy = 98.33%
- ค่าเฉลี่ย SUS = 89.00 (ระดับ Excellent)
- ค่าเฉลี่ยความพึงพอใจ = 4.80/5.00

จากผลลัพธ์ทั้งหมดโดยภาพรวมสอดคล้องกับ Requirements และแผนการประเมินที่กำหนด จึงสามารถถือได้ว่าแอปพลิเคชันต้นแบบ VocabEyes ประสบความสำเร็จตามกรอบการวิจัยที่กำหนดไว้

บทที่ 5

บทสรุป อภิปราย และข้อเสนอแนะ

การวิจัยเรื่องการพัฒนาแอปพลิเคชัน VocabEyes สำหรับการรู้จำวัตถุ เพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) สำหรับผู้พิการทางสายตา สามารถสรุปผลการดำเนินงาน อภิปรายผล และให้ข้อเสนอแนะเพื่อการพัฒนาในอนาคตได้ดังนี้

5.1 สรุปผลการวิจัย

จากการพัฒนาแอปพลิเคชันต้นแบบและทดสอบประสิทธิภาพตามที่ระบุในบทที่ 4 ผลการวิจัยสรุปได้ตามประเด็นสำคัญดังนี้

1. ด้านประสิทธิภาพการรู้จำวัตถุ

ระบบมีความแม่นยำในการทำนายอันดับแรก (Top-1 Accuracy) เฉลี่ยที่ 93.83% และความแม่นยำใน 5 อันดับแรก (Top-5 Accuracy) สูงถึง 98.33% ซึ่งถือว่าบรรลุผลตาม Functional Requirement ที่กำหนดไว้

2. ด้านความสามารถในการใช้งาน (Usability)

จากการประเมินด้วยแบบวัด SUS (System Usability Scale) พบว่าแอปพลิเคชันได้คะแนนเฉลี่ย 89 คะแนน ซึ่งจัดอยู่ในระดับ Excellent (Grade A) สะท้อนว่าระบบใช้งานง่าย ไม่ซับซ้อน และผู้ใช้ไม่ต้องเรียนรู้มากก่อนใช้งาน

3. ด้านความพึงพอใจของผู้ใช้

ผู้ใช้งานกลุ่มตัวอย่างมีความพึงพอใจต่อระบบในภาพรวมในระดับ สูงมาก โดยมีค่าเฉลี่ยอยู่ที่ 4.80 จาก 5.00 คะแนน

5.2 การอภิปรายผล

จากการดำเนินงานวิจัย มีประเด็นที่น่าสนใจในการอภิปรายผลดังนี้

1. ประสิทธิภาพของโมเดลการรู้จำวัตถุ

ความแตกต่างระหว่างค่า Top-1 และ Top-5 Accuracy ที่มีเพียงประมาณ 4.5% แสดงให้เห็นว่าความคลาดเคลื่อนส่วนใหญ่ของระบบเกิดจากการจัดลำดับความสำคัญของวัตถุที่มีลักษณะใกล้เคียงกัน ไม่ใช่การจำแนกประเภทผิดพลาดอย่างสิ้นเชิง สิ่งนี้ยืนยันว่าโมเดลที่เลือกใช้มีความเหมาะสมกับการนำไปเป็นเครื่องมือช่วยเรียนรู้คำศัพท์เบื้องต้น

2. การออกแบบส่วนติดต่อผู้ใช้ (UI) และประสบการณ์ผู้ใช้ (UX)

คะแนน SUS ที่สูงถึง 89 คะแนน และค่าส่วนเบี่ยงเบนมาตรฐานที่ต่ำ (SD 13.06) บ่งบอกว่าการออกแบบอินเทอร์เฟซของ VocabEyes มีความสอดคล้องกับความต้องการของผู้ใช้ที่หลากหลาย กลุ่ม ตั้งแต่ผู้เชี่ยวชาญด้านระบบสารสนเทศไปจนถึงบุคลากรทางการศึกษา ทำให้ระบบสามารถตอบโจทย์การใช้งานจริงได้โดยไม่สร้างภาระแก่ผู้ใช้

3. การตอบสนองต่อ Functional Requirements ของระบบ

สามารถทำงานได้ตามที่ออกแบบไว้ครบถ้วน ทั้งการรู้จำวัตถุ การแสดงผลในภาษาต่างๆ และการอ่านออกเสียงโดยไม่พึ่งพาอินเทอร์เน็ต ซึ่งมีความพึงพอใจที่สูงมาก อยู่ที่ 4.80 คะแนน จากคะแนนเต็ม 5 คะแนน เป็นการยืนยันว่าแอปพลิเคชันมีแนวโน้มที่สามารถเป็นตัวช่วยในการเรียนรู้ภาษาต่างประเทศได้จริง

5.3 ปัญหา อุปสรรค และข้อจำกัดในการวิจัย

ในการดำเนินงานวิจัยครั้งนี้ ผู้วิจัยพบปัญหา อุปสรรค และมีข้อจำกัดบางประการที่ส่งผลกระทบต่อกระบวนการพัฒนาและประสิทธิภาพของแอปพลิเคชันเฉพาะประเด็นหลัก ดังนี้

1. ข้อจำกัดด้านการเข้าถึงกลุ่มเป้าหมายและการทดสอบระบบ เนื่องจากข้อจำกัดในการหากลุ่มตัวอย่างที่เป็นผู้พิการทางสายตา ผู้วิจัยจึงได้ใช้วิธีการจำลองสถานการณ์โดยให้บุคคลที่มีสายตาปกติ ทดสอบการใช้งานแอปพลิเคชันในสภาวะปิดตา เพื่อเทียบเคียงพฤติกรรมและความยากลำบากในการใช้งาน ซึ่งแม้จะช่วยให้เห็นปัญหาหรือข้อแนะนำในเบื้องต้น แต่อาจมีรายละเอียดความต้องการเชิงลึกบางประการที่แตกต่างจากผู้พิการทางสายตาจริง งานวิจัยนี้จึงเป็นการประเมินความสามารถในการใช้งานเบื้องต้น (Preliminary Usability Evaluation) เพื่อเตรียมความพร้อมในการนำไปสู่การทดสอบทางคลินิก (Clinical validation) กับผู้พิการจริงในอนาคต

2. ข้อจำกัดของการไม่มีเกณฑ์ขั้นต่ำของค่าความมั่นใจ (Absence of Confidence Thresholds) เนื่องจากวัตถุประสงค์หลักของแอปพลิเคชัน VocabEyes ให้ความสำคัญกับการเรียนรู้คำศัพท์อย่างต่อเนื่อง มากกว่าการระบุวัตถุในโลกจริงอย่างเข้มงวดตายตัว ระบบจึงจงใจละเว้นการตั้งเกณฑ์คะแนนความมั่นใจขั้นต่ำที่จะไปปิดกั้นการแสดงผล เช่น การแจ้งว่า ‘ไม่พบวัตถุ’ หรือไม่สามารถระบุได้ เมื่อค่าความมั่นใจอยู่ในระดับต่ำ ทั้งนี้ การให้ความสำคัญกับการแสดงผลคำศัพท์ออกมอย่างต่อเนื่องจะช่วยรักษากระแสการเรียนรู้ (Learning Flow) ของผู้ใช้ให้ดำเนินไปอย่างไม่ขาดตอน

3. ข้อจำกัดด้านอุปกรณ์และการประมวลผลแบบออฟไลน์ เนื่องจากการประมวลผลทั้งหมด ทั้งการรู้จำวัตถุและการการสังเคราะห์เสียง เกิดขึ้นบนอุปกรณ์สมาร์ตโฟนโดยตรง (On-Device Processing) เพื่อให้สามารถใช้งานได้โดยไม่ต้องเชื่อมต่ออินเทอร์เน็ต (Offline) ความเร็วในการตรวจจับวัตถุจึงขึ้นอยู่กับประสิทธิภาพของหน่วยประมวลผล (CPU/GPU) ของสมาร์ตโฟน ซึ่งอาจส่งผลให้เกิดความร้อนสะสม

หรือการใช้พลังงานแบตเตอรี่ที่สูงกว่าปกติหากมีการใช้งานต่อเนื่องเป็นเวลานาน เพื่อให้เห็นภาพในเชิงปริมาณ จากการทดสอบเบื้องต้น ณ อุณหภูมิห้องปกติโดยไม่เปิดเครื่องปรับอากาศในประเทศไทยช่วงเดือนเมษายน พบว่าแบตเตอรี่ลดลงไป 7% (จาก 86% เหลือ 79%) ภายในระยะเวลาการใช้งานต่อเนื่อง 10 นาที โดยมีการจับภาพและประมวลผลประมาณ 10 ภาพต่อนาที นอกจากนี้ ตัวเครื่องสมาร์ทโฟนยังมีความร้อนเพิ่มขึ้นอย่างเห็นได้ชัดเมื่อสัมผัส

4. ข้อจำกัดด้านชุดข้อมูลและคลาสของวัตถุ เนื่องจากระบบถูกออกแบบมาให้ทำงานแบบออฟไลน์ ชุดข้อมูลที่ใช้ในการฝึกสอนจึงถูกจำกัดอยู่เพียงคลาสที่มีอยู่ในชุดข้อมูล ImageNet 21K-P (11,221 Classes) เท่านั้น การเพิ่มความสามารถในการรู้จำวัตถุใหม่ๆ นอกเหนือจากที่กำหนดไว้ ไม่สามารถทำได้ในทันทีผ่านการอัปเดตข้อมูลบนระบบคลาวด์ แต่จำเป็นต้องผ่านกระบวนการปรับปรุงโครงสร้างของแอปพลิเคชันและอัปเดตตัวไฟล์โมเดลภายในเครื่องใหม่อีกครั้ง

5. ข้อจำกัดด้านเครื่องมือวิจัย ในขั้นตอนการจัดทำรูปเล่มวิทยานิพนธ์ก็เกิดปัญหาการจัดหน้ากระดาษเมื่อใช้โปรแกรม Microsoft Word ในคอมพิวเตอร์โน้ตบุ๊ก MacBook ระบบปฏิบัติการ MacOS ที่เกิด Error และโดนบังคับปิดโปรแกรมทุกครั้ง ทำให้ไม่สามารถใช้งานได้ จึงทำให้มีความยากลำบากเป็นอย่างมากในการจัดทำเอกสาร เนื่องจากต้องจัดหน้ากระดาษให้ออกมาสวยงาม แต่ไม่สามารถมองเห็นรูปลักษณะในการพิมพ์ออกมาจริงได้ทันทีดังเช่นการใช้งานในเครื่องระบบปฏิบัติการ Windows ผู้วิจัยจึงพยายามหาวิธีแก้ไขเฉพาะหน้า เนื่องจากทดลองใช้ใน Microsoft Word แบบออนไลน์แล้วก็ใช้งานไม่ได้เช่นกัน จึงลองประยุกต์การใช้งานใน Google Docs ซึ่งเป็นแบบออนไลน์ แต่ทว่าขนาดตัวอักษร (Font Size) ที่แสดงข้อมูลกับตอนที่พิมพ์ออกมาจริง จะมีขนาดไม่เท่ากัน เช่น ที่หน้าจอการพิมพ์เห็นว่ามีประมาณ 100 หน้ากระดาษ แต่เมื่อสั่งปริ้นการพิมพ์จะแสดงให้เห็นว่าอยู่ที่ประมาณไม่ถึง 80 หน้า ซึ่งมีความต่างกันเป็นอย่างมาก ทำให้ต้องกดปุ่มปริ้นทุกครั้งเพื่อตรวจสอบผลลัพธ์จริง จึงเป็นอีกหนึ่งอุปสรรคที่สำคัญซึ่งทำให้เสียเวลาเป็นอย่างมากและเพิ่มขั้นตอนในการดำเนินงานจัดทำรูปเล่มวิทยานิพนธ์ด้วยตลอดระยะเวลาที่ผ่านมา

6. ข้อจำกัดด้านความสมบูรณ์ของงานวิจัย จากข้อจำกัดด้านที่กล่าวมาข้างต้น อาจส่งผลให้รายละเอียดบางประการหรือความสมบูรณ์ของแอปพลิเคชันในเวอร์ชันต้นแบบยังมีส่วนที่บกพร่องหรือไม่ครอบคลุมทุกมิติตามที่คาดหวัง ผู้วิจัยจึงได้ตระหนักถึงข้อบกพร่องดังกล่าวและนำมาสรุปเป็นแนวทางไว้ในส่วนของข้อจำกัดและข้อเสนอแนะเพื่อให้ผู้ที่สนใจสามารถนำไปพัฒนาต่อยอดให้มีความสมบูรณ์มากยิ่งขึ้นในอนาคต

5.4 ข้อเสนอแนะ

5.4.1 ข้อเสนอแนะในการนำไปใช้งาน

1. การใช้งานในสถานการณ์ที่ซับซ้อน เช่น เห็นวัตถุบางส่วน มีวัตถุหลายอย่างในภาพ วัตถุเล็กพื้นหลังใหญ่ เนื่องจากผลการทดสอบความแม่นยำ (Accuracy) จากเงื่อนไขที่กำหนดสำหรับเงื่อนไขดังกล่าว มีค่าความแม่นยำที่ลดลงเมื่อเทียบกับการถ่ายภาพในสถานการณ์อื่นในการใช้งานจริง หากเป็นผู้มีสายตาศักดิ์สิทธิ์ถึงสภาพการณ์ดังกล่าว ที่อาจส่งผลกระทบต่อความแม่นยำของการประมวลผลของแอปพลิเคชันซึ่งใช้โมเดลสำเร็จรูปในการแยกแยะวัตถุ

2. การฝึกฝนภาษา ไม่เพียงผู้พิการทางสายตาเท่านั้น สำหรับผู้ใช้สายตาศักดิ์สิทธิ์สามารถใช้แอปพลิเคชันนี้เป็นเครื่องมือเสริมในการเรียนรู้คำศัพท์ด้วยตนเองได้เช่นกัน โดยไม่ว่าจะเป็นวัยเด็กที่มีความอยากรู้อยากลอง หรือวัยผู้ใหญ่ ทั้งชาวไทยหรือชาวต่างชาติที่ต้องการเรียนรู้คำศัพท์ภาษาต่างประเทศจากสิ่งรอบตัว โดยสามารถเรียนรู้คำศัพท์ 3 ภาษาพร้อมกันซึ่งเป็นจุดเด่นของระบบ หรืออย่างน้อยหนึ่งในสามภาษาเป้าหมายตามคลังคำศัพท์ที่มีอยู่

5.4.2 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

1. การขยายผลการทดสอบกับกลุ่มผู้พิการทางสายตาโดยตรง ควรมีการทดสอบระบบกับกลุ่มผู้พิการทางสายตาในหลากหลายช่วงวัยและระดับความบกพร่องทางการมองเห็น เช่น ตาบอดสนิทและผู้มีสายตาเลือนราง ในสภาพแวดล้อมที่หลากหลาย เพื่อเก็บข้อมูลเชิงพฤติกรรมและการตอบสนองต่อระบบเสียง หรือการใช้งาน เพื่อนำมาปรับปรุง User Experience (UX) ให้ตรงตามความต้องการจริงมากที่สุด เนื่องจากจินตนาการเชิงพื้นที่และพฤติกรรมการโต้ตอบของผู้ที่ถูกจำลองการปิดตานั้น มีความแตกต่างจากผู้พิการทางสายตาจริงอย่างสิ้นเชิง ดังนั้นข้อมูลพฤติกรรมจากการศึกษานี้จึงต้องถูกตีความด้วยความระมัดระวัง และจำเป็นอย่างยิ่งที่จะต้องมีการทดสอบทางคลินิก (Clinical validation) กับผู้พิการจริงในอนาคต

2. การยกระดับการแจ้งผลลัพธ์ (Enhanced Result Communication) เพื่อป้องกันการได้รับข้อมูลที่ผิดพลาด การพัฒนาในอนาคตควรเพิ่มฟังก์ชันการอ่านออกเสียงค่าเปอร์เซ็นต์ความมั่นใจของ AI (AI confidence percentage) ควบคู่ไปกับผลการทำนาย แนวทางดังกล่าวซึ่งเป็นข้อเสนอแนะโดยตรงจากผู้ทดสอบเบื้องต้นชาวญี่ปุ่น จะช่วยให้มั่นใจได้ว่าผู้ใช้สามารถเรียนรู้คำศัพท์ใหม่ๆ ได้อย่างต่อเนื่อง ในขณะที่เดียวกันก็ยังคงตระหนักถึงระดับความน่าเชื่อถือในการทำนายของระบบได้อย่างชัดเจน

3. การปรับปรุงพจนานุกรมคลังคำศัพท์ จากการที่ขณะเก็บข้อมูลในการทำแบบสอบถามได้รับข้อคิดเห็นจากชาวญี่ปุ่นเรื่องคำที่ใช้ว่า คำศัพท์ของวัตถุบางคำ อาจไม่ใช่คำที่มีความหมายในการใช้งานทั่วไป ดังนั้น การใช้คลังคำศัพท์โดยอ้างอิงจาก WordNet อย่างเดียวอาจไม่ตอบโจทย์การใช้งาน

จริงในแต่ละภาษา เนื่องจากคำที่เชื่อมโยงไว้ต่อหนึ่งคำ มีกรณีที่จำนวนมากกว่าหนึ่ง จึงจำเป็นต้องพึ่งพาผู้เชี่ยวชาญ หรือเจ้าของภาษาในการสะสางคลังคำศัพท์ ให้ถูกต้องตามการใช้งานจริง

4. การปรับแต่งอุปกรณ์ (Device Optimization) เพื่อแก้ไขปัญหาการสิ้นเปลืองแบตเตอรี่อย่างรวดเร็วและการสะสมความร้อนที่พบระหว่างการใช้งานต่อเนื่องโดยตรง ควรมีการนำเทคนิคการบีบอัดโมเดล หรือ Model Quantization ซึ่งแนวทางนี้จะช่วยบีบอัดขนาดของโมเดล เพิ่มประสิทธิภาพในการประมวลผล และลดการใช้พลังงานแบตเตอรี่ของอุปกรณ์สมาร์ทโฟนได้อย่างมีประสิทธิภาพ โดยไม่ทำให้ความแม่นยำของระบบลดลงมากนัก ซึ่งจะช่วยให้อุปกรณ์มีความเสถียรและสามารถใช้งานได้ยาวนานยิ่งขึ้น

5. คุณสมบัติการปรับแต่งตามความต้องการ (Customization Features) เพิ่มฟังก์ชันการทำงานต่างๆ เช่น การอนุญาตให้ผู้ใช้สามารถปรับความเร็วของเสียงอ่าน และเลือกเปิดหรือปิดการอ่านเฉพาะบางภาษาได้ เพื่อให้สอดคล้องกับความต้องการในการเรียนรู้ของแต่ละบุคคล

6. การขยายการรองรับภาษา (Extended Language Support) การขยายขอบเขตของโครงสร้างระบบเพื่อเพิ่มการรองรับภาษาอื่น ๆ เช่น ภาษาจีน ซึ่งจะเป็นการยกระดับและขยายขอบเขตประสบการณ์การเรียนรู้ให้กว้างขวางยิ่งขึ้น

7. การเพิ่มจำนวนวัตถุและฐานข้อมูล หากสามารถเพิ่มชุดข้อมูล (Dataset) ของวัตถุให้มีความหลากหลายมากขึ้นโดยที่ไม่เพิ่มขนาดของข้อมูลมากนักและไม่ส่งผลกระทบต่อเวลาในการประมวลผล เพื่อขยายขอบเขตการเรียนรู้ให้กว้างขึ้น

8. การขยายฐานข้อมูลคำศัพท์ เพิ่มจำนวนคำแปลในคลังคำศัพท์ เนื่องจาก WordNet ในภาษาญี่ปุ่นและภาษาไทยยังมีจำนวนคำแปลที่น้อยกว่าในภาษาอังกฤษเพื่อให้สามารถแสดงผลได้ครบทั้งสามภาษาในทุกคำที่โมเดลสามารถทำนายได้

การพัฒนาแอปพลิเคชัน VocabEyes ในครั้งนี้ ไม่เพียงแต่เป็นการพิสูจน์ถึงประสิทธิภาพของเทคโนโลยีการรู้จำวัตถุ (Object Recognition) และการประมวลผลภาษา (Natural Language Processing) ที่สามารถทำงานร่วมกันได้อย่างลงตัวบนแพลตฟอร์มโมบายล์โดยการบูรณาการเทคโนโลยีที่มีอยู่แล้วเข้าด้วยกันเท่านั้น แต่ยังเป็นการสร้างนวัตกรรมที่ตอบโจทย์การเรียนรู้ภาษาแบบไร้พรมแดน ผลการวิจัยที่ปรากฏชัดทั้งในด้านความแม่นยำทางเทคนิคที่สูงถึงร้อยละ 93.83 และระดับความพึงพอใจของผู้ใช้งานที่อยู่ในเกณฑ์ดีเยี่ยมจากกลุ่มตัวอย่าง โดยมีค่าเฉลี่ยอยู่ที่ 4.80 จากคะแนนเต็ม 5 คะแนน เป็นสิ่งยืนยันว่าเทคโนโลยีสามารถเป็นเครื่องมือสำคัญในการลดช่องว่างทางการศึกษา (Education) การเข้าถึง (Accessability) และเป็นการส่งเสริมการเรียนรู้ด้วยตนเองได้อย่างมีประสิทธิภาพสำหรับผู้พิการทางสายตาและทุกคนผ่านการออกแบบเพื่อมวลชน (Universal Design) ที่ใช้งานได้ง่าย และรองรับการใช้งานของมนุษย์ทุกกลุ่มในสังคมอย่างเท่าเทียม

นอกจากนี้ แม้ว่าผู้วิจัยสำเร็จการศึกษาระดับปริญญาตรีในสาขาวิชาภาษาญี่ปุ่น และไม่ได้มีพื้นฐานทางด้านวิทยาการคอมพิวเตอร์มาก่อน หลังจากนั้น ได้เข้าศึกษาต่อในระดับปริญญาโท สาขาวิชาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีไทย-ญี่ปุ่น แต่ตัวแอปพลิเคชันก็ได้รับการพัฒนาจนสำเร็จลุล่วงด้วยการใช้ประโยชน์จากเครื่องมือ Generative AI ที่ก้าวหน้าทางเทคโนโลยีปัจจุบัน เช่น ChatGPT, Gemini เพื่อช่วยในส่วนของการเขียนโค้ดและการแก้ไขจุดบกพร่อง (Debugging) ของแอปพลิเคชัน Flutter ในรูปแบบข้ามแพลตฟอร์มด้วยตนเอง แนวทางนี้แสดงให้เห็นถึงกรอบการทำงานที่ทันสมัยและนำไปใช้ได้จริงสำหรับการพัฒนาผลงานต้นแบบข้ามสาขาวิชา อีกทั้งยังเป็นการเน้นย้ำถึงองค์ความรู้เชิงปฏิบัติที่มีคุณค่าในการประยุกต์ใช้ AI เพื่อการบูรณาการระบบต่อไป

ผู้วิจัยหวังเป็นอย่างยิ่งว่าองค์ความรู้และกระบวนการพัฒนาที่นำเสนอในวิทยานิพนธ์ฉบับนี้จะเป็นส่วนหนึ่งของรากฐานสำคัญและเป็นแรงบันดาลใจให้เกิดการต่อยอดงานวิจัยด้านสื่อการสอนอัจฉริยะเพื่อสร้างสรรค์สังคมแห่งการเรียนรู้ที่ยั่งยืนและก้าวทันต่อการเปลี่ยนแปลงของเทคโนโลยีในอนาคตสืบไป



TNI

บรรณานุกรม

TNI

บรรณานุกรม

- [1] กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ, “สถานการณ์คนพิการในประเทศไทย,” [Online]. Available : <https://app.powerbi.com/view?r=eyJrIjojNTJhMWEYjAtZDBjMi00OTYLTlkOGltNUMzllNzgMDVjliwidCijEZTIYTZhLWUxYzYtNDNhMihNjUwLWIOTYOWY2OGEyMCIsImMiOjEwfQ%3D%3D>. [Accessed : 2 กันยายน 2568].
- [2] World Health Organization, “Blindness and Vision Impairment,” [Online]. Available : <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> [Accessed : September 2, 2025].
- [3] A. Quatraro and M. Paiano, “New ways of language learning for blind or visually impaired children and teenagers,” *In Proc. 3rd Int. Conf. ICT for Language Learning, ICT4LL 2010*, Florence, Italy, November 11-12, 2010, pp. 1-4.
- [4] สถานเอกอัครราชทูตญี่ปุ่นประจำประเทศไทย, “รัฐบาลญี่ปุ่นให้ความร่วมมือสนับสนุนโครงการ “The Project for Constructing a Media Production and Educational Technology Center for the Blind in Khonkaen Province,” [Online]. Available : https://www.th.emb-japan.go.jp/itpr_th/2020_27.html. [Accessed : 5 มกราคม 2569].
- [5] ประกาศกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์, “ประเภทและหลักเกณฑ์ความพิการ,” [Online]. Available : <https://www.ratchakitcha.soc.go.th/DATA/PDF/2552/E/077/2.PDF>. [Accessed : 2 กันยายน 2568].
- [6] สำนักงานส่งเสริมเศรษฐกิจสร้างสรรค์ (องค์การมหาชน), “Generative AI ผู้ช่วยคนใหม่ในยุคแห่งอนาคตสำหรับนักสร้างสรรค์,” [Online]. Available : <https://www.cea.or.th/en/single-research/Generative-AI>. [Accessed : 17 กันยายน 2568].
- [7] IBM, “What is Object Detection?,” [Online]. Available : <https://www.ibm.com/think/topics/object-detection>. [Accessed : September 2, 2025].
- [8] Sciencedirect, “Object Recognition,” [Online]. Available : <https://www.sciencedirect.com/topics/computer-science/object-recognition>. [Accessed : September 2, 2025].
- [9] A. G. Howard and M. Zhu, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” [Online]. Available : <https://arxiv.org/pdf/1704.04861>. [Accessed : September 3, 2025].
- [10] O. Russakovsky et al., “ImageNet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, December 2015.

- [11] R. Szeliski, *Computer vision : Algorithms and applications*, New Jersey : Pearson Prentice Hall, 2022.
- [12] เอมี่ สูงศิริ, “การจำแนกเกรตริงนกแอ่นกินรังด้วยขั้นตอนวิธีการเรียนรู้เชิงลึก,” วิทยานิพนธ์ วท.ม. (เทคโนโลยีสารสนเทศ), สถาบันเทคโนโลยีไทย-ญี่ปุ่น, กรุงเทพฯ, 2565.
- [13] J. Singh and S. Singh, “A review on machine learning aspect in physics and mechanics of glasses,” *Journal of Imaging*, vol. 11, no. 1, pp. 9, January 2025.
- [14] A. G. Howard and M. Zhu, “MobileNets : Open-Source Models for Efficient On-Device Vision,” [Online]. Available : <https://research.google/blog/mobilenets-open-source-models-for-efficient-on-device-vision/>. [Accessed : September 10, 2025].
- [15] Imagent Official Website, “ImageNet,” [Online]. Available : <https://www.image-net.org/>. [Accessed : October 3, 2025].
- [16] J. Deng et al., “ImageNet : A Large-Scale Hierarchical Image Database,” [Online]. Available : https://www.image-net.org/static_files/papers/imagenet_cvpr09056.pdf. [Accessed : September 10, 2025].
- [17] ธนัท ลีลาน้อย, “การสร้างเครือข่ายคำไทยของมโนทัศน์พื้นฐาน,” วิทยานิพนธ์ วท.ม. (การจัดการอุตสาหกรรม), จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2551.
- [18] Global WordNet Association, “Open English Wordnet,” [Online]. Available : <https://github.com/globalwordnet/english-wordnet?tab=readme-ov-file>. [Accessed : February 6, 2026].
- [19] Princeton University, “Open English Wordnet,” [Online]. Available : <https://wordnet.princeton.edu/>. [Accessed : February 6, 2026].
- [20] Flutter, “Flutter Architectural Overview,” [Online]. Available : <https://docs.flutter.dev/resources/architectural-overview>. [Accessed : October 12, 2025].
- [21] Google Cloud, “Supported Voices and Languages,” [Online]. Available : <https://cloud.google.com/text-to-speech/docs/list-voices-and-types>. [Accessed : October 3, 2025].
- [22] S. Saychum et al., “A bi-lingual thai-english TTS system on android mobile devices,” *International Conference on Electrical Engineering/Electronics*, ICIT 2012, Jeju, Korea (South), May 17-19, 2012, pp. 1,482-1,484.
- [23] วราภรณ์ มามี, “เกณฑ์มาตรฐานในการออกแบบห้องน้ำสาธารณะสำหรับผู้พิการด้านการเดินภายใต้แนวคิด การออกแบบเพื่อมวลชน,” วิทยานิพนธ์ วศ.ม. (วิศวกรรมสารสนเทศ), สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, กรุงเทพฯ, 2554.

- [24] J. Brooke, (1996). “*SUS-A Quick and Dirty Usability Scale*,” [Online]. Available : https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale. [Accessed : October 3, 2025].
- [25] เกริกเกียรติ สมศรี, “การปรับปรุงประสิทธิภาพการจัดการลูกค้าสัมพันธ์ (CRM) ในธุรกิจร้านขายยา ด้วยการประยุกต์ใช้เทคโนโลยี Chatbot และ Chat GPT,” วิทยานิพนธ์ วท.บ. (เทคโนโลยีสารสนเทศ), สถาบันเทคโนโลยีไทย-ญี่ปุ่น, กรุงเทพฯ, 2567.
- [26] พัชรวดี เดชะ และคณะ, “การพัฒนาต้นแบบโปรแกรมประเมินและออกแบบการรักษาเฉพาะบุคคลสำหรับโรคออฟฟิศซินโดรม : การศึกษานำร่อง,” *วารสารกายภาพบำบัด*, ปีที่ 47, ฉบับที่ 2, หน้า 100, พฤษภาคม-สิงหาคม 2568.
- [27] A. Howard et al., “Multi-stage local path planning algorithm based on obstacles as anchor points,” *International Conference on Advanced Algorithms and Control Engineering, ICAACE 2024*, Shanghai, China, March 1-3, 2024, pp. 1,440-1,445.
- [28] M. Bahaa et al., “A Smart AIoT-Based Mobile Application for Plant Disease Detection and Environment Management in Small-Scale Farms Using MobileViT,” [Online]. Available : https://www.researchgate.net/publication/399281972_A_Smart_AIoT-Based.. [Accessed : January 30, 2026].
- [29] T. Ridnik et al., “*ImageNet-21K Pretraining for the Masses*,” [Online]. Available : <https://arxiv.org/pdf/2104.10972>. [Accessed : November 11, 2025].
- [30] H. M Fadhil et al., “Smartphone-based object recognition application for educational engagement,” *In Proceedings of the 4th International Conference on Distributed Sensing and Intelligent Systems, ICDSIS 2023*, Shenyang, China, July 27-29, 2023, pp. 212-217.
- [31] ประเมศวร์ เบญจวรรณ, “แนวทางการออกแบบโมบายแอปพลิเคชันในการเรียนรู้เชิงรุกสำหรับผู้พิการทางการเห็น,” วิทยานิพนธ์ วศ.ม. (เทคโนโลยีและสื่อสารการศึกษา), จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2561.
- [32] เสกสรร อามาศย์มนตรี และอัญชลี สุวัชนน, “สมาร์ทโฟน เทคโนโลยีสิ่งอำนวยความสะดวกสำหรับผู้เรียนที่บกพร่องทางการเห็น,” *วารสารเทคโนโลยีและสื่อสารการศึกษา ECT Journal*, ปีที่ 7, ฉบับที่ 17, หน้า 92-104, กรกฎาคม-ธันวาคม 2562.
- [33] R. Budiu, “*Why 5 Participants Are Okay in a Qualitative Study*,” [Online]. Available : <https://www.nngroup.com/articles/5-test-usersqual-quant/>. [Accessed : March 14, 2026].

- [34] American Occupational Therapy Association, "Occupational therapy practice framework : Domain and process American," *Journal of Occupational Therapy*, vol. 34, no. 3, pp. 267-274, March 2025.
- [35] S. J. Lederman et al., "Haptic perception : A tutorial," *International Journal of Computer & Organization Trends*, vol. 71, no. 7, pp. 1,439-1,459, September 2025.
- [36] W. R. Wiener et al., *Foundations of Orientation and Mobility*, New Jersey : Pearson Prentice Hall, 2010.



ภาคผนวก

TNI

ภาคผนวก ก.

เอกสารนำเสนอสอบป้องกันวิทยานิพนธ์

TNI

**การพัฒนาแอปพลิเคชันแบบรู้จำวัตถุด้วยสมาร์ตโฟน “VocabEyes”
เพื่อการเรียนรู้คำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) สำหรับผู้พิการทางสายตา**
DEVELOPMENT OF A SMARTPHONE-BASED OBJECT RECOGNITION PROTOTYPE APPLICATION “VOCABEYES”
FOR TRILINGUAL VOCABULARY LEARNING (THAI-JAPANESE-ENGLISH) FOR THE VISUALLY IMPAIRED

ผู้วิจัย: วรพจน์ บุตรอัน
อาจารย์ประจำภา: อ.ดร.พงษ์สิทธิ์ นฤกุลสาร
สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น

1

บทที่ 1 บทนำ

การพัฒนาแอปพลิเคชัน VocabEyes (แอปพลิเคชันรู้จำวัตถุและอ่านออกเสียง 3 ภาษา)

2

ที่มาและความสำคัญของปัญหา :
ข้อจำกัดของการมองเห็น
และกำแพงแห่งการเรียนรู้ภาษา

171,953 คน

จำนวนผู้พิการทางตาของตนเองในไทย
(7.54% ของผู้พิการทั้งหมด)
จำนวนผู้พิการในประชาชาติทั้งหมด 2,280,409 คน

อุปสรรคทางกายภาพ
การรับรู้สิ่งรอบตัวด้วยสัมผัสผ่านเครื่องช่วยฟังเป็นปกติ

อุปสรรคทางการศึกษา
การเรียนรู้ภาษาต่างประเทศ (โดยเฉพาะคำศัพท์ที่ซับซ้อน)
ขาดเครื่องมือที่ใช้ช่วยจำหรือจดจำสิ่งที่ได้ฟัง

ทางออก (The Solution)
การผสาน Computer Vision เข้ากับ Text-to-Speech (TTS) บนอุปกรณ์พกพาที่เข้าถึงได้ง่าย

อ้างอิงข้อมูล: กรมส่งเสริมการค้าระหว่างประเทศ กระทรวงพาณิชย์ (พ.ศ. 2560)

3

โอกาสทางเทคโนโลยี: การบูรณาการเพื่อขยายข้อจำกัด

สมาร์ตโฟน (Smartphone)
อุปกรณ์พกพาพื้นฐานที่เข้าถึงง่าย
มีกล้องคุณภาพดี ครอบคลุม
ประมวลผลข้อมูล

เทคโนโลยีรู้จำวัตถุ
(Object Recognition)
เป็นแอปพลิเคชันที่ทำงานกับข้อมูล
“ของตา” ให้ข้อมูลที่จำเป็น

เทคโนโลยีเสียงสังเคราะห์
(Text-to-Speech; TTS)
แปลงข้อมูลภาพและข้อความให้เป็นเสียง
เพื่อให้ผู้พิการทางสายตาสามารถรับรู้ได้ทันที

การผสานเทคโนโลยีทั้ง 3 ส่วนนี้ นำไปสู่การพัฒนาเครื่องมือพกพา
ที่ช่วยให้ผู้พิการทางสายตาสามารถเรียนรู้ภาษาต่างประเทศที่สนใจได้ด้วยตนเอง

4

แนะนำ : แอปพลิเคชัน “VocabEyes”

- ระบบปฏิบัติการ Android**
ใช้งานบนเครื่องสมาร์ตโฟน
- รองรับ 3 ภาษา (Trilingual)**
แสดงผลเป็นเสียงภาษาไทย, อังกฤษ และอังกฤษ ตามลำดับ
- ออกแบบเพื่อทุกคน (Accessibility)**
ใช้งานง่ายด้วยตัวอักษรใหญ่ มีการแสดงข้อความบนหน้าจอสำหรับบุคคลที่ตาบอด
- ทำงานแบบออฟไลน์ (Offline Mode)**
เพื่อความรวดเร็ว ความเป็นส่วนตัว และใช้งานได้ทุกที่

5

วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาแอปพลิเคชันต้นแบบสมาร์ตโฟน ระบบปฏิบัติการแอนดรอยด์เร่งด่วนเสียงคำศัพท์สามภาษา สำหรับการรู้จำวัตถุและอ่านเสียงคำศัพท์สามภาษา (ไทย-ญี่ปุ่น-อังกฤษ) เพื่อประโยชน์ทางการศึกษา ภาษาลำบากสำหรับผู้พิการทางการได้ยิน

6

วัตถุประสงค์และขอบเขตหลักของการพัฒนาระบบ

- พัฒนาแอปพลิเคชันต้นแบบระบบปฏิบัติการปฏิบัติการ Android สำหรับใช้งานบนสมาร์ตโฟน
- อ่านออกเสียงคำศัพท์ 3 ภาษาแบบต่อเนื่อง โดยอิงจากข้อมูล WordNet (6,000-10,000 คำ)
- ประมวลผลแบบออฟไลน์ (On-device) เพื่อความเร็ว ความเป็นส่วนตัว และประหยัดค่าใช้จ่าย

7

ขอบเขตการวิจัย: ด้านผู้ใช้งานและเนื้อหา (User & Content Scope)

กลุ่มผู้ใช้งาน (User Scope):

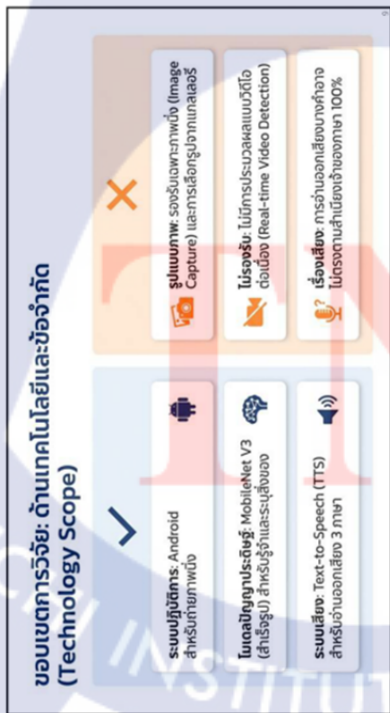
- ผู้พิการทางสายตา (ทั้งตาบอด Blind และตาบอดเห็นเลือนราง Low Vision)
- บุคคลทั่วไปที่ต้องการเรียนรู้คำศัพท์ ภาษาด้านประเทศเบื้องต้น

อ้างอิงข้อมูล ImageNet-21K-P
(จำแนกภาพสิ่งของและสิ่งมีชีวิตทั่วไป 11,221 คลาส)

ภาษาอังกฤษ:	~10,000 คำ
ภาษาไทย:	~8,500 คำ
ภาษาญี่ปุ่น:	~6,000 คำ

(ข้อมูลตามเว็บไซต์ของ ImageNet ไม่ครอบคลุมภาษาญี่ปุ่นและภาษาเกาหลี)

8



9



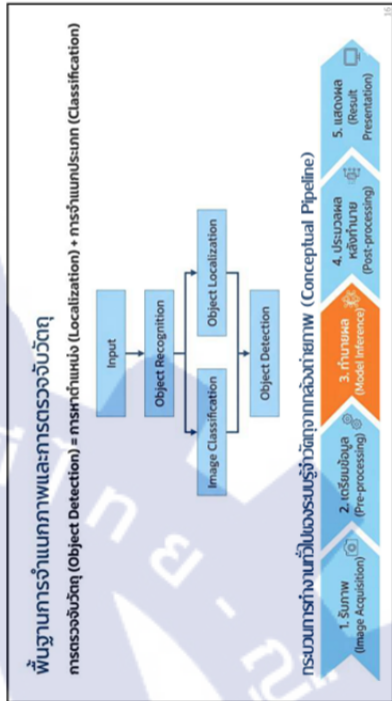
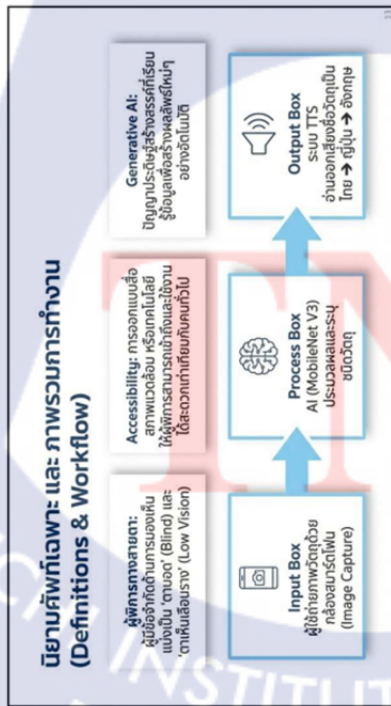
10



11



12



เทคโนโลยีเบื้องหลัง : Deep Learning Convolutional Neural Network (CNN)

จำลองการทำงานของระบบประสาทมนุษย์ (Artificial Neural Network) สำหรับการรับรู้เชิงลึก

- ออกแบบเฉพาะสำหรับจัดการรูปภาพ
- สกัดคุณลักษณะ (Feature Extraction) จากภาพโดยอัตโนมัติ
- เรียนรู้แบบลำดับขั้น (Hierarchical Approach) จากการละเลียดเลี้ยวสู่ภาพรวม

17

สถาปัตยกรรมสำหรับอุปกรณ์พกพา: MobileNet

Depthwise separable convolution

1 x 1 pointwise convolution

Adaptive average pooling

Output

18

สถาปัตยกรรมสำหรับอุปกรณ์พกพา: MobileNet

ข้อจำกัดของฮาร์ดแวร์: ทรัพยากรประมวลผล พลังงาน และหน่วยความจำจำกัด

Standard Convolution

ประมวลผล และใช้ทรัพยากรจำนวนมาก

VS

Depthwise Separable Convolution

✓ - ลดขนาดพารามิเตอร์และลดการใช้หน่วยความจำ
- เพิ่มประสิทธิภาพการทำงาน (On-device Processing)

19

สถาปัตยกรรมสำหรับอุปกรณ์พกพา: MobileNet

Accuracy vs MACCs vs model size

Model Size (MB)

Accuracy (Top-1)

MACCs (Millions of MACCs)

Legend: InceptionV2, MobileNet-A, MobileNetV1, MobileNetV2, MobileNetV3, NasNet, ProxylessNAS, ResNet-50, ShuffleNetV2

*Model Size (MB): ขนาดของโมเดลที่บันทึกไว้บนฮาร์ดแวร์. การคำนวณค่า MACCs: จำนวนการคูณและบวกที่ใช้ในการประมวลผล. การคำนวณค่า Accuracy: ความแม่นยำของโมเดลในการจำแนกประเภท. ข้อมูลอ้างอิง: Deep Learning Library: Convolutional Neural Networks (CNN)

20

ชุดข้อมูลและคลังคำศัพท์ (Datasets & Lexicon)



14,000,000+ ภาพ

ฐานข้อมูลภาพธรรมชาติสำหรับการฝึกฝนโมเดล

รวมข้อมูล ≈ 21,000 คำ (ImageNet-21K)



Open English Wordnet

คลังคำศัพท์และคลังวลี

• ต้องการคำศัพท์ทางจิตวิทยาและจิตวิทยา

• ครอบคลุมคำศัพท์ (Synsets) และวลี

• ใช้สำหรับพัฒนาความรู้เชิงความหมายเพื่อประมวลผลภาษา

21

เครื่องมือพัฒนาแอปพลิเคชัน: Flutter

UI Toolkit จาก Google สำหรับสร้างแอปพลิเคชันประสิทธิภาพสูง

Cross-platform

เขียนโค้ดเดียว (Dart) รองรับทั้ง iOS และ Android

Native Performance

คอมไพล์เป็น Machine Code โดยตรงทำงานไวเหมือนจริง

Hot Reload

เห็นผลลัพธ์การแก้ไขโค้ดทันทีโดยไม่ต้องรีสตาร์ทแอปพลิเคชัน

22

เทคโนโลยีแปลงข้อความเป็นเสียง (Text-to-Speech: TTS)

ใช้ AI/ML เพื่อสังเคราะห์เสียงพูดที่เป็นธรรมชาติ

Text Input

➔

Text Processing

➔

Speech Synthesis

➔

Audio Stream

- รับรู้ข้อความ, ข้อความ, แปลงเป็นเสียงพูด (G2P)
- สังเคราะห์เสียงพูดจากข้อมูลเสียงพูด
- ส่งออกเป็นไฟล์เสียงพูดที่ใช้งานได้

23

แนวคิดการออกแบบเพื่อมวลชน (Universal Design)

เป้าหมาย: สร้างสรรค์สิ่งต่างๆ ให้ทุกคนเข้าถึงและใช้งานได้โดยไม่ต้องปรับเปลี่ยน



1. เสมอภาค (Equitable Use)
2. ยืดหยุ่น (Flexibility in Use)
3. เรียบง่าย (Simple and Intuitive Use)
4. ข้อมูลที่รับรู้ได้ (Perceptible Information)
5. อดทนต่อข้อผิดพลาด (Tolerance for Error)
6. พยายามใช้แรงกายให้น้อยที่สุด (Low Physical Effort)
7. ขนาดพื้นที่และระยะทาง (Size and Space)

24

มาตรวัดประสิทธิภาพของโมเดล AI (Evaluation Metrics)

➤ **Top-1 Accuracy:**
อัตราความถูกต้องเมื่อพิจารณาเฉพาะผลลัพธ์ในอันดับที่ 1
สูงสุดอันดับแรก

Top-1 Accuracy = $\frac{\text{จำนวนการทำนายที่ถูกต้องอันดับที่ 1}}{\text{จำนวนภาพทดสอบทั้งหมด}}$

➤ **Top-k Accuracy:**
อัตราความถูกต้องเมื่อพิจารณาถึงอันดับ k อันดับแรก
ของผลการทำนาย (พิจารณาถึงจุดในการเลือกคำตอบ)

Top-k Accuracy = $\frac{\text{จำนวนการทำนายที่ถูกต้องอันดับที่ 1 ถึง k}}{\text{จำนวนภาพทดสอบทั้งหมด}}$

1. แมว (Cat) 60%
2. สุนัข (Dog) 15%
3. หมี (Bear) 5%

มาตรวัดความสามารถในการใช้งาน (System Usability Scale: SUS)

แบบสอบถามมาตรฐานสากล 10 ข้อ (Likert Scale 5 ระดับ)
เพื่อประเมินภาพรวมความใช้งานง่ายของระบบ

งานวิจัยที่เกี่ยวข้อง: การพัฒนาโมเดล AI

ความก้าวหน้าของ MobileNetV3:
- สามารถปรับ NAS หรือ NetAdapt หรือใช้กับ h-swish
กลไก Squeeze-and-Excitation (SE)**

สนับสนุนด้วยกราฟิก: (ลด Latency ลง 20%)

Google: **Neural Architecture Search (NAS)** คือ AI ที่ออกแบบโครงสร้างโมเดลที่ดีที่สุดโดยอัตโนมัติ
Google: **Guided Neural Architecture Search (GNAS)** เป็น NAS ที่ใช้ความรู้ก่อนหน้าเพื่อปรับปรุงการค้นหาโครงสร้างโมเดล
Practical: **Practical Neural Architecture Search (Practical NAS)** เป็น NAS ที่ใช้ความรู้ก่อนหน้าเพื่อปรับปรุงการค้นหาโครงสร้างโมเดล

การเพิ่มประสิทธิภาพของ ImageNet-21K Pretraining:
- ใช้ฐานข้อมูลขนาดใหญ่ที่ครอบคลุมสัตว์ป่า
- สนับสนุน Semantic Software ที่รองรับกับ WordNet

เพิ่มประสิทธิภาพให้โมเดลโดยลดขนาดของโมเดล

ประสิทธิภาพของ ImageNet-21K Pretraining: ImageNet-21K Pretraining เป็นชุดข้อมูลขนาดใหญ่ที่ครอบคลุมสัตว์ป่า ซึ่งช่วยให้โมเดลสามารถเรียนรู้ลักษณะที่กว้างขวางของสัตว์ป่าได้ดีกว่าชุดข้อมูล ImageNet-1000

****กลไก Squeeze-and-Excitation (SE)**
กลไก Squeeze-and-Excitation (SE) เป็นกลไกที่ช่วยลดขนาดของโมเดลโดยลดจำนวนพารามิเตอร์ลง โดยยังคงรักษาประสิทธิภาพไว้ได้

Efficiency

มาตรฐานและแนวทางการออกแบบ UI สำหรับผู้พิการทางสายตา

อ้างอิงจาก WCAG (WCAG 2.0) และ BBC Mobile Accessibility:

การได้ยิน (Audio Interaction): ต้องมีการแจ้งข้อมูลเสียงเมื่อมีการโต้ตอบ	การมองเห็น (Visual Interaction): ต้องมีการแจ้งข้อมูลภาพเมื่อมีการโต้ตอบ
การได้ยิน (Audio Interaction): ต้องมีการแจ้งข้อมูลเสียงเมื่อมีการโต้ตอบ	การมองเห็น (Visual Interaction): ต้องมีการแจ้งข้อมูลภาพเมื่อมีการโต้ตอบ

แอปพลิเคชันช่วยเหลือผู้พิการทางสายตาในปัจจุบัน

**Seeing AI (Microsoft)** สแกนและอ่านข้อความ/วัตถุ

**Lookout (Google)** ตรวจสอบสิ่งรอบตัวและอ่านข้อความแบบออฟไลน์

**Pannana (ไทยพีบีเอส)** เสียงบรรยายภาพสำหรับสื่อวีดิโอ/ภาพยนตร์

**Read for the Blind** คลังหนังสือเสียงจากอาสาสมัคร

**Be My Eyes** วิดีโอคอลเชื่อมต่อผู้พิการทางสายตากับอาสาสมัคร

บทสรุปและกรอบแนวคิดเพื่อการวิจัย (Conclusion & Synthesis)

การบูรณาการองค์ความรู้สู่การพัฒนาจริง:

**AI Engine:** ใช้ MobileNet V3 กับโมเดลล่วงหน้าเพื่อความแม่นยำมากขึ้น

**Interface:** ออกแบบตามหลัก Universal Design เน้นเสียงนำทางและ Contrast สูง

**แอปพลิเคชัน ขาววิจัย**

**Application Layer:** พัฒนาด้วย Flutter มาตรฐาน TTS แบบ Multilingual

**Validation:** ประเมินความแม่นยำด้วย Top-1/Top-k และวัดความพึงพอใจผู้ทรงรอบ SUS

VocabEyes

แอปพลิเคชันเพื่อส่งเสริมการเรียนรู้คำศัพท์สำหรับผู้พิการทางสายตา

บทที่ 3 วิธีการวิจัย

Chapter 3: Research Methodology

ภาพรวมและวัตถุประสงค์ของแอปพลิเคชัน

**ผู้พิการทางสายตา**

**ทำงานบนซอฟต์แวร์ MobileNetV3**

**ฐานข้อมูล 3 ภาษา (ไทย, ญี่ปุ่น, อังกฤษ)**

**ระบบสังเคราะห์เสียง (TTS)**

มุ่งเน้นความเร็ว ความแม่นยำ และส่งเสริมการเรียนรู้ภาษาเบื้องต้นผ่านการรู้จำเสียง (Audio-based learning)

ช่องว่างของงานวิจัย (Research Gap)		
App	ฟีเจอร์ (Function)	ข้อจำกัด/จุดอ่อน (Limitation/Advantage)
Seeing AI (Microsoft) / Lookout (Google)	ฟีเจอร์: ตรวจจับวัตถุและบอกความหมาย	ข้อจำกัด: พึ่งพา Cloud, ระบุภาษาหลักได้เฉพาะภาษาอังกฤษ
Pannana / Read for the Blind	ฟีเจอร์: เป็นระบบอ่านภาพ / คิววีอาร์ทีที (Real-time detection)	ข้อจำกัด: ไม่รองรับการตรวจจับวัตถุแบบเรียลไทม์
Be My Eyes	ฟีเจอร์: บริการอาสาสมัครช่วยเหลือ	ข้อจำกัด: พึ่งพาบุคคล, ควบคุมไม่ได้ด้วยตัวเอง
VocabEyes (Our Proposal)	ฟีเจอร์: เรียนรู้ภาษาพูดแบบเรียลไทม์	จุดเด่น: (รองรับข้อจำกัด Privacy, เสถียร TTS 3 ทางเลือก, ควบคุม UI ที่หลากหลายผ่าน AI) อนาคต: AI ที่ฉลาดยิ่งขึ้น

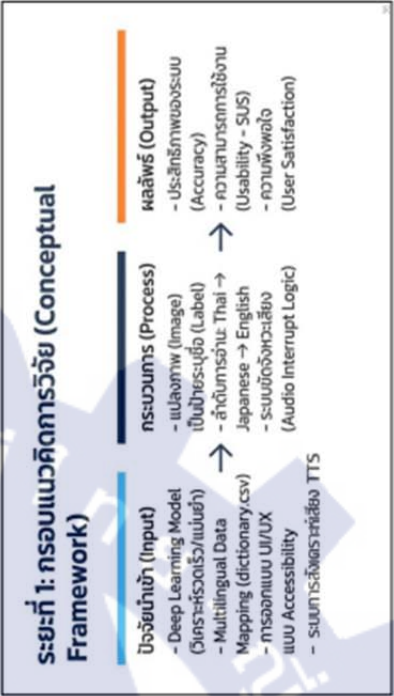
33

ช่องว่างของงานวิจัย (Research Gap)				
ชื่อระบบ	ตรวจสอบวัตถุ	เรียนรู้ภาษาพูด	ออฟไลน์ / ความเป็นส่วนตัว	
Seeing AI / Lookout	✓ มี	✗ ไม่รองรับ	✗ ต้องใช้อินเทอร์เน็ต	
Pannana / Read for the Blind	✗ เป็นเสียงบรรยาย	✗ ไม่รองรับ	✗ ต้องใช้อินเทอร์เน็ต	
Be My Eyes	✗ ใช้อาสาสมัคร	✗ ไม่สนับสนุน	✗ เสี่ยงความเป็นส่วนตัว	
AEYE	✓ มี	✗ ไม่รองรับ	✗ ต้องใช้อินเทอร์เน็ต	
VocabEyes	✓ ตรวจสอบแบบออฟไลน์	✓ เรียนรู้ภาษาพูด 3 ทางเลือก	✓ ความเป็นส่วนตัว (Privacy)	

34



35



36

Functional Requirements (ฟังก์ชันการทำงาน)

- Object Recognition: ตรวจจับวัตถุจากภาพหรือวิดีโอ
- Vocabulary Output: แสดงคำศัพท์ 3 ทักษะ (ฟัง พูด อ่าน)
- Text-to-Speech: สามารถแปลงคำพูดเป็นเสียง

User Features (คุณสมบัติผู้ใช้)

- Full-Screen Gesture: สามารถใช้ท่าทางควบคุมระบบ
- High Contrast UI: ธีมสีตัดกันสูง
- Tutorial Guide: คู่มือการใช้งาน

Non-Functional Requirements

- Offline Capability: สามารถใช้งานได้โดยไม่ต้องเชื่อมต่ออินเทอร์เน็ต

ระบบจะรองรับการใช้งานบนมือถือและแท็บเล็ต

37

ระบบจะรองรับการใช้งานบนมือถือและแท็บเล็ต

Single Tap to Return / Interrupt Audio

Smartphone Input (Camera or Gallery)

Object Recognition Model (MobileNetV2)

Data Mapping (dictionary.csv)

Text-to-Speech (TTS)

Display Results

Object Name + Confidence Score

Hold to Repeat Audio

Single Tap to Return / Interrupt Audio

38

ขั้นตอนการทำงานของระบบ (System Pipeline)

Input: ภาพจากกล้องมือถือ / แอปแกลเลอรี

Process 1: Inference (MobileNetV2)

Process 2: Data Mapping (Dictionary.csv)

Output: Visual (คำศัพท์, ความมั่นใจ)

Output: Audio (เสียงพูด)

Interaction: แอปพลิเคชัน / ผู้ใช้

39

ซอฟต์แวร์ (Software)

- Flutter (Dart) สำหรับพัฒนา UI/UX
- Google Colab (คลาวด์ไฟล์ .html) ใช้สำหรับ Mapping

ฮาร์ดแวร์ (Hardware)

- สมาร์ตโฟน Android
- RAM ≥ 4GB
- เครื่องพัฒนา: Mac M3

ฐานข้อมูล (Database)

- WordNet ID Mapping (dictionary EN-JA)
- ข้อมูลคำศัพท์ประมาณ 6,000-10,000 terms

โมเดล AI (Models)

- MobileNetV3 Large (TFLite Interpreter) - Pretrained on ImageNet ZIK-P (1,221 classes)
- Google TTS (Offline Speech Synthesis)

40

ระยะที่ 3: เงื่อนไขการประเมินความแม่นยำ (Testing Scenarios)

1. สถานการณ์ปกติ (Simple Case - S1)

พื้นหลังเรียบ, วัตถุอยู่กึ่งกลาง, ไม่มีส่วนบัง, แสงปกติ (ใช้เป็นค่า Baseline)

2. สถานการณ์ซับซ้อน (Complex Cases)

C1: วัตถุบังพื้นหลังบางส่วน

C2: พื้นหลังบางส่วน / วัตถุบางส่วน

C3: วัตถุหลายชิ้นในภาพ

C4: วัตถุเล็ก พื้นหลังใหญ่

C5: ไฟล์ / แสงฉาบ

41

ระเบียบวิธีวิจัยและเงื่อนไขการทดสอบ (Evaluation Methodology)

ข้อมูลทดสอบ: วัตถุเป็นภาพ 10 ชนิด x 10 ภาพ x 6 สถานการณ์ = รวม 600 ภาพ

Technical Evaluation (Accuracy)

- S1 (Simple Case): วัตถุเด่นชัด จากหลังเรียบ ไม่มีส่วนบัง
- C1 (Background Clutter): วัตถุเดี่ยว แต่พื้นหลังมีความซับซ้อน
- C2 (Partial Occlusion): วัตถุถูกบังบางส่วน (เต็ม 70-80%)
- C3 (Multiple Objects): วัตถุอยู่ในกรอบมากกว่า 3 ชิ้นขึ้นไป
- C4 (Small Object): วัตถุเล็ก พื้นหลังที่พบบ่อย (ประมาณ 30%)
- C5 (Low-light): ถ่ายในที่แสงน้อย ไฟฉาย

Usability Evaluation (SUS)

เครื่องมือ: System Usability Scale (10 ข้อ) & แบบสอบถามความพึงพอใจ (7 ข้อ)

ผู้ตอบแบบสอบถาม: ผู้วิจัย 3 คน และบุคลากรทางการแพทย์ (N=5)

42

ระยะที่ 3: เป้าหมายและสถิติการประเมิน (Test Targets & Metrics)

Target Data Breakdown

10 ชนิดวัตถุ x 10 ภาพ / 30รูป / สถานการณ์ = รวมทั้งหมด 600 ภาพ

เป้าหมาย: 90% Accuracy, 95% Precision, 95% Recall, 95% F1 Score

Top-1 Accuracy

ส่วนที่ทำคะแนนได้ถูกต้องที่สุดของโมเดล

Top-k Accuracy

ส่วนที่ทำคะแนนได้ถูกต้องปรากฏอยู่ในผลการทำนาย k อันดับแรก เพื่อประเมินความแม่นยำในการจำแนก

43

แอปพลิเคชันต้นแบบ 'VocabEyes' (The Prototype)

หน้าจอแสดงภาพวัตถุ

แอปพลิเคชัน VocabEyes

หน้าจอแสดงผลลัพธ์

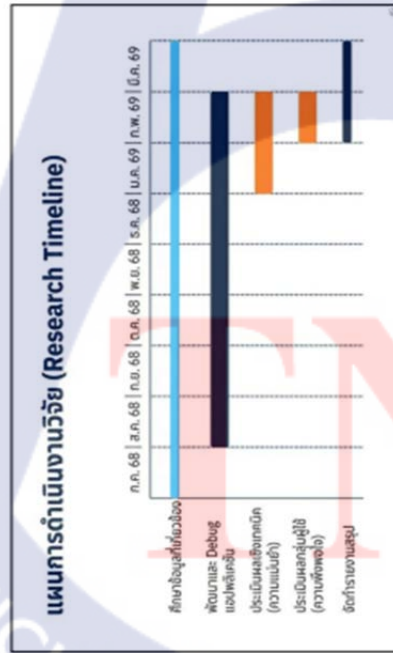
ผลการทำนายจากโมเดล

หน้าจอแสดงคำศัพท์

คำศัพท์ที่เกี่ยวข้องกับภาพ

ทำงานบนแอปไฟล์ 100% (On-device processing) เพื่อความเป็นส่วนตัว

44

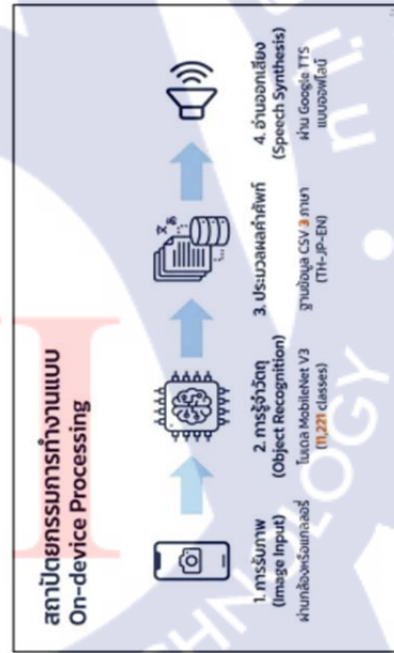


49

บทที่ 4 ผลการวิจัย

การพัฒนาระบบกับ VocabBytes (แอปพลิเคชันสำหรับผู้ใช้และงานเอกสาร 3 ภาษา)

50



51



52



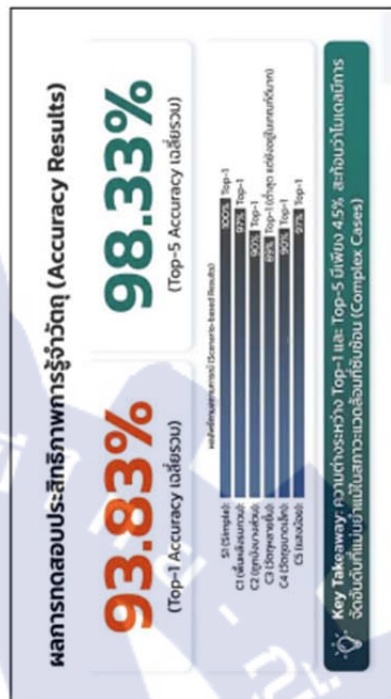
53



54



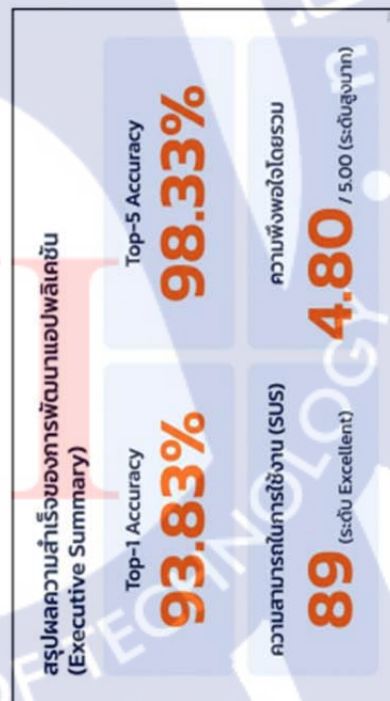
55



95



57



59



58



60

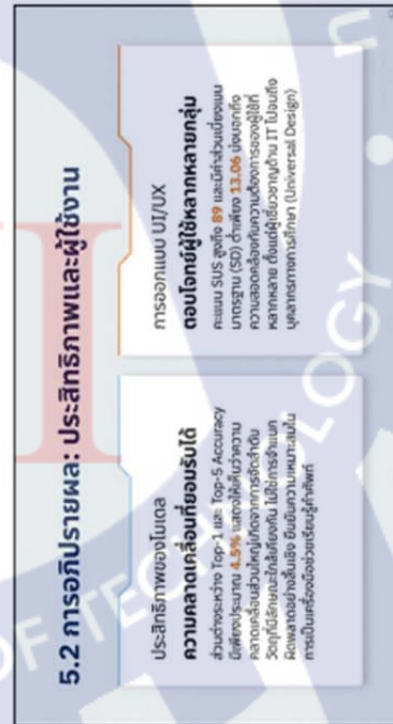
บทที่ 5: บทสรุป การอภิปรายผล และข้อเสนอแนะ

แอปพลิเคชัน VocabEyes (Chapter 5: Conclusions, Discussions, and Recommendations)

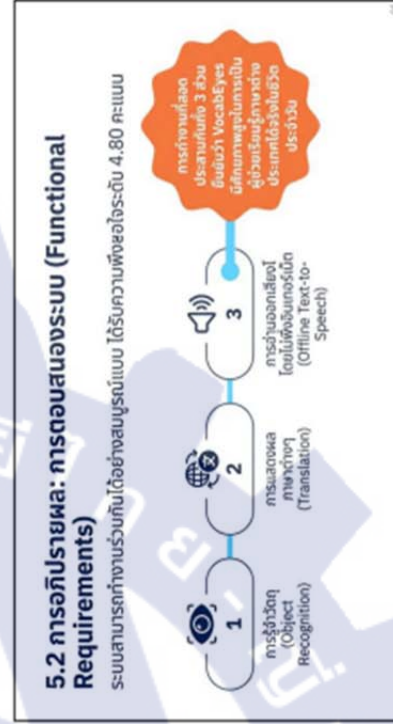
61



62



63



64

5.3 ปัญหา อุปสรรค และข้อจำกัดในการวิจัย

	ข้อจำกัดด้านกลุ่มเป้าหมาย	- การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้ - การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้ - การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้
	ข้อจำกัดด้านวิธีการวิจัย	- การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้ - การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้ - การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้
	ข้อจำกัดด้านข้อมูลเชิงลึก	- การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้ - การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้ - การทดลองใช้ได้ในสถานที่ที่ไม่สามารถเข้าถึงได้

5.4.1 ข้อเสนอแนะในการนำไปใช้งาน (Practical Applications)


ข้อควรระวังในการใช้งาน (Environmental Constraints)

- ความแม่นยำอาจลดลงในสถานการณ์ที่ซับซ้อน เช่น ในบริเวณทางข้าม มีวัตถุหลายอย่างในภาพ หรือวัตถุมีขนาดเล็ก/กลืนไปกับพื้นหลัง


เครื่องมือเสริมสำหรับผู้พิการ (Beyond Visually Impaired Users)

- VocabEyes ไม่ได้จำกัดเฉพาะผู้พิการทางสายตา
- เป็นเครื่องมือที่ช่วยในการเรียนรู้คำศัพท์หรือข้อความต่าง ๆ ที่ต้องการเรียนรู้คำศัพท์ 3 ภาษา (ไทย-ญี่ปุ่น-อังกฤษ) พร้อมกันจากสิ่งรอบตัว

5.4.2 ข้อเสนอแนะสำหรับการวิจัยในอนาคต (Suggestions for Future Research)

- การปรับปรุงประสิทธิภาพของระบบ**
- การปรับปรุงประสิทธิภาพของระบบ**
- การปรับปรุงประสิทธิภาพของระบบ**
- การปรับปรุงประสิทธิภาพของระบบ**
- การปรับปรุงประสิทธิภาพของระบบ**

ข้อจำกัด และทิศทางในอนาคต (Limitations & Future Roadmap)


ข้อจำกัดปัจจุบัน (Limitations)

- Hardware Limits:** การประมวลผลภาพที่ใช้เวลานานเกินไป
- Dataset Bound:** ระบบที่ฝึกด้วยชุดข้อมูล 11,221 ภาพ อาจไม่สามารถทำงานได้ดีกับชุดข้อมูลอื่น
- WordNet Nuance:** ระบบที่ฝึกด้วย WordNet อาจไม่สามารถทำงานได้ดีกับคำที่มีความหมายใกล้เคียง


ทิศทางต่อไป (Future Vision)

- User Expansion:** ขยายการรองรับผู้พิการทางสายตาที่มีความหลากหลาย (Blind & Low Vision)
- Model Fine-Tuning:** ปรับปรุงโมเดล (Fine-tune) เพื่อรองรับชุดข้อมูลที่หลากหลายยิ่งขึ้น
- Global Scaling:** ขยายการใช้งานไปยังภูมิภาคอื่น ๆ เช่น ภาษาอื่น

The Mission
เชื่อมโยงทางการศึกษา
สร้างเครื่องมือการเรียนรู้ที่ฝึก 3
ภาษาผ่านเสียงและข้อความ

The Engine
ใช้ MobileNetV2 ขนาดเล็กและ
แผนที่ค่า ประมวลผลบนอุปกรณ์
(On-device) 100%

The Result
ความแม่นยำ (Top-1) 92.83%
และคะแนนความสามารถในการใช้งาน
(SUS) ระดับ Excellent (89.0)

VocabEyes: เปลี่ยนสมาร์ตโฟนเป็น "ดวงตา" เพื่อการเรียนรู้ภาษา
การพัฒนาแอปพลิเคชันบนมือถือ สำหรับผู้พิการทางสายตา (ไทย-ญี่ปุ่น-อังกฤษ)

บทสรุปสัปดาห์: นวัตกรรมเพื่อการเรียนรู้ที่ยั่งยืน

VocabEyes เป็นมากกว่าแอปพลิเคชัน แต่คือบทพิสูจน์ถึงศักยภาพในการบูรณาการเทคโนโลยี Object Recognition เข้ากับ Natural Language Processing

ช่องว่างทางการศึกษา (Education)
ส่งเสริมการเรียนรู้ด้วยตนเองอย่างมีประสิทธิภาพสำหรับผู้พิการทางสายตา

การออกแบบเพื่อมวลชน (Universal Design)
รองรับการใช้งานของมนุษย์ทุกกลุ่มในสังคมอย่างเท่าเทียมและเข้าถึงได้ (Accessibility)

รากฐานสำคัญสำหรับการต่อยอดงานวิจัยด้านสื่อการสอนอัจฉริยะในอนาคต