การพัฒนาระบบผู้ช่วยจำแนกวัตถุภายในครัวเรือนแบบโต้ตอบเรียลไทม์ ผ่านเว็บแอปพลิเคชันสำหรับผู้พิการทางสายตา โดยประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์

โดย

นายเจษฎากร ม่วงวิจิตร

นายพงศกร เพิ่มคำ

นายฐิติพงษ์ สงกะสิน

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ วิทยาเขต หันตรา

ปีการศึกษา 2568

DEVELOPMENT OF AN INTERACTIVE REAL-TIME HOUSEHOLD OBJECT CLASSIFICATION ASSISTANT SYSTEM VIA WEB APPLICATION FOR VISUALLY IMPAIRED PERSONS USING GENERATIVE ARTIFICIAL INTELLIGENCE

BY

MR. JETSADAKORN MUANGWICHIT

MR. PONGSAKORN PHOEMKHAM

MR. THITIPONG SONGKASIN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE BACHELOR OF SCIENCE DEGREE

IN COMPUTER SCIENCE FACUITY OF SCIENCE AND TECHNOLOGY

RAJAMANGALA UNIVERSITY OF TECHNOLOGY SUVARNABHUMI

HANTRA CAMPUS

ACADEMIC YEAR 2025

หัวข้อปริญญานิพนธ์	การพัฒนาระบบผู้ช่วยจำแนกวัตถุภายในครัวเรือนแบบโต้ตอบเรียลไทม์ผ่านเว็บแอปพลิเคชันสำหรับผู้พิการทางสายตาโดยปัญญาประดิษฐ์เชิงสร้างสรรค์
ชื่อผู้เขียน	เจษฎากร ม่วงวิจิตร พงศกร เพิ่มคำ ฐิติพงษ์ สงกะสิน
ชื่อปริญญา	วิทยาศาสตรบัณฑิต
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ วิทยาเขต หันตรา
อาจารย์ที่ปรึกษาปริญญานิพนธ์	นาย วิเชพ ใจบุญ นาง ทักศิณา คงสมลาภ
ปีการศึกษา	2568

บทคัดย่อ

โครงงานนี้มีวัตถุประสงค์เพื่อพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของเพื่อยกระดับคุณภาพชีวิตของผู้พิการทางสายตาโดยประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์ Artificial Intelligence เป็นแกนหลักในการส่งเสริมความสามารถด้านการดำรงชีวิตประจำวันอย่างอิสระและลดการพึ่งพาผู้อื่น ซึ่งสอดคล้องกับพระราชบัญญัติส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ พ.ศ.2550 และยุทธศาสตร์ชาติ 20 ปี ด้านการสร้างโอกาสและความเสมอภาคทางสังคม

การทำงานของระบบขับเคลื่อนด้วยเทคโนโลยีปัญญาประดิษฐ์ในสองส่วนหลัก โดยมุ่งเน้นการใช้งานเพื่อช่วยเหลือการดำเนินชีวิตประจำวันภายในบ้านหรือที่พักอาศัยเป็นหลัก ส่วนแรกคือ ระบบผู้ช่วยการมองเห็นอัจฉริยะซึ่งรองรับการสั่งงานด้วยเสียงภาษาไทย โดยบูรณาการโมเดล Llama 4 Vision เพื่อทำ AI Visual Assistant หน้าที่บรรยายภาพ อ่านฉลากสินค้า จำแนกรายละเอียดของสิ่งของเครื่องใช้พื้นฐานในชีวิตประจำวัน ตลอดจนตรวจสอบความถูกต้องของวัตถุหลังการหยิบจับเพื่อลดความคลาดเคลื่อน อีกส่วนหนึ่งคือ ระบบตรวจจับและระบุตำแหน่งวัตถุ Object Detection & Localization ที่ทำหน้าที่สแกนหาวัตถุพื้นฐานผ่านกล้องสมาร์ตโฟน พร้อมประมวลผลและให้คำแนะนำด้านทิศทางด้วยเสียงแบบเรียลไทม์ เช่น ซ้าย ขวา บน ล่าง เพื่อช่วยให้ผู้ใช้เล็งกล้องให้ตรงและทราบตำแหน่งของวัตถุได้อย่างถูกต้อง นอกจากนี้ เพื่ออุดช่องโหว่และข้อจำกัดของปัญญาประดิษฐ์ ผู้ดำเนินโครงงานได้พัฒนา

ระบบเครือข่ายอาสาสมัคร Volunteer Network System ขึ้นมาใหม่เป็นการภายใน เพื่อให้ทำงานร่วมกันอย่างบูรณาการ โดยทำหน้าที่เป็นระบบสำรอง Fallback System ในกรณีที่ AI ไม่สามารถจำแนกวัตถุได้อย่างแน่ชัด ระบบจะเปิดให้ผู้พิการทางสายตาสามารถวิดีโอคอลเพื่อเชื่อมต่อกับกลุ่มอาสาสมัครที่ลงทะเบียนไว้ในระบบ เพื่อขอรับความช่วยเหลือและคำแนะนำจากมนุษย์ได้ทันที

การดำเนินโครงงานมุ่งเน้นการพัฒนาระบบทั้งหมดในรูปแบบ เว็บแอปพลิเคชันแบบก้าวหน้า Progressive Web Application PWA บนสมาร์ตโฟน โดยออกแบบให้สามารถทำงานร่วมกับฟังก์ชันการเข้าถึงสำหรับผู้พิการทางสายตา Accessibility Features เช่น TalkBack Android หรือ VoiceOver iOS ได้อย่างสมบูรณ์แบบ เพื่อให้ผู้ใช้สามารถควบคุมระบบและรับฟังข้อมูลผ่านโปรแกรมอ่านหน้าจอ Screen Reader ของอุปกรณ์ได้อย่างราบรื่น จากนั้นได้ดำเนินการทดสอบประสิทธิภาพในสภาพแวดล้อมภายในบ้าน

คำสำคัญ : ผู้พิการทางสายตา,ปัญญาประดิษฐ์,ระบบผู้ช่วยการมองเห็น,การตรวจจับและระบุตำแหน่งวัตถุ,เครือข่ายอาสาสมัคร,เว็บแอปพลิเคชันแบบก้าวหน้า

Thesis Title	DEVELOPMENT OF AN INTERACTIVE REAL-TIME HOUSEHOLD OBJECT CLASSIFICATION ASSISTANT SYSTEM VIA WEB APPLICATION FOR VISUALLY IMPAIRED PERSONS USING GENERATIVE ARTIFICIAL INTELLIGENCE
Author	JETSADAKORN MUANGWICHIT PONGSAKORN PHOEMKHAM THITIPONG SONGKASIN
Degree	Bachelor of Science
Major field/Faculty/University	Computer Science Faculty of Science and Technology Rajamangala University of Technology Suvarnabhumi Hantra campus
Thesis Advisor	นาย นาง นาง
Academic Year	2025

ABSTRACT

This research aims to develop an intelligent assistant system for object classification and localization to enhance the quality of life for visually impaired individuals. By leveraging Generative Artificial Intelligence (AI) as a core technology, the system promotes independent living and reduces dependency, aligning with the Empowerment of Persons with Disabilities Act, B.E. 2550 (2007) and the 20-Year National Strategy regarding social opportunity and equality.

The system's operations are driven by two primary AI components specifically designed for domestic assistance. The first component is an intelligent visual assistant that supports Thai voice commands and integrates the Llama 4 Vision model to function as an AI Visual Assistant; its duties include providing image descriptions, reading product labels,

classifying essential household items, and verifying objects after handling to minimize errors. The second component is an Object Detection and Localization system that scans for items via smartphone cameras and provides real-time audio directional guidance (e.g., left, right, up, down) to assist users in accurate positioning. Furthermore, to address the inherent limitations of AI, the researchers developed an internal Volunteer Network System as a fallback mechanism. In cases where the AI cannot definitively identify an object, the system allows visually impaired users to initiate video calls to registered volunteers for immediate human assistance.

The entire system is developed as a Progressive Web Application (PWA) on smartphones, designed for seamless compatibility with accessibility features such as TalkBack (Android) or VoiceOver (iOS) to ensure smooth interaction through device screen readers. The system's efficiency was subsequently evaluated through performance testing with a sample group of visually impaired participants in a home environment.

Keywords: Visually Impaired, Artificial Intelligence, Visual Assistant System, Object Detection and Localization, Volunteer Network, Progressive Web Application (PWA)

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาของ นายวิเชพ ใจบุญ และ นาง ทักศิณา คงสมลาภ อาจารย์ที่ปรึกษาปริญญานิพนธ์ ที่ได้กรุณาให้คำแนะนำ ตรวจแก้ไข และให้ข้อเสนอแนะอันเป็นประโยชน์ตลอดระยะเวลาการดำเนินงานโครงงาน ผู้ดำเนินโครงงานขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบคุณคณาจารย์และเจ้าหน้าที่สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ทุกท่านที่ให้ความอนุเคราะห์และอำนวยความสะดวกตลอดการศึกษา

สุดท้ายนี้ ขอกราบขอบพระคุณบิดามารดาและครอบครัว ที่ให้การสนับสนุน ให้กำลังใจ และเป็นแรงผลักดันสำคัญในการศึกษาและการจัดทำปริญญานิพนธ์ฉบับนี้จนสำเร็จลุล่วงด้วยดี

คณะผู้จัดทำโครงงาน

ปีการศึกษา 2568

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(3)
กิตติกรรมประกาศ	(5)
สารบัญตาราง	(8)
สารบัญภาพ	(9)
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของโครงงาน	1
1.3 ขอบเขตของโครงงาน	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
1.5 นิยามศัพท์เฉพาะ	3
1.6 กรอบแนวคิดโครงงาน	4
1.7 สมมติฐานของโครงงาน	4
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	5
2.1 แนวคิดเกี่ยวกับความพิการทางสายตาประเภทตาบอดสนิท และเทคโนโลยีสิ่งอำนวยความสะดวก	5
2.2 เทคโนโลยีปัญญาประดิษฐ์และการเรียนรู้เชิงลึก (AI & Deep Learning)	8
2.3 เทคโนโลยีโมเดลภาษาขนาดใหญ่และความสามารถด้านการมองเห็น (Large Multimodal Models)	8

2.4 เทคโนโลยีเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA) 9
2.5 แนวคิดระบบมนุษย์ร่วมในวงรอบ (Human-in-the-Loop) และเครือข่ายอาสาสมัคร 9
2.6 เทคโนโลยีที่ใช้ในการพัฒนาระบบ 9
2.7 งานวิจัยที่เกี่ยวข้อง 11

บทที่ 3 ระเบียบวิธีการดำเนินงาน 13

3.1 รูปแบบโครงงาน 13
3.2 เครื่องมือที่ใช้ในโครงงาน 20
3.3 การออกแบบและพัฒนาระบบ 22
3.4 การเก็บรวบรวมข้อมูล 28
3.5 การวิเคราะห์ข้อมูล 28
3.6 นิยามปฏิบัติการของตัวแปร 30

รายการอ้างอิง 32

สารบัญตาราง

ตารางที่ **หน้า**

1.1 ขอบเขตด้านระยะเวลา ระยะเวลาในการดำเนินการวิจัย 3
2.1 การจำแนกระดับความบกพร่องทางสายตาตามเกณฑ์องค์การอนามัยโลก (WHO) 6
3.1 ความต้องการของระบบแยกตามส่วนการทำงาน 18
3.2 เกณฑ์การแปลผลค่าประสิทธิภาพของระบบ 29

สารบัญภาพ

ภาพที่	หน้า
3.1 การออกแบบและพัฒนาระบบ (หน้าแรก)	13
3.2 การออกแบบและพัฒนาระบบ	14
3.3 การออกแบบและพัฒนาระบบ	14
3.4 การออกแบบและพัฒนาระบบ	15
3.5 การออกแบบและพัฒนาระบบ	16
3.6 กรอบแนวคิดการดำเนินโครงงาน	17
3.7 ภาพรวมสถาปัตยกรรมระบบ	23
3.8 ลำดับการทำงานของระบบ	26

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การมองเห็นเป็นประสาทสัมผัสที่มีบทบาทสำคัญสูงสุดในการรับรู้ข้อมูลข่าวสารและสภาพแวดล้อมรอบตัว สำหรับผู้พิการทางสายตา ข้อจำกัดด้านการมองเห็นส่งผลกระทบโดยตรงต่อความสามารถในการดำเนินชีวิตประจำวัน โดยเฉพาะอย่างยิ่งปัญหาในการระบุชนิด (Classification) และการระบุตำแหน่ง (Localization) ของสิ่งของเครื่องใช้ภายในบ้าน ซึ่งเป็นอุปสรรคสำคัญที่ทำให้ผู้พิการทางสายตาต้องพึ่งพาผู้อื่นในการทำกิจกรรมพื้นฐาน ลดทอนความเป็นอิสระ และอาจนำไปสู่ความเสี่ยงด้านความปลอดภัยหากมีการหยิบจับวัตถุผิดประเภท เช่น การหยิบยารับประทาน หรือสารเคมีในครัวเรือนผิดพลาด

แม้ในปัจจุบันจะมีการพัฒนาเทคโนโลยีสิ่งอำนวยความสะดวก (Assistive Technology) เพื่อช่วยเหลือผู้พิการทางสายตา แต่จากการศึกษาและทบทวนวรรณกรรมพบว่า ระบบที่มีอยู่ส่วนใหญ่ยังคงประสบปัญหาและข้อจำกัดหลายประการ อาทิ การขาดการรองรับภาษาไทยอย่างสมบูรณ์ ความคลาดเคลื่อนในการระบุพิกัดตำแหน่งของวัตถุในระยะใกล้เพื่อการหยิบจับ และข้อจำกัดของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) ที่อาจเกิดความผิดพลาดเมื่อต้องประมวลผลในสภาพแวดล้อมที่มีแสงน้อยหรือวัตถุมีความซับซ้อน นอกจากนี้ แอปพลิเคชันส่วนใหญ่ถูกพัฒนาในรูปแบบแอปพลิเคชันพื้นฐาน (Native Application) ที่มีความซับซ้อนในการติดตั้งและอัปเดต ซึ่งอาจไม่สอดคล้องกับพฤติกรรมการใช้งานผ่านโปรแกรมอ่านหน้าจอ (Screen Reader) ของผู้พิการทางสายตา

จากประเด็นปัญหาดังกล่าว ผู้ดำเนินโครงงานจึงมีแนวคิดในการพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของ โดยมุ่งเน้นการใช้งานภายในที่พักอาศัย ระบบนี้ใช้สถาปัตยกรรมแบบผสมผสาน (Hybrid Architecture) โดยประยุกต์ใช้ เทคโนโลยีการตรวจจับวัตถุ (Object Detection) เพื่อระบุตำแหน่งและทิศทางแบบเรียลไทม์ ร่วมกับ โมเดลภาษาขนาดใหญ่ที่รองรับภาพ (Large Multimodal Model: LMM) เพื่อให้รายละเอียดและบริบทของสิ่งของนั้นๆ และเพื่อเป็นการปิดช่องโหว่ความผิดพลาดของระบบอัตโนมัติ ผู้ดำเนินโครงงานได้พัฒนา "ระบบเครือข่ายอาสาสมัคร (Volunteer Network System)" ขึ้นมาใหม่เป็นการภายใน เพื่อทำหน้าที่เป็นระบบสำรอง (Fallback System) ให้ผู้ใช้สามารถเชื่อมต่อขอความช่วยเหลือจากมนุษย์ได้ทันทีผ่านวิดีโอคอล โดยพัฒนาระบบทั้งหมดให้อยู่ในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA) ที่ออกแบบมาให้ทำงานร่วมกับฟังก์ชันการเข้าถึงบนสมาร์ตโฟนได้อย่างสมบูรณ์แบบ ซึ่งสอดคล้องกับเจตนารมณ์ของพระราชบัญญัติส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ พ.ศ. 2550 และยุทธศาสตร์ชาติ 20 ปี ในด้านการสร้างโอกาสและความเสมอภาคทางสังคม

1.2 วัตถุประสงค์ของโครงงาน

1. เพื่อพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของสำหรับผู้พิการทางสายตา ในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (PWA)

2. เพื่อประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) ร่วมกับระบบเครือข่ายอาสาสมัคร (Volunteer Network System) ในการส่งเสริมความสามารถด้านการดำรงชีวิตประจำวัน

3. เพื่อศึกษาและประเมินประสิทธิภาพของระบบต้นแบบ

1.3 ขอบเขตของโครงงาน

เพื่อให้โครงงานดำเนินไปอย่างมีประสิทธิภาพและบรรลุวัตถุประสงค์ที่ตั้งไว้ ผู้ดำเนินโครงงานได้กำหนดขอบเขตของเนื้อหาไว้ดังนี้

1.3.1 ขอบเขตด้านเนื้อหาและพื้นที่ ระบบถูกออกแบบมาเพื่อช่วยเหลือการดำเนินชีวิตประจำวัน ภายในบ้านหรือที่พักอาศัย (Indoor Environment) เป็นหลัก โดยมุ่งเน้นการตรวจจับ จำแนก และระบุตำแหน่งของสิ่งของเครื่องใช้พื้นฐานในชีวิตประจำวัน เช่น อุปกรณ์รับประทานอาหาร ยาสามัญประจำบ้าน และของใช้ส่วนตัว

1.3.2 ขอบเขตด้านระบบและเทคโนโลยี ผู้ดำเนินโครงงานพัฒนาระบบในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (PWA) บนสมาร์ตโฟน ที่รองรับการทำงานร่วมกับโปรแกรมอ่านหน้าจอ (TalkBack หรือ VoiceOver) โดยมีองค์ประกอบหลัก 3 ส่วน ดังนี้

(1) ระบบผู้ช่วยการมองเห็นอัจฉริยะ (AI Visual Assistant)

พัฒนาโดยบูรณาการโมเดล Llama 4 Vision รองรับการสั่งงานด้วยเสียงภาษาไทย ทำหน้าที่ประมวลผลภาพถ่ายเพื่อบรรยายรายละเอียด อ่านฉลากสินค้า และตอบคำถามเกี่ยวกับวัตถุในภาพ

(2) ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization)

ทำหน้าที่สแกนวัตถุและให้คำแนะนำทิศทาง (ซ้าย ขวา บน ล่าง) ด้วยเสียงแบบเรียลไทม์ เพื่อช่วยในการจัดเฟรมภาพและนำทางมือของผู้ใช้ให้ตรงกับตำแหน่งวัตถุ

(3) ระบบเครือข่ายอาสาสมัคร (Volunteer Network System)

เป็นระบบที่ผู้ดำเนินโครงงานพัฒนาขึ้นเอง เพื่อทำหน้าที่เป็นระบบสำรอง (Fallback System) รองรับการวิดีโอคอลระหว่างผู้พิการทางสายตากับอาสาสมัครในระบบ ในกรณีที่ AI ไม่สามารถให้คำตอบที่ชัดเจนได้

1.3.3 ขอบเขตด้านระยะเวลา ระยะเวลาในการดำเนินการ

ตารางที่ 1.1

ขอบเขตด้านระยะเวลา ระยะเวลาในการดำเนินโครงงาน

ขั้นตอนการดำเนินงาน	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.
1.การวางแผนโครงงาน	↔											
2.ศึกษาและรวบรวมข้อมูล	↔	→										
3.วิเคราะห์ระบบ		↔	→									
4.ออกแบบระบบและ UI			↔	↔	↔							
5.พัฒนาระบบ (Video & AI)				↔	↔	→						
6.ทดสอบและปรับปรุง					↔	↔	↔	↔	→			
7.เผยแพร่									↔	↔	→	
8.นำเสนอโครงงาน												↔

1.4 ประโยชน์ที่คาดว่าจะได้รับ

(1) ผู้พิการทางสายตามีเครื่องมือที่ช่วยเพิ่มความเป็นอิสระ ความมั่นใจ และความปลอดภัยในการจำแนกและหยิบจับสิ่งของภายในบ้าน ลดการพึ่งพาผู้อื่น

(2) ได้ต้นแบบระบบผู้ช่วยอัจฉริยะ (PWA) ที่มีการทำงานแบบผสมผสานระหว่างปัญญาประดิษฐ์และมนุษย์ (Human-in-the-Loop) ซึ่งมีประสิทธิภาพและรองรับการใช้งานภาษาไทย

(3) เป็นแนวทางในการพัฒนาเทคโนโลยีสิ่งอำนวยความสะดวกที่สอดคล้องกับบริบทสังคมไทยและนโยบายการพัฒนาคุณภาพชีวิตคนพิการของประเทศ

1.5 นิยามศัพท์เฉพาะ

เพื่อให้เกิดความเข้าใจที่ตรงกันในการดำเนินโครงงานครั้งนี้ ผู้ดำเนินโครงงานได้กำหนดนิยามศัพท์เฉพาะไว้ดังนี้

1.5.1 ระบบผู้ช่วยอัจฉริยะ หมายถึง เว็บแอปพลิเคชันที่ผู้ดำเนินโครงงานพัฒนาขึ้น เพื่อทำหน้าที่จำแนกสิ่งของและบอกทิศทางตำแหน่งของวัตถุผ่านเสียงสังเคราะห์

1.5.2 ผู้พิการทางสายตา หมายถึง บุคคลที่มีข้อจำกัดในการมองเห็น ซึ่งจำเป็นต้องใช้เทคโนโลยีสิ่งอำนวยความสะดวกในการเข้าถึงข้อมูลและการดำเนินชีวิต

1.5.3 เว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA) หมายถึง เว็บไซต์ที่ได้รับการพัฒนาให้มีคุณสมบัติและการทำงานเสมือนแอปพลิเคชันบนสมาร์ตโฟน สามารถติดตั้งลงบนหน้าจอหลักและใช้งานได้โดยไม่ต้องผ่านร้านค้าแอปพลิเคชัน

1.5.4 ระบบผู้ช่วยการมองเห็นอัจฉริยะ (AI Visual Assistant) หมายถึง ส่วนของโปรแกรมที่ใช้โมเดลปัญญาประดิษฐ์ Llama 4 Vision ในการประมวลผลภาพถ่ายเพื่อระบุรายละเอียดเชิงลึกของวัตถุและอ่านข้อความภาษาไทย

1.5.5 ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization) หมายถึง ส่วนของโปรแกรมที่ทำหน้าที่ตรวจจับวัตถุในภาพแบบทันที (Real-time) และส่งเสียงเตือนเพื่อระบุทิศทางของวัตถุเทียบกับกึ่งกลางของหน้าจอ

1.5.6 ระบบเครือข่ายอาสาสมัคร (Volunteer Network System) หมายถึง ระบบฐานข้อมูลและการสื่อสารที่ผู้ดำเนินโครงงานพัฒนาขึ้น เพื่อเชื่อมโยงระหว่างผู้พิการทางสายตาและอาสาสมัคร สำหรับการขอความช่วยเหลือผ่านวิดีโอคอลในกรณีที่ระบบปัญญาประดิษฐ์ไม่สามารถระบุวัตถุได้

1.6 กรอบแนวคิดโครงงาน

ในการดำเนินโครงงานครั้งนี้ ผู้ดำเนินโครงงานได้กำหนดกรอบแนวคิดในการพัฒนาระบบผู้ช่วยอัจฉริยะ โดยแบ่งออกเป็นส่วนของตัวแปรต้นและตัวแปรตาม ดังนี้

1.6.1 ตัวแปรต้น (Independent Variables): คือ ระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของ (PWA) ซึ่งประกอบด้วย

(1) ระบบผู้ช่วยการมองเห็นอัจฉริยะ (AI Visual Assistant)

(2) ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization)

(3) ระบบเครือข่ายอาสาสมัคร (Volunteer Network System)

1.6.2 ตัวแปรตาม (Dependent Variables): คือ ประสิทธิภาพและผลสัมฤทธิ์จากการใช้งานระบบ ได้แก่ ความแม่นยำในการจำแนกและระบุตำแหน่งสิ่งของ

1.7 สมมติฐานของโครงงาน

การจัดทำโครงงานครั้งนี้ได้กำหนดสมมติฐานไว้ดังนี้

1.7.1 ระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของที่พัฒนาขึ้นในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (PWA) มีความแม่นยำในการจำแนกวัตถุภายในครัวเรือนไม่ต่ำกว่าร้อยละ 80

1.7.2 ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization) สามารถให้คำแนะนำด้านทิศทางแบบเรียลไทม์ได้อย่างถูกต้อง โดยมีค่าความคลาดเคลื่อนของตำแหน่งอยู่ในระดับที่ยอมรับได้

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของเพื่อผู้ที่ตาบอดสนิท (Smart Assistant for Object Classification and Localization for the Blind) ผู้ดำเนินโครงงานได้ศึกษาค้นคว้าแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อเป็นพื้นฐานในการออกแบบและพัฒนาระบบให้มีประสิทธิภาพ โดยแบ่งหัวข้อการนำเสนอออกเป็น 6 ส่วน ดังนี้

2.1 แนวคิดเกี่ยวกับความพิการทางสายตาประเภทตาบอดสนิทและเทคโนโลยีสิ่งอำนวยความสะดวก

2.2 เทคโนโลยีปัญญาประดิษฐ์และการเรียนรู้เชิงลึก (AI & Deep Learning)

2.3 เทคโนโลยีโมเดลภาษาขนาดใหญ่และความสามารถด้านการมองเห็น (Large Multimodal Models)

2.4 เทคโนโลยีเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application)

2.5 แนวคิดระบบมนุษย์ร่วมในวงรอบ (Human-in-the-Loop) และเครือข่ายอาสาสมัคร

2.6 งานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดเกี่ยวกับความพิการทางสายตาประเภทตาบอดสนิทและเทคโนโลยีสิ่งอำนวยความสะดวก

2.1.1 นิยามและการจำแนกประเภทความพิการ

องค์การอนามัยโลก (World Health Organization: WHO) และพระราชบัญญัติส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ พ.ศ. 2550 ได้กำหนดนิยามของ "ผู้ที่ตาบอดสนิท (Blind)" หมายถึง บุคคลที่สูญเสียการมองเห็นโดยสิ้นเชิง หรือมีการมองเห็นน้อยมากจนไม่สามารถใช้สายตาในการรับรู้ภาพหรือแสงเพื่อการดำเนินชีวิตประจำวันได้ (No Light Perception)

โดยตามเกณฑ์การจำแนกขององค์การอนามัยโลก ซึ่งอ้างอิงระบบการจำแนกโรคระหว่างประเทศ (International Classification of Diseases: ICD) ภาวะตาบอด (Blindness) หมายถึง บุคคลที่มีค่าความชัดของการมองเห็น (Visual Acuity) ในตาที่ดีกว่าหลังการแก้ไขที่ดีที่สุด ต่ำกว่า 3/60 หรือ มีความบกพร่องของลานสายตา (Visual Field) แคบกว่า 10 องศาจากจุดศูนย์กลางการมองเห็น ในกรณีที่รุนแรงที่สุด ผู้ป่วยอาจอยู่ในภาวะไม่รับรู้แสง (No Light Perception: NLP) ซึ่งจัดเป็นระดับตาบอดสนิทอย่างสมบูรณ์ (WHO, 2019) การกำหนดเกณฑ์เชิงปริมาณดังกล่าวมีความสำคัญต่อการคัดเลือกกลุ่มตัวอย่างทางวิชาการ และช่วยให้การจำแนกระดับความพิการเป็นไปตามมาตรฐานสากล เพื่อให้เห็นภาพรวมของการจำแนกระดับความบกพร่องทางสายตาอย่างเป็นระบบ สามารถสรุปได้ดังตารางต่อไปนี้

ตารางที่ 2.1

การจำแนกระดับความบกพร่องทางสายตาตามเกณฑ์องค์การอนามัยโลก (WHO)

ระดับความบกพร่อง	ค่า Visual Acuity (ตาที่ดีกว่าหลังแก้ไขแล้ว)	ลานสายตา (Visual Field)	คำอธิบายลักษณะการมองเห็น
การมองเห็นปกติ (Normal Vision)	≥ 6/18	ปกติ	มองเห็นรายละเอียดได้ตามปกติ
ความบกพร่องระดับเล็กน้อย (Mild Visual Impairment)	< 6/18 ถึง ≥ 6/60	อาจเริ่มแคบลง	มองเห็นไม่ชัด ต้องเพ่งหรือใช้อุปกรณ์ช่วย
ความบกพร่องระดับรุนแรง (Severe Visual Impairment)	< 6/60 ถึง ≥ 3/60	อาจแคบกว่า 20°	มองเห็นเพียงวัตถุขนาดใหญ่หรือใกล้มาก
ตาบอด (Blindness)	< 3/60	< 10°	ไม่สามารถใช้สายตาในการดำเนินชีวิตประจำวันได้
ตาบอดสนิท (Total Blindness / No Light Perception: NLP)	ไม่สามารถวัดค่าได้	ไม่มีการรับรู้ลานสายตา	ไม่สามารถรับรู้แสงได้เลย

จากตารางข้างต้น สำหรับโครงงานนี้ มุ่งเน้นการแก้ปัญหาให้แก่กลุ่มผู้ที่ตาบอดสนิทโดยเฉพาะ ซึ่งเผชิญอุปสรรคขั้นวิกฤต (Critical Barriers) ในการรับรู้สภาพแวดล้อมภายในอาคาร (Indoor Perception) การขาดข้อมูลทางสายตาอย่างสิ้นเชิงส่งผลให้ไม่สามารถระบุชนิด (Classification) และตำแหน่ง (Localization) ของวัตถุได้ด้วยตนเอง จำเป็นต้องพึ่งพาประสาทสัมผัสอื่น เช่น การได้ยินและการสัมผัส หรือความช่วยเหลือจากผู้อื่นในการทำกิจวัตรพื้นฐาน เช่น การหยิบยารับประทาน การค้นหาของใช้ส่วนตัว หรือการหลีกเลี่ยงสิ่งกีดขวาง ซึ่งสะท้อนให้เห็นถึงความจำเป็นของเทคโนโลยีสิ่งอำนวยความสะดวกที่สามารถทดแทนข้อมูลทางสายตาได้อย่างแม่นยำและปลอดภัย

2.1.2 เทคโนโลยีสิ่งอำนวยความสะดวก (Assistive Technology)

เทคโนโลยีสำหรับผู้ที่ตาบอดสนิทจำเป็นต้องเปลี่ยนรูปแบบข้อมูลภาพ (Visual Information) ให้เป็นข้อมูลเสียง (Auditory Information) หรือข้อมูลสัมผัส (Tactile Information) อย่างสมบูรณ์ ปัจจุบันสมาร์ตโฟนได้กลายเป็นเครื่องมือหลัก (Primary Aid) เนื่องจากมีระบบปฏิบัติการที่รองรับการใช้งานแบบไร้การมองเห็น (Eyes-free Operation) ผ่านฟังก์ชันการเข้าถึง (Accessibility Features) ได้แก่ TalkBack (Android) และ VoiceOver (iOS) ซึ่งทำหน้าที่อ่านออกเสียงทุกองค์ประกอบบนหน้าจอ (Screen Reader) ช่วยให้ผู้ที่ตาบอดสนิทสามารถโต้ตอบกับแอปพลิเคชันได้โดยไม่ต้องมองเห็นหน้าจอ (Google, 2023; Apple, 2023)

2.1.3 ยุทธศาสตร์ชาติ 20 ปี กับการพัฒนาคุณภาพชีวิตคนพิการ

ยุทธศาสตร์ชาติ 20 ปี (พ.ศ. 2561–2580) ซึ่งประกาศใช้ตามพระราชบัญญัติการจัดทำยุทธศาสตร์ชาติ พ.ศ. 2560 ถือเป็นแผนแม่บทสูงสุดในการพัฒนาประเทศ ประกอบด้วย 6 ด้านหลัก โดยยุทธศาสตร์ที่เกี่ยวข้องโดยตรงกับโครงงานนี้คือ ยุทธศาสตร์ด้านที่ 4 การสร้างโอกาสและความเสมอภาคทางสังคม ซึ่งมีเป้าหมายสำคัญในการลดความเหลื่อมล้ำและสร้างความเป็นธรรมในสังคม โดยเฉพาะการเพิ่มโอกาสในการเข้าถึงบริการและเทคโนโลยีสำหรับกลุ่ม

ผู้ด้อยโอกาส ผู้พิการ และผู้สูงอายุ ให้สามารถพึ่งพาตนเองได้และมีส่วนร่วมในสังคมอย่างเต็มศักยภาพ (สำนักงานคณะกรรมการฯ, 2561)

ยุทธศาสตร์ดังกล่าวได้กำหนดแนวทางการพัฒนาที่สอดคล้องกับโครงงานนี้ใน 2 ประเด็นหลัก ได้แก่

ประเด็นแรกคือ การเสริมสร้างพลังทางสังคม มุ่งเน้นให้คนพิการมีความสามารถในการดำรงชีวิตอย่างอิสระ ลดการพึ่งพาผู้อื่นในชีวิตประจำวัน และสามารถเข้าถึงข้อมูลและบริการต่างๆ ได้อย่างเท่าเทียมกับคนทั่วไป

ประเด็นที่สองคือ การใช้เทคโนโลยีเพื่อลดความเหลื่อมล้ำ ส่งเสริมการนำนวัตกรรมและเทคโนโลยีดิจิทัลมาประยุกต์ใช้เพื่อยกระดับคุณภาพชีวิตของกลุ่มเปราะบางในสังคม รวมถึงการพัฒนาเทคโนโลยีสิ่งอำนวยความสะดวกที่สอดคล้องกับบริบทและภาษาของสังคมไทย โครงงานนี้จึงถือเป็นการตอบสนองต่อยุทธศาสตร์ชาติดังกล่าวอย่างเป็นรูปธรรม โดยการพัฒนาระบบผู้ช่วยอัจฉริยะที่ช่วยให้ผู้พิการทางสายตาสามารถดำเนินชีวิตประจำวันภายในบ้านได้อย่างอิสระและปลอดภัยมากขึ้น

2.1.4 แนวคิดการอยู่ร่วมกันอย่างเท่าเทียมและครอบคลุม (Inclusion and Equity)

แนวคิดการอยู่ร่วมกันอย่างเท่าเทียมและครอบคลุม (Social Inclusion) เป็นแนวคิดที่ได้รับการยอมรับในระดับสากล โดยองค์การสหประชาชาติ (United Nations) ได้บัญญัติไว้ในอนุสัญญาว่าด้วยสิทธิของคนพิการ (Convention on the Rights of Persons with Disabilities: CRPD) (United Nations, 2006) ซึ่งประเทศไทยได้ให้สัตยาบันเมื่อปี พ.ศ. 2551 โดยมีหลักการสำคัญ 3 ประการที่เกี่ยวข้องกับโครงงานนี้ดังนี้

หลักการที่ 1 การเคารพศักดิ์ศรีและความเป็นอิสระ (Respect for Dignity and Autonomy) กำหนดว่าคนพิการมีสิทธิในการตัดสินใจด้วยตนเองและดำรงชีวิตอย่างอิสระ โดยไม่จำเป็นต้องพึ่งพาผู้อื่นในกิจกรรมพื้นฐาน ซึ่งเป็นแรงจูงใจหลักในการพัฒนาระบบผู้ช่วยอัจฉริยะในโครงงานนี้ เพื่อให้ผู้พิการทางสายตาสามารถหยิบจับและจำแนกสิ่งของภายในบ้านได้ด้วยตนเอง

หลักการที่ 2 การเข้าถึงได้อย่างเท่าเทียม (Accessibility) กำหนดให้รัฐและภาคส่วนต่างๆ มีหน้าที่จัดให้มีเทคโนโลยีและสภาพแวดล้อมที่คนพิการสามารถเข้าถึงได้อย่างเท่าเทียม ซึ่งสอดคล้องกับการพัฒนาระบบในรูปแบบ PWA ที่ทำงานร่วมกับ Screen Reader และรองรับภาษาไทยอย่างสมบูรณ์

หลักการที่ 3 การมีส่วนร่วมในสังคม (Full Participation) ส่งเสริมให้คนพิการสามารถมีส่วนร่วมในกิจกรรมทางสังคมได้อย่างเต็มที่ โดยไม่ถูกกีดกันหรือเลือกปฏิบัติ การพัฒนาเทคโนโลยีสิ่งอำนวยความสะดวกจึงเป็นกลไกสำคัญในการลดอุปสรรคและเพิ่มโอกาสการมีส่วนร่วมของผู้พิการทางสายตาในสังคม

2.1.5 สถานการณ์ผู้พิการทางสายตาในประเทศไทย

จากข้อมูลของกรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ พบว่าในปี พ.ศ. 2566 ประเทศไทยมีผู้พิการที่ได้รับการจดทะเบียนทั้งสิ้นกว่า 2 ล้านคน โดยผู้พิการทางการมองเห็นเป็นกลุ่มที่มีจำนวนมากเป็นอันดับต้นๆ คิดเป็นสัดส่วนประมาณร้อยละ 12 ของผู้พิการทั้งหมด ซึ่งส่วนใหญ่ยังขาดการเข้าถึงเทคโนโลยีสิ่งอำนวยความสะดวกที่รองรับภาษาไทยอย่างแท้จริง สะท้อนให้เห็นถึงความจำเป็นเร่งด่วนในการพัฒนาเทคโนโลยีที่ตอบสนองต่อความต้องการของกลุ่มนี้โดยเฉพาะ (กรมส่งเสริมฯ, 2566)

2.2 เทคโนโลยีปัญญาประดิษฐ์และการเรียนรู้เชิงลึก (AI & Deep Learning)

2.2.1 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs)

โครงงานนี้ประยุกต์ใช้เทคโนโลยี Deep Learning ประเภท CNNs ซึ่งมีความสามารถในการทำหน้าที่แทน "ดวงตา" ของมนุษย์ โดยโมเดลจะทำการสกัดคุณลักษณะ (Feature Extraction) จากภาพที่รับผ่านกล้องสมาร์ตโฟน เพื่อแปลงเป็นข้อมูลที่ระบบคอมพิวเตอร์สามารถเข้าใจและนำไปประมวลผลต่อได้ (LeCun et al., 2015)

2.2.2 การตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization)

สำหรับผู้ที่ตาบอดสนิท การรู้เพียงว่า "มีอะไรอยู่ตรงหน้า" (Classification) นั้นไม่เพียงพอ แต่จำเป็นต้องทราบ "ตำแหน่งที่แน่ชัด" (Localization) เพื่อให้สามารถยื่นมือไปสัมผัสหรือหยิบจับวัตถุได้ เทคโนโลยี Object Detection จึงเข้ามามีบทบาทสำคัญในการทำหน้าที่ 2 ประการ

(1) ระบุชนิดวัตถุ (Identification): บอกชื่อวัตถุเพื่อให้ผู้ใช้มั่นใจว่าเป็นสิ่งที่ต้องการหาจริง

(2) ระบุพิกัด (Localization): สร้างกรอบสี่เหลี่ยม (Bounding Box) ล้อมรอบวัตถุ ซึ่งโครงงานนี้จะนำพิกัดดังกล่าวมาคำนวณเพื่อสร้าง ระบบนำทางมือ (Hand Guidance System) โดยส่งเสียงแจ้งเตือนทิศทาง (ซ้าย, ขวา, บน, ล่าง) แบบเรียลไทม์ ช่วยให้ผู้ที่ตาบอดสนิทสามารถเข้าถึงวัตถุเป้าหมายได้อย่างแม่นยำแม้จะมองไม่เห็นเลยก็ตาม

2.3 เทคโนโลยีโมเดลภาษาขนาดใหญ่และความสามารถด้านการมองเห็น (Large Multimodal Models)

2.3.1 ความสำคัญของ LMMs ต่อผู้ที่ตาบอดสนิท

โมเดลภาษาขนาดใหญ่แบบพหุรูปแบบ (Large Multimodal Models: LMMs) เข้ามาแก้ปัญหาข้อจำกัดของระบบ AI แบบดั้งเดิมที่มักให้คำตอบสั้นห้วน (เช่น "ขวดน้ำ") ซึ่งไม่เพียงพอสำหรับผู้ที่ตาบอดสนิทที่ต้องการรายละเอียดบริบท (Context) เพิ่มเติมเพื่อความปลอดภัย เช่น "ขวดน้ำพลาสติกใส ฝาสีแดง วางอยู่ริมโต๊ะ" ความสามารถในการบรรยายภาพอย่างละเอียด (Dense Captioning) และการอ่านข้อความในภาพ (OCR) ของ LMMs จึงเปรียบเสมือนการสร้างภาพจำลองในจินตนาการให้แก่ผู้ใช้งาน (Meta AI, 2025)

2.3.2 โมเดล Llama 4 Vision

โครงงานนี้เลือกใช้ Llama 4 Vision เนื่องจากมีความสามารถโดดเด่นในการทำความเข้าใจบริบทภาพและรองรับภาษาไทย ซึ่งเป็นสิ่งจำเป็นอย่างยิ่งสำหรับการอ่านฉลากยา หรือเอกสารกำกับสินค้าที่ผู้ที่ตาบอดสนิทไม่สามารถเข้าถึงได้ด้วยวิธีการอื่น การใช้ Llama 4 Vision จึงช่วยลดความเสี่ยงในการหยิบจับสิ่งของผิดประเภทที่เป็นอันตราย (Meta AI, 2025)

2.4 เทคโนโลยีเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA)

2.4.1 การเข้าถึงที่ไร้รอยต่อ (Seamless Accessibility)

Progressive Web Application (PWA) เป็นเทคโนโลยีที่เหมาะสมอย่างยิ่งสำหรับผู้ที่ตาบอดสนิท เนื่องจากช่วยลดความซับซ้อนในการติดตั้งแอปพลิเคชัน (Installation Friction) ผ่าน App Store ซึ่งมักมีขั้นตอนที่ยุ่งยากและอาจมีปุ่มที่ไม่รองรับ Screen Reader การใช้ PWA อนุญาตให้ผู้ใช้เข้าถึงผ่านลิงก์และติดตั้งลงหน้าจอหลักได้ทันที ทำให้กระบวนการเริ่มต้นใช้งาน (Onboarding) ง่ายและรวดเร็ว

2.4.2 การทำงานร่วมกับ Screen Reader PWA ที่พัฒนาตามมาตรฐานเว็บ (Web Standards) สามารถทำงานร่วมกับซอฟต์แวร์อ่านหน้าจอ (TalkBack/VoiceOver) ได้อย่างมีประสิทธิภาพ โดยใช้ Semantic HTML และ ARIA Labels ที่ถูกต้อง ช่วยให้ผู้ที่ตาบอดสนิทสามารถนำทางในแอปพลิเคชันได้โดยไม่หลงทาง (Roumeliotis & Tselikas, 2022)

2.5 แนวคิดระบบมนุษย์ร่วมในวงรอบ (Human-in-the-Loop) และเครือข่ายอาสาสมัคร

2.5.1 ความจำเป็นของระบบสำรองสำหรับผู้ที่ตาบอดสนิท

เนื่องจากผู้ที่ตาบอดสนิทไม่สามารถตรวจสอบความถูกต้องของผลลัพธ์จาก AI ได้ด้วยสายตาตนเอง (No Visual Verification) หาก AI เกิดความผิดพลาด (Hallucination) อาจนำไปสู่อันตรายร้ายแรงได้ แนวคิด Human-in-the-Loop (HITL) (Wu et al., 2022) จึงมีความจำเป็นสูงสุดสำหรับกลุ่มเป้าหมายนี้

2.5.2 บทบาทของเครือข่ายอาสาสมัคร

โครงงานนี้ได้บูรณาการ "ระบบเครือข่ายอาสาสมัคร (Volunteer Network)" เพื่อทำหน้าที่เป็น "ดวงตามนุษย์" ในยามจำเป็น ผ่านระบบวิดีโอคอล (Video Call) การมีมนุษย์คอยตรวจสอบและให้คำแนะนำในสถานการณ์ที่ซับซ้อน ช่วยสร้างความมั่นใจ (Trust) และความปลอดภัย (Safety) ในระดับที่ AI เพียงอย่างเดียวยังไม่สามารถทำได้

2.6 เทคโนโลยีที่ใช้ในการพัฒนาระบบ

การพัฒนาระบบผู้ช่วยอัจฉริยะในโครงงานนี้ได้ประยุกต์ใช้เทคโนโลยีและเครื่องมือหลายประเภทร่วมกัน โดยสามารถจำแนกออกเป็น 3 กลุ่มหลัก ได้แก่ เทคโนโลยีสำหรับการประมวลผลภาพบนอุปกรณ์ เทคโนโลยีสำหรับการสื่อสารแบบเรียลไทม์ และเทคโนโลยีสำหรับการโต้ตอบด้วยเสียงและแพลตฟอร์มการพัฒนา

2.6.1 TensorFlow.js และโมเดล COCO-SSD

TensorFlow.js เป็นไลบรารีโอเพนซอร์สที่พัฒนาโดย Google ช่วยให้สามารถนำโมเดลการเรียนรู้เชิงลึก (Deep Learning) มาทำงานบนเบราว์เซอร์ผ่านภาษา JavaScript ได้โดยตรง โดยไม่ต้องพึ่งพาเซิร์ฟเวอร์ภายนอก ส่งผลให้ข้อมูลภาพถูกประมวลผลบนอุปกรณ์โดยตรง (On-device Processing) ซึ่งช่วยลดความหน่วงและรักษาความเป็นส่วนตัวของผู้ใช้

โมเดล COCO-SSD (Common Objects in Context - Single Shot MultiBox Detector) เป็นโมเดลตรวจจับวัตถุที่ออกแบบมาสำหรับอุปกรณ์ที่มีทรัพยากรจำกัด สามารถตรวจจับและจำแนกวัตถุในชีวิตประจำวันได้ 80

ประเภท โดยให้ผลลัพธ์เป็นพิกัดกรอบสี่เหลี่ยม (Bounding Box) พร้อมค่าความเชื่อมั่น (Confidence Score) แบบเรียลไทม์ (Smilkov et al., 2019)

ในโครงงานนี้ TensorFlow.js และโมเดล COCO-SSD ถูกนำมาใช้ในระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization) โดยรับภาพวิดีโอจากกล้องหลังของสมาร์ตโฟนผ่าน getUserMedia API และประมวลผลบนเบราว์เซอร์โดยตรงในอัตรา 10 เฟรมต่อวินาที เมื่อตรวจพบวัตถุแล้ว ระบบจะคำนวณพิกัดกึ่งกลางของวัตถุเทียบกับจุดกึ่งกลางหน้าจอด้วยสูตร Euclidean Distance และแปลงเป็นคำแนะนำทิศทาง ได้แก่ ซ้าย ขวา บน หรือล่าง ผ่านเสียงสังเคราะห์แบบเรียลไทม์ โดยกำหนด Tolerance Zone ที่ 20% ของจุดกึ่งกลางหน้าจอเป็นเกณฑ์แจ้งเตือนว่าวัตถุอยู่ในตำแหน่งที่เหมาะสมแล้ว

2.6.2 Groq API

Groq API เป็นบริการประมวลผลโมเดลภาษาขนาดใหญ่ผ่าน RESTful API โดยใช้สถาปัตยกรรมฮาร์ดแวร์เฉพาะทางที่เรียกว่า Language Processing Unit (LPU) ซึ่งออกแบบมาเพื่อลดความหน่วงในการประมวลผลโมเดลภาษาขนาดใหญ่ได้อย่างมีประสิทธิภาพ ทำให้เหมาะสมกับการใช้งานที่ต้องการการตอบสนองแบบเรียลไทม์ โดยในโครงงานนี้ใช้เป็นช่องทางในการเรียกใช้โมเดล Llama 4 Vision สำหรับการวิเคราะห์ภาพเชิงลึก (Groq, Inc., 2024)

ในโครงงานนี้ Groq API ถูกนำมาใช้เป็นช่องทางหลักในการเรียกใช้โมเดล Llama 4 Vision สำหรับระบบผู้ช่วยการมองเห็นอัจฉริยะ (AI Visual Assistant) โดยเมื่อผู้ใช้สั่งงานด้วยเสียง ระบบจะแปลงภาพเป็นข้อมูล Base64 รูปแบบ JPEG และส่งพร้อมกับคำถามและประวัติการสนทนาสูงสุด 6 ข้อความล่าสุดไปยัง Groq API เพื่อวิเคราะห์และสร้างคำตอบเป็นภาษาไทยแบบเรียลไทม์

2.6.3 WebRTC และ PeerJS

Web Real-Time Communication (WebRTC) เป็นมาตรฐานเทคโนโลยีเว็บแบบโอเพนซอร์สที่รองรับการสื่อสารด้วยเสียงและวิดีโอแบบ Peer-to-Peer โดยตรงระหว่างเบราว์เซอร์ โดยไม่ต้องผ่านเซิร์ฟเวอร์กลาง ช่วยลดความหน่วงและภาระของระบบ WebRTC ได้รับการรองรับโดยเบราว์เซอร์หลักทั้งหมด เช่น Chrome, Firefox และ Safari ทำให้สามารถใช้งานข้ามแพลตฟอร์มได้โดยไม่ต้องติดตั้งซอฟต์แวร์เพิ่มเติม (Jennings et al., 2021)

PeerJS เป็นไลบรารีที่ห่อหุ้ม (Wrapper) WebRTC API ให้ใช้งานได้ง่ายขึ้น โดยทำหน้าที่จัดการกระบวนการ Signaling และการสร้างการเชื่อมต่อ Peer-to-Peer ให้โดยอัตโนมัติ ลดความซับซ้อนในการพัฒนาระบบวิดีโอคอลระหว่างผู้พิการทางสายตากับอาสาสมัคร (PeerJS, 2024)

ในโครงงานนี้ WebRTC และ PeerJS ถูกนำมาใช้ในระบบเครือข่ายอาสาสมัคร (Volunteer Network System) สำหรับสถาปนาการเชื่อมต่อวิดีโอคอลแบบ Peer-to-Peer ระหว่างผู้พิการทางสายตาและอาสาสมัครโดยตรง เมื่ออาสาสมัครตอบรับคำร้องขอ ระบบจะเชื่อมต่อทันทีโดยใช้ PeerJS เป็นตัวกลางจัดการ Signaling ทำให้อาสาสมัครสามารถรับชมภาพสดจากกล้องของผู้ใช้ สื่อสารด้วยเสียง และควบคุมอุปกรณ์ระยะไกล เช่น การเปิดไฟแฟลช ได้แบบเรียลไทม์

2.6.4 Pusher

Pusher เป็นบริการ WebSocket แบบ Hosted ที่ช่วยให้สามารถรับส่งข้อมูลแบบเรียลไทม์ระหว่างเซิร์ฟเวอร์และไคลเอนต์ได้อย่างรวดเร็ว ในระบบนี้ Pusher ทำหน้าที่เป็น Signaling Layer สำหรับการตรวจสอบสถานะออนไลน์ของอาสาสมัคร (Presence Channel) และการส่งคำร้องขอความช่วยเหลือไปยังอาสาสมัครที่พร้อมให้บริการ ก่อนที่ระบบจะสถาปนาการเชื่อมต่อวิดีโอคอลผ่าน WebRTC (Pusher Ltd., 2024)

ในงโครงงานนี้ Pusher ถูกนำมาใช้เป็น Signaling Layer ของระบบเครือข่ายอาสาสมัคร โดยทำหน้าที่ตรวจสอบและแสดงสถานะออนไลน์ของอาสาสมัครแบบเรียลไทม์ผ่าน Presence Channel และส่งสัญญาณแจ้งเตือนไปยังอาสาสมัครที่พร้อมให้บริการทันทีเมื่อผู้พิการทางสายตากดขอความช่วยเหลือ ก่อนที่ระบบจะส่งต่อให้ WebRTC สถาปนาการเชื่อมต่อวิดีโอคอลในขั้นตอนถัดไป

2.6.5 Web Speech API

Web Speech API เป็น API มาตรฐานของเว็บเบราว์เซอร์ที่รองรับทั้งการแปลงเสียงพูดเป็นข้อความ (Speech Recognition) และการแปลงข้อความเป็นเสียงพูด (Speech Synthesis) โดยประมวลผลบนเบราว์เซอร์ได้โดยตรง ในโครงงานนี้ใช้สำหรับรับคำสั่งเสียงภาษาไทยจากผู้ใช้ และแปลงผลลัพธ์จากโมเดล AI ให้กลับมาเป็นเสียงพูดที่ผู้พิการทางสายตาสามารถรับฟังได้ทันที (Shires & Wennerstrom, 2023)

ในโครงงานนี้ Web Speech API ถูกนำมาใช้ใน 2 ส่วนหลัก ส่วนแรกคือการรับคำสั่งเสียงภาษาไทยจากผู้ใช้ (Speech Recognition) เพื่อส่งต่อให้ Groq API ประมวลผลร่วมกับภาพ และส่วนที่สองคือการแปลงผลลัพธ์ที่ได้จากโมเดล Llama 4 Vision ให้กลับมาเป็นเสียงพูดภาษาไทย (Speech Synthesis) เพื่อให้ผู้พิการทางสายตาสามารถรับฟังข้อมูลได้ทันทีโดยไม่ต้องมองหน้าจอ

2.6.6 Next.js

Next.js เป็น React Framework ที่รองรับการ Render ทั้งแบบ Server-Side Rendering (SSR) และ Static Site Generation (SSG) ช่วยให้เว็บแอปพลิเคชันโหลดได้รวดเร็วและสามารถกำหนดค่าให้ทำงานในรูปแบบ Progressive Web Application (PWA) ได้อย่างสมบูรณ์ โดยการเพิ่ม Service Worker สำหรับการทำงานแบบ Offline บางส่วน และ Web App Manifest สำหรับการติดตั้งบนหน้าจอหลักของสมาร์ตโฟน (Vercel, 2024)

ในโครงงานนี้ Next.js ถูกนำมาใช้เป็นแพลตฟอร์มหลักในการพัฒนาเว็บแอปพลิเคชันแบบก้าวหน้า (PWA) โดยใช้ Server-Side Rendering (SSR) เพื่อให้ระบบโหลดได้รวดเร็วและรองรับการทำงานร่วมกับโปรแกรมอ่านหน้าจอ TalkBack และ VoiceOver ได้อย่างสมบูรณ์ นอกจากนี้ยังกำหนดค่า Service Worker และ Web App Manifest เพื่อให้ผู้ใช้สามารถติดตั้งแอปพลิเคชันบนหน้าจอหลักของสมาร์ตโฟนได้โดยไม่ต้องผ่านร้านค้าแอปพลิเคชัน

2.7 งานวิจัยที่เกี่ยวข้อง

จากการทบทวนวรรณกรรม ผู้ดำเนินโครงงานได้จำแนกงานวิจัยที่เกี่ยวข้องกับเทคโนโลยีสำหรับผู้พิการทางสายตาออกเป็น 3 กลุ่ม ดังนี้

2.7.1 กลุ่มงานวิจัยด้านระบบตรวจจับวัตถุและนำทาง

Kuriakose และ Shajahan (2020) พัฒนาระบบนำทางด้วยเสียงสำหรับผู้พิการทางสายตาโดยใช้ Computer Vision ร่วมกับระบบเสียง 3 มิติ (3D Audio Feedback) ผลการวิจัยพบว่าการให้ข้อมูลเสียงแบบเรียลไทม์ช่วยเพิ่มความเร็วในการค้นหาวัตถุและการนำทางได้อย่างมีนัยสำคัญ แต่ยังพบข้อจำกัดเรื่องความแม่นยำในสภาพแวดล้อมที่มีแสงน้อย และความซับซ้อนในการติดตั้งอุปกรณ์

Kumar และคณะ (2022) พัฒนาระบบตรวจจับวัตถุสำหรับผู้พิการทางสายตาผ่านสมาร์ตโฟน โดยใช้กล้องเป็นอุปกรณ์นำเข้าข้อมูลและแจ้งผลผ่านระบบเสียง ระบบสามารถช่วยระบุและหลีกเลี่ยงวัตถุทั้งในสภาพแวดล้อมภายใน

และภายนอกอาคาร อย่างไรก็ตาม ระบบดังกล่าวยังไม่รองรับการประมวลผลแบบ Real-time บนเบราว์เซอร์โดยตรง และไม่มีระบบสำรองจากมนุษย์

Islam และคณะ (2023) พัฒนาระบบช่วยเหลือสำหรับผู้พิการทางสายตาโดยใช้ TensorFlow Object Detection API ร่วมกับโมเดล SSDLite MobileNetV2 ที่ผ่านการฝึกบนชุดข้อมูล COCO เพื่อตรวจจับวัตถุและบรรยายสภาพแวดล้อมโดยรอบ แสดงให้เห็นถึงศักยภาพของโมเดลในตระกูล SSD/COCO สำหรับงานด้าน Assistive Technology โดยตรง

2.7.2 กลุ่มงานวิจัยด้านโมเดลภาษาขนาดใหญ่และการอ่านข้อความ

Tepsan และคณะ (2025) ศึกษาการประยุกต์ใช้ Generative AI ในการอ่านฉลากยาสำหรับผู้สูงอายุและผู้พิการทางสายตาในประเทศไทย พบว่าการใช้โมเดลภาษาขนาดใหญ่ (LLMs) ช่วยลดความผิดพลาดในการระบุชื่อยาและวิธีการรับประทานได้อย่างมีประสิทธิภาพ แต่ระบบ OCR แบบดั้งเดิมที่ใช้ในงานวิจัยก่อนหน้ายังประสบปัญหาในการรองรับฟอนต์และภาษาไทยที่มีความหลากหลายบนฉลากยาได้อย่างสมบูรณ์

2.7.3 กลุ่มงานวิจัยด้านการช่วยเหลือทางไกลและ PWA

Avila และคณะ (2016) ทำการสำรวจและวิจัยเชิงคุณภาพเกี่ยวกับการใช้งานแอปพลิเคชัน Be My Eyes โดยระบุว่าผู้พิการทางสายตามีความต้องการความช่วยเหลือจากมนุษย์สูงในสถานการณ์ที่ไม่คุ้นเคยหรือต้องการความละเอียดอ่อน เช่น การดูวันหมดอายุอาหาร หรือการเลือกเสื้อผ้า แต่ข้อจำกัดสำคัญที่พบคือความกังวลเรื่องความเป็นส่วนตัวและความไม่แน่นอนของระยะเวลาในการรอคอยอาสาสมัคร

Chaudary และคณะ (2023) พัฒนาระบบนำทางระยะไกล (Teleguidance) สำหรับผู้พิการทางสายตา โดยให้ผู้ดูแลที่อยู่ห่างไกลรับชมภาพวิดีโอสดจากกล้องสมาร์ตโฟนของผู้ใช้ และให้คำแนะนำทิศทางผ่านการสั่นสะเทือน (Haptic Feedback) และเสียงพูด ผลการศึกษาพบว่าการผสานระหว่างการนำทางด้วยมนุษย์ระยะไกลและ Haptic Feedback ช่วยเพิ่มความมั่นใจและความปลอดภัยของผู้ใช้ได้อย่างมีนัยสำคัญ ซึ่งสอดคล้องกับแนวคิด Human-in-the-Loop ในโครงงานนี้

Roumeliotis และ Tselikas (2022) ศึกษาความสามารถในการเข้าถึงของ Progressive Web Application (PWA) สำหรับผู้พิการในรูปแบบการทบทวนวรรณกรรมอย่างเป็นระบบ พบว่า PWA ที่พัฒนาตามมาตรฐานเว็บสามารถทำงานร่วมกับ Assistive Technology ได้อย่างมีประสิทธิภาพ และมีศักยภาพสูงในการทดแทนแอปพลิเคชันแบบ Native สำหรับผู้ใช้ที่มีความพิการ อย่างไรก็ตาม ยังพบช่องว่างด้านการรองรับโปรแกรมอ่านหน้าจอในบาง PWA Framework

2.7.4 สรุปและช่องว่างของงานวิจัย (Research Gap)

งานวิจัยส่วนใหญ่ในปัจจุบันมักแยกส่วนกันระหว่างระบบอัตโนมัติ (AI-only) และระบบมนุษย์ (Human-only) อีกทั้งแอปพลิเคชันจำนวนมากยังขาดการรองรับภาษาไทยและการออกแบบ PWA ที่คำนึงถึงการใช้งานร่วมกับ Screen Reader อย่างแท้จริง โครงงานนี้จึงมุ่งเน้นที่จะพัฒนาระบบผู้ช่วยอัจฉริยะบนแพลตฟอร์ม PWA ที่ผสานพลังของ Llama 4 Vision และเครือข่ายอาสาสมัคร เพื่อสร้างโซลูชันที่ครบวงจรและปลอดภัยที่สุดสำหรับผู้พิการทางสายตาในประเทศไทย

บทที่ 3
ระเบียบวิธีดำเนินงาน

การดำเนินโครงงานครั้งนี้มุ่งเน้นการพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับผู้พิการทางสายตาในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (PWA) โดยผู้ดำเนินโครงงานได้กำหนดระเบียบวิธีดำเนินงานเพื่อให้การดำเนินงานเป็นไปอย่างมีระบบและบรรลุวัตถุประสงค์ที่ตั้งไว้ ดังมีรายละเอียดต่อไปนี้

3.1 รูปแบบโครงงาน

การดำเนินโครงงานครั้งนี้เป็น โครงงานและการพัฒนา ซึ่งมีลักษณะผสมผสานระหว่างโครงงานเชิงประยุกต์และการพัฒนาระบบ (System Development) โดยมีขั้นตอนการดำเนินงานแบ่งออกเป็น 4 ระยะหลัก ดังนี้

ระยะที่ 1 การศึกษาและวิเคราะห์ความต้องการ (Requirement Analysis)
ผู้ดำเนินโครงงานทำการศึกษาปัญหาและความต้องการของผู้พิการทางสายตา ทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง รวมถึงวิเคราะห์เทคโนโลยีที่เหมาะสมสำหรับการพัฒนาระบบ เพื่อกำหนดขอบเขตและคุณสมบัติของระบบที่จะพัฒนา

ระยะที่ 2 การออกแบบและพัฒนาระบบ (System Design and Development)
ผู้ดำเนินโครงงานออกแบบสถาปัตยกรรมระบบ (System Architecture) และพัฒนาองค์ประกอบทั้ง 3 ส่วน ได้แก่ ระบบผู้ช่วยการมองเห็นอัจฉริยะ (AI Visual Assistant) ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization) และระบบเครือข่ายอาสาสมัคร (Volunteer Network System) โดยพัฒนาในรูปแบบ Progressive Web Application บนสมาร์ตโฟน

ภาพที่ 3.1 การออกแบบและพัฒนาระบบ (หน้าแรก)

จากภาพที่ 3.1 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

(1) หน้าเริ่มต้น (Landing Page) สำหรับกดเริ่มใช้งานแอปพลิเคชัน

(2) I Need Help ฟังชั่นของผู้พิการทางสายตา

(3) Volunteer ฟังชั่นของอาสา

ภาพที่ 3.2 การออกแบบและพัฒนาระบบ

จากภาพที่ 3.2 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

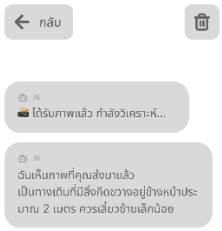
(1) เมื่อเล็งกล้องตามคำแนะนำของ AI จนเหมาะสมแล้ว ให้แตะกึ่งกลางหน้าจอเพื่อทำการบันทึกภาพนิ่ง

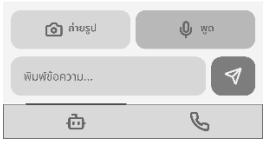
(2) ระบบจะส่งเสียงชัตเตอร์และสั่นเตือน (Haptic) เพื่อยืนยัน พร้อมสลับเข้าสู่หน้าจอแชท AI ทันที

(3) ภาพที่ถ่ายจะปรากฏในช่องแชทพร้อมสถานะ กำลังวิเคราะห์ภาพ เพื่อส่งข้อมูลไปยัง Cloud-AIรองรับ

(4) เมื่อประมวลผลเสร็จสิ้น AI จะส่งข้อความบรรยายรายละเอียดวัตถุพร้อมระบบเสียงอ่านอัตโนมัติ

ภาพที่ 3.3 การออกแบบและพัฒนาระบบ





จากภาพที่ 3.3 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

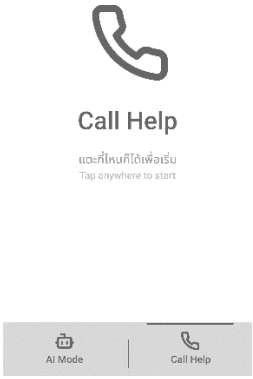
(1) แสดงข้อความวิเคราะห์ภาพจากโมเดล AI พร้อมระบบอ่านออกเสียง (TTS) อัตโนมัติเพื่อให้ผู้ใช้รับรู้ข้อมูลได้ทันที

(2) ปุ่มควบคุมหลัก ประกอบด้วยปุ่มถ่ายรูปเพื่อเริ่มวิเคราะห์ภาพใหม่ ปุ่มกดค้างเพื่อถามด้วยเสียง (Hold-to-talk) และปุ่มล้างประวัติการสนทนา

(3) AI สามารถจำประวัติการสนทนาและภาพก่อนหน้าได้สูงสุด 6 ข้อความ ทำให้ผู้ใช้สามารถสอบถามข้อมูลต่อเนื่องได้ทันที

(4) ระบบมีการสั่นเตือนเพื่อยืนยันสถานะการส่งข้อความและการประมวลผลเสร็จสิ้น เพื่อสร้างความมั่นใจในการใช้งาน

ภาพที่ 3.4 การออกแบบและพัฒนาระบบ



จากภาพที่ 3.4 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

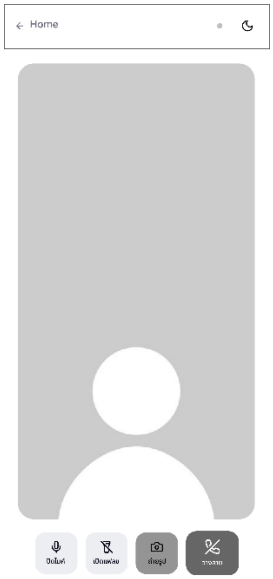
(1) ออกแบบปุ่มไอคอนโทรศัพท์ขนาดใหญ่กึ่งกลางหน้าจอ รองรับการ "แตะพื้นที่ใดก็ได้" เพื่อเริ่มส่งสัญญาณขอความช่วยเหลือ ช่วยให้ผู้พิการใช้งานได้ทันทีโดยไม่ต้องควานหาปุ่ม

(2) แถบเมนูด้านล่างสำหรับสลับการทำงานระหว่างระบบวิเคราะห์ด้วย AI และระบบขอความช่วยเหลือจากอาสาสมัคร

(3) เมื่อผู้ใช้กดปุ่มเรียก ระบบจะส่งคำขอ (Signaling) ไปยังอาสาสมัครที่กำลังออนไลน์อยู่ในขณะนั้นแบบทันที

(4) เมื่ออาสาสมัครกดรับสาย ระบบจะสถาปนาการเชื่อมต่อแบบวิดีโอคอลผ่านโปรโตคอล WebRTC เพื่อให้เห็นภาพจากกล้องและสื่อสารด้วยเสียงได้แบบเรียลไทม์

ภาพที่ 3.5 การออกแบบและพัฒนาระบบ



จากภาพที่ 3.5 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

(1) พื้นที่แสดงภาพสดแบบเต็มจอจากกล้องของผู้พิการทางสายตา เพื่อให้อาสาสมัครช่วยตรวจสอบวัตถุหรือสภาพแวดล้อมได้อย่างชัดเจน

(2) อาสาสมัครสามารถสั่งเปิดไฟแฟลช (Flashlight Toggle) บนสมาร์ตโฟนของผู้พิการได้โดยตรงเมื่ออยู่ในที่แสงน้อย

(3) ประกอบด้วยปุ่มสั่งถ่ายภาพนิ่งเพื่อเก็บรายละเอียดที่ชัดเจนขึ้น ปุ่มควบคุมไมโครโฟน และปุ่มยุติการเชื่อมต่อ (End Session)

(4) ใช้เทคโนโลยี WebRTC ผ่าน PeerJS ในการส่งคำสั่งควบคุมฮาร์ดแวร์และสื่อสารภาพและเสียงแบบความหน่วงต่ำ (Real-time)

ระยะที่ 3 การทดสอบและประเมินประสิทธิภาพ (Testing and Evaluation)

ผู้ดำเนินโครงงานนำระบบต้นแบบไปทดสอบกับตัวชี้วัดประสิทธิภาพของระบบในสภาพแวดล้อมภายในบ้านจริง โดยกำหนดตัวชี้วัด ดังนี้

(1) ประสิทธิภาพในการจำแนกวัตถุ (Object Classification Performance)

(2) ความถูกต้องในการระบุตำแหน่งวัตถุ (Localization Accuracy)

(3) ระยะเวลาเฉลี่ยในการค้นหาวัตถุ

ระยะที่ 4 การสรุปผลและปรับปรุงระบบ (Conclusion and Improvement)

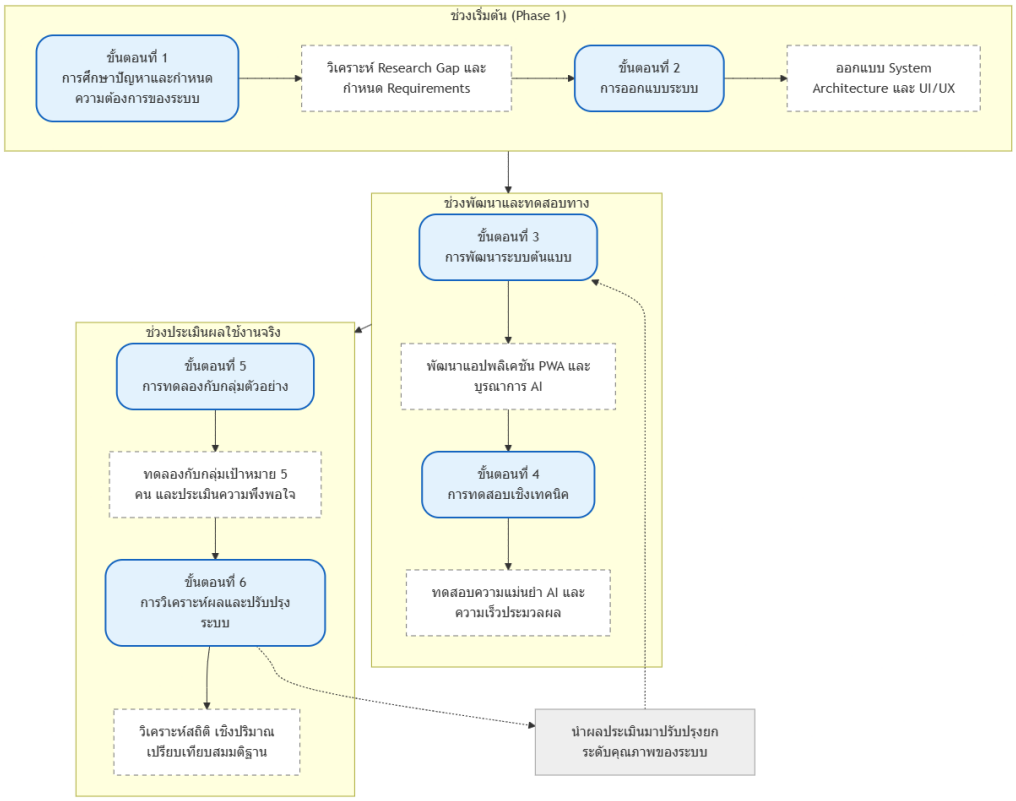
ผู้ดำเนินโครงงานนำผลการทดสอบมาวิเคราะห์ สรุปผล และปรับปรุงระบบให้มีประสิทธิภาพมากยิ่งขึ้น พร้อมเสนอแนวทางการพัฒนาในอนาคต กระบวนการดำเนินงานดังกล่าวมีลักษณะเป็นวงจร (Iterative Process) โดยผลลัพธ์จากระยะที่ 3 จะถูกนำกลับไปปรับปรุงในระยะที่ 2 เพื่อพัฒนาเวอร์ชันที่มีประสิทธิภาพมากยิ่งขึ้น

การเลือกใช้รูปแบบโครงงานและพัฒนา (R&D) มีความเหมาะสมกับลักษณะของโครงงาน เนื่องจากเป็นงานที่มุ่งสร้างนวัตกรรมเชิงเทคโนโลยี ควบคู่กับการประเมินประสิทธิภาพในสถานการณ์จริง เพื่อให้ได้ต้นแบบที่สามารถนำไปใช้งานได้จริงในบริบทของผู้พิการทางสายตาในประเทศไทย

3.1.1 กรอบแนวคิดการดำเนินงาน (Operational Framework)

การดำเนินงานโครงงานครั้งนี้ใช้แนวคิดการทำโครงงานและพัฒนา โดยกำหนดกระบวนการดำเนินงานเป็นลำดับขั้นอย่างเป็นระบบ ตั้งแต่การวิเคราะห์ปัญหาไปจนถึงการประเมินผลและปรับปรุงระบบ โดยสามารถสรุปเป็นกรอบแนวคิดการดำเนินงานได้ดังนี้

ภาพที่ 3.6 กรอบแนวคิดการดำเนินงานโครงงาน



จากภาพที่ 3.6 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

(1) การศึกษาปัญหาและวิเคราะห์ความต้องการ

(2) การออกแบบระบบ

(3) การพัฒนาระบบ

(4) การทดสอบระบบ

(5) การประเมินผลกับกลุ่มตัวอย่าง

(6) การวิเคราะห์ผลและปรับปรุงระบบ

3.1.2 ขั้นตอนการดำเนินงานโครงงาน

การดำเนินงานโครงงานครั้งนี้ดำเนินการตามแนวทางการทำโครงงานและพัฒนา โดยกำหนดขั้นตอนการดำเนินงานอย่างเป็นลำดับ เพื่อให้สอดคล้องกับวัตถุประสงค์ของการพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับผู้พิการทางสายตา ทั้งในด้านเทคนิคและการประเมินผลการใช้งานจริง โดยมีรายละเอียดดังต่อไปนี้

ขั้นตอนที่ 1 การศึกษาปัญหาและกำหนดความต้องการของระบบ ผู้จัดทำโครงงานศึกษาปัญหาและข้อจำกัดในการดำเนินชีวิตประจำวันของผู้พิการทางสายตา โดยเฉพาะปัญหาการจำแนกและระบุตำแหน่งสิ่งของภายในบ้าน จากนั้นทบทวนเอกสาร แนวคิด และงานวิจัยที่เกี่ยวข้องกับเทคโนโลยีปัญญาประดิษฐ์ การตรวจจับวัตถุ ระบบช่วยเหลือผู้พิการ และแนวคิดมนุษย์ร่วมในวงรอบ (Human-in-the-Loop) เพื่อวิเคราะห์ช่องว่างของงานวิจัย (Research Gap) และกำหนดคุณลักษณะของระบบ (System Requirements) ทั้งในด้านความสามารถหลักของระบบ (Functional Requirements) และข้อกำหนดด้านประสิทธิภาพและการเข้าถึง (Non-functional Requirements) โดยได้กำหนดความต้องการของระบบ (System Requirements) แบ่งออกเป็น 2 ส่วนหลัก ดังตารางที่ 3.1

ตารางที่ 3.1

ความต้องการของระบบแยกตามส่วนการทำงาน

ส่วนของระบบ	Functional Requirements	Non-Functional Requirements
กล้องและการรับภาพ	ระบบต้องเปิดกล้องสมาร์ตโฟนและสตรีมภาพแบบ Real-time ได้อย่างต่อเนื่อง	ระบบต้องประมวลผลภาพได้ไม่ต่ำกว่า 10 เฟรมต่อวินาที (FPS) เพื่อให้การนำทางด้วยเสียงลื่นไหล
การตรวจจับวัตถุ (COCO-SSD)	ระบบต้องตรวจจับและจำแนกประเภทวัตถุภายในครัวเรือนจากภาพกล้องได้แบบ Real-time และระบุตำแหน่งวัตถุพร้อมส่งเสียงแจ้งทิศทาง (ซ้าย ขวา บน ล่าง)	ระบบต้องมีความแม่นยำในการจำแนกวัตถุไม่ต่ำกว่าร้อยละ 80 และประมวลผลบนอุปกรณ์โดยตรง (On-device) โดยไม่ต้องใช้อินเทอร์เน็ต
การวิเคราะห์ภาพด้วย AI (Llama 4 Vision	ระบบต้องบรรยายรายละเอียดวัตถุและบริบทในภาพเป็นภาษาไทย อ่านฉลากสินค้าภาษาไทย (OCR) และยืนยันความถูกต้องของวัตถุหลังหยิบจับได้	ระบบต้องตอบสนองต่อคำสั่งเสียงและส่งผลลัพธ์กลับมาเป็นเสียงพูดได้ภายในเวลาที่เหมาะสม และรองรับการจำประวัติการสนทนาสูงสุด 6 ข้อความล่าสุด
การรับคำสั่งเสียง (Web Speech API)	ระบบต้องรับคำสั่งเสียงภาษาไทยจากผู้ใช้และแปลงเป็นข้อความได้ รวมถึงแปลงผลลัพธ์จาก AI กลับมาเป็นเสียงพูดได้	ระบบต้องทำงานร่วมกับโปรแกรมอ่านหน้าจอ TalkBack (Android) และ VoiceOver (iOS) ได้อย่างสมบูรณ์
ระบบอาสาสมัคร (WebRTC/Pusher)	ระบบต้องส่งสัญญาณแจ้งเตือนอาสาสมัครที่ออนไลน์อยู่ได้ทันที และสถาปนาการเชื่อมต่อวิดีโอคอลแบบ Peer-to-Peer ระหว่างผู้พิการทางสายตาและ	ระบบต้องมีระบบสำรอง (Fallback System) ที่พร้อมทำงานได้ทันทีเมื่อ AI ไม่สามารถจำแนกวัตถุได้อย่าง

	อาสาสมัครได้ รวมถึงให้อาสาสมัครควบคุมไฟแฟลชระยะไกลได้	มั่นใจ และรองรับการเชื่อมต่อได้อย่างเสถียร
ส่วนติดต่อผู้ใช้ (PWA/Next.js)	ระบบต้องติดตั้งบนหน้าจอหลักของสมาร์ตโฟนได้โดยไม่ต้องผ่านร้านค้าแอปพลิเคชัน และรองรับการสั่งงานด้วยการสัมผัสเพียงครั้งเดียว (One-Tap)	ระบบต้องรองรับ Android และ iOS แจ้งสถานะผ่านการสั่นสะเทือน (Haptic Feedback) และป้องกันหน้าจอดับระหว่างใช้งานวิดีโอคอลผ่าน Wake Lock API

ขั้นตอนที่ 2 การออกแบบระบบ ภายหลังจากกำหนดความต้องการของระบบ ผู้จัดทำโครงงานได้ดำเนินการออกแบบสถาปัตยกรรมระบบ (System Architecture) และโครงสร้างการทำงานขององค์ประกอบหลักทั้ง 3 ส่วน ได้แก่ ระบบผู้ช่วยการมองเห็นอัจฉริยะ ระบบตรวจจับและระบุตำแหน่งวัตถุ และระบบเครือข่ายอาสาสมัคร พร้อมทั้งกำหนดกระบวนการไหลของข้อมูล (Data Flow) ระหว่างส่วนประมวลผลภาพ ส่วนประมวลผลภาษา และระบบสื่อสารแบบเรียลไทม์ นอกจากนี้ ได้ออกแบบส่วนติดต่อผู้ใช้ (User Interface: UI) ให้สอดคล้องกับหลักการออกแบบเพื่อการเข้าถึง (Accessibility Design) เพื่อให้สามารถใช้งานร่วมกับโปรแกรมอ่านหน้าจอ (Screen Reader) ได้อย่างมีประสิทธิภาพ

ขั้นตอนที่ 3 การพัฒนาระบบต้นแบบ ในขั้นตอนนี้ ผู้ดำเนินโครงงานพัฒนาเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA) สำหรับสมาร์ตโฟน โดยบูรณาการเทคโนโลยีปัญญาประดิษฐ์ในการประมวลผลภาพและเสียง ระบบตรวจจับวัตถุถูกพัฒนาเพื่อทำหน้าที่ระบุชนิดและตำแหน่งของวัตถุแบบเรียลไทม์ ขณะที่โมเดลภาษาขนาดใหญ่ที่รองรับภาพถูกนำมาใช้เพื่อบรรยายรายละเอียดของวัตถุและอ่านข้อความภาษาไทยจากภาพ นอกจากนี้ ได้พัฒนาระบบเครือข่ายอาสาสมัครเพื่อทำหน้าที่เป็นระบบสำรอง (Fallback System) ในกรณีที่ระบบอัตโนมัติไม่สามารถให้ผลลัพธ์ที่ชัดเจนหรือมีค่าความเชื่อมั่นต่ำกว่าค่าที่กำหนด

ขั้นตอนที่ 4 การทดสอบเชิงเทคนิค หลังจากพัฒนาระบบต้นแบบเสร็จสิ้น ผู้ดำเนินโครงงานได้ทำการทดสอบประสิทธิภาพของระบบในเชิงเทคนิค โดยประเมินความแม่นยำในการจำแนกวัตถุ (Object Classification Accuracy) ความถูกต้องในการระบุตำแหน่งวัตถุ (Localization Accuracy) และความรวดเร็วในการประมวลผล (Latency) เพื่อให้มั่นใจว่าระบบสามารถทำงานได้อย่างเสถียรและเหมาะสมต่อการใช้งานจริง

ขั้นตอนที่ 5 การวิเคราะห์ผลและปรับปรุงระบบ ภายหลังจากการเก็บรวบรวมข้อมูล ผู้ดำเนินโครงงานดำเนินการวิเคราะห์ข้อมูลเชิงปริมาณโดยใช้สถิติพื้นฐาน เช่น ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน เพื่อประเมินประสิทธิภาพของระบบตามตัวชี้วัดที่กำหนด และเปรียบเทียบกับสมมติฐาน จากนั้นนำผลการวิเคราะห์มาปรับปรุงและพัฒนาระบบให้มีประสิทธิภาพและความเหมาะสมต่อการใช้งานมากยิ่งขึ้น

ทั้งนี้ กระบวนการดำเนินงานดังกล่าวมีลักษณะเป็นวงจร (Iterative Process) โดยผลการประเมินจากขั้นตอนการทดลองสามารถนำกลับไปใช้ในการปรับปรุงในขั้นตอนการพัฒนา เพื่อยกระดับคุณภาพของระบบอย่างต่อเนื่อง

3.2 เครื่องมือที่ใช้ในโครงงาน

การจัดทำโครงงานครั้งนี้เป็นการพัฒนาและประเมินระบบผู้ช่วยอัจฉริยะสำหรับผู้พิการทางสายตาในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA) โดยเครื่องมือที่ใช้ในโครงงานแบ่งออกเป็น 3 ส่วนหลัก ได้แก่ เครื่องมือด้านการพัฒนาระบบ เครื่องมือด้านปัญญาประดิษฐ์ และเครื่องมือด้านการเก็บรวบรวมข้อมูล ดังรายละเอียดต่อไปนี้

3.2.1 เครื่องมือด้านการพัฒนาระบบ (System Development Tools)

3.2.1.1 แพลตฟอร์มการพัฒนา

ระบบถูกพัฒนาในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (PWA) เพื่อให้สามารถใช้งานผ่านเว็บเบราว์เซอร์บนสมาร์ตโฟน และติดตั้งบนหน้าจอหลักได้โดยไม่ต้องผ่านร้านค้าแอปพลิเคชัน

3.2.1.2 ภาษาและเทคโนโลยีที่ใช้พัฒนา

(1) ภาษา JavaScript สำหรับพัฒนาส่วนติดต่อผู้ใช้ (Frontend)

(2) HTML5 และ CSS3 สำหรับโครงสร้างและการจัดรูปแบบหน้าจอ

(3) WebRTC สำหรับการสื่อสารวิดีโอคอลระหว่างผู้ใช้และอาสาสมัคร

3.2.1.3 สถาปัตยกรรมระบบ

ระบบใช้สถาปัตยกรรมแบบ Client–Server โดยส่วนของ PWA ทำหน้าที่รับภาพจากกล้องสมาร์ตโฟนและส่งข้อมูลไปยังเซิร์ฟเวอร์เพื่อประมวลผล จากนั้นส่งผลลัพธ์กลับมาในรูปแบบเสียงสังเคราะห์

3.2.2 เครื่องมือด้านปัญญาประดิษฐ์ (Artificial Intelligence Tools)

(1) ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization)

ระบบใช้ไลบรารี TensorFlow.js ร่วมกับโมเดล COCO-SSD (Common Objects in Context - Single Shot MultiBox Detector) ซึ่งเป็นโมเดลปัญญาประดิษฐ์ขนาดเล็กที่สามารถตรวจจับและจำแนกวัตถุพื้นฐานได้แบบเรียลไทม์ ระบบรับภาพวิดีโอจากกล้องหลังของสมาร์ตโฟนผ่าน getUserMedia API และประมวลผลในอัตรา 10 เฟรมต่อวินาที เพื่อรักษาสมดุลระหว่างความเร็วในการตอบสนองและการประหยัดพลังงานแบตเตอรี่ เมื่อโมเดลตรวจพบวัตถุและคืนค่าพิกัดกรอบสี่เหลี่ยม (Bounding Box) ระบบจะคำนวณจุดกึ่งกลางของวัตถุ (objCenterX, objCenterY) และเปรียบเทียบกับจุดกึ่งกลางของหน้าจอ (frameCenterX, frameCenterY) โดยใช้สูตรระยะทางแบบยุคลิด (Euclidean Distance) เพื่อระบุวัตถุที่อยู่ใกล้จุดโฟกัสมากที่สุด

(2) โมเดลภาษาขนาดใหญ่ที่รองรับภาพ (Large Multimodal Model: LMM)

ระบบใช้โมเดล Llama 4 Vision ผ่าน Groq API ซึ่งเป็น Large Multimodal Model (LMM) ที่มีความสามารถทั้งในการรับรู้ภาพ (Computer Vision) และประมวลผลภาษาธรรมชาติ (Natural Language Processing) รองรับการตอบสนองเป็นภาษาไทย เมื่อผู้ใช้สั่งงานด้วยเสียง ระบบจะแปลงเสียงเป็นข้อความ (Speech-to-Text) ผ่าน Web Speech API พร้อมกันนั้นจะดึงภาพจากวิดีโอในขณะนั้นมาวาดลงบน Canvas ชั่วคราว และแปลงเป็นข้อมูล Base64 รูปแบบ JPEG เพื่อส่งไปยัง Groq API จากนั้นระบบจะรวมภาพและคำถามของผู้ใช้เข้ากับ System Prompt ที่กำหนด Priority Framework ไว้ล่วงหน้า โดยเรียงลำดับความสำคัญดังนี้

ลำดับที่ 1 ตรวจสอบสิ่งบดบังหน้ากล้อง เช่น นิ้วมือบังเลนส์

ลำดับที่ 2 อ่านข้อความและฉลากสินค้าภาษาไทย (OCR)

ลำดับที่ 3 ตรวจสอบสภาพแวดล้อมด้านความปลอดภัย เช่น สิ่งกีดขวาง

ลำดับที่ 4 บรรยายบริบททั่วไปของวัตถุในภาพ

นอกจากนี้ระบบยังรองรับ Contextual Memory โดยเก็บประวัติการสนทนาสูงสุด 6 ข้อความล่าสุด เพื่อให้โมเดลสามารถตอบคำถามต่อเนื่องได้อย่างสอดคล้องกับบริบท เมื่อโมเดลประมวลผลเสร็จสิ้น ผลลัพธ์จะถูกแสดงบนหน้าจอเพื่อให้โปรแกรมอ่านหน้าจอ (Screen Reader) อ่านออกเสียงให้ผู้ใช้รับฟังได้ทันที

(3) ระบบกำหนดค่าความเชื่อมั่น (Confidence Threshold)

ระบบกำหนดค่าความเชื่อมั่นขั้นต่ำในการจำแนกวัตถุโดยใช้ Tolerance Zone ที่ 20% ของจุดกึ่งกลางหน้าจอเป็นเกณฑ์ในการพิจารณาความแม่นยำของการระบุตำแหน่ง แบบเรียลไทม์ เช่น ซ้าย ขวา บน หรือล่าง เพื่อช่วยให้ผู้ใช้เล็งกล้องไปยังตำแหน่งของวัตถุได้อย่างแม่นยำ และเมื่อวัตถุเข้าสู่พื้นที่เป้าหมาย ระบบจะแจ้งเตือนว่าพร้อมถ่ายภาพ หากระบบไม่สามารถตรวจพบวัตถุเป้าหมายหรือค่าความเชื่อมั่นต่ำกว่าเกณฑ์ที่กำหนด ระบบจะเปิดใช้งานระบบเครือข่ายอาสาสมัคร (Volunteer Network System) เพื่อเป็นกลไกสำรอง (Fallback Mechanism) ให้ผู้ใช้สามารถขอความช่วยเหลือจากมนุษย์ได้ทันที

3.2.3 เครื่องมือด้านอุปกรณ์และสภาพแวดล้อมในการทดลอง

(1) สมาร์ตโฟนระบบปฏิบัติการ Android หรือ iOS ที่รองรับกล้องถ่ายภาพและโปรแกรมอ่านหน้าจอ (TalkBack หรือ VoiceOver)

(2) เครือข่ายอินเทอร์เน็ตสำหรับการเชื่อมต่อแบบเรียลไทม์

(3) สภาพแวดล้อมภายในบ้าน (Indoor Environment) สำหรับทดสอบการค้นหาและจำแนกวัตถุ

3.2.4 เครื่องมือด้านการเก็บรวบรวมข้อมูล (Data Collection Instruments)

แบบบันทึกผลการทดลอง ใช้สำหรับบันทึกผลการจำแนกวัตถุ ความถูกต้องของตำแหน่ง และระยะเวลาในการค้นหาแต่ละรายการ

3.2.5 ตัวชี้วัดประสิทธิภาพของระบบ (Performance Metrics)

เพื่อประเมินประสิทธิภาพของระบบ ผู้ดำเนินโครงงานได้กำหนดตัวชี้วัดดังนี้

(1) ประสิทธิภาพในการจำแนกวัตถุ (Object Classification Performance) ประเมินผ่านตารางความสับสน (Confusion Matrix) เพื่อวิเคราะห์ความถูกต้องของการทำนาย โดยมีพารามิเตอร์พื้นฐาน 4 ค่า ได้แก่

True Positive (TP) : ระบบทำนายว่าพบวัตถุที่ต้องการ และวัตถุนั้นมีอยู่จริง

True Negative (TN) : ระบบทำนายว่าไม่พบวัตถุ และวัตถุนั้นไม่มีอยู่จริง

False Positive (FP) : ระบบทำนายว่าพบวัตถุ แต่วัตถุนั้นไม่มีอยู่จริง (ทำนายผิดว่าเจอ)

False Negative (FN) : ระบบทำนายว่าไม่พบวัตถุ ทั้งที่วัตถุนั้นมีอยู่จริง (หาไม่เจอ) จากนั้นนำพารามิเตอร์ดังกล่าวมาคำนวณเป็นตัวชี้วัดประสิทธิภาพ ดังนี้

1. ความแม่นยำ (Accuracy) : อัตราส่วนของการทำนายถูกต้องทั้งหมด

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. ความเที่ยงตรง (Precision) : อัตราส่วนของการทำนายว่าพบวัตถุที่ถูกต้องจริง

$$Precision = \frac{TP}{TP + FP}$$

3. ความระลึก (Recall) : อัตราส่วนของวัตถุที่มีอยู่จริงและระบบสามารถหาเจอ

$$Recall = \frac{TP}{TP + FN}$$

4. ค่าเอฟวัน (F1-Score) : ค่าเฉลี่ยฮาร์โมนิกระหว่างความเที่ยงตรงและความระลึก เพื่อประเมินความสมดุลของโมเดล

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

(2) ความถูกต้องในการระบุตำแหน่งวัตถุ (Localization Accuracy) ประเมินจากความถูกต้องของตำแหน่งกรอบวัตถุเทียบกับตำแหน่งจริง โดยพิจารณาความสอดคล้องของทิศทางที่ระบบแนะนำกับตำแหน่งวัตถุจริงในภาพ

(3) ระยะเวลาเฉลี่ยในการค้นหาวัตถุ วัดจากระยะเวลาตั้งแต่ผู้ใช้เริ่มเปิดกล้องจนสามารถระบุวัตถุได้สำเร็จ โดยคำนวณค่าเฉลี่ยจากการทดลองทั้งหมด

3.3 การออกแบบและพัฒนาระบบ

การพัฒนาระบบผู้ช่วยอัจฉริยะสำหรับผู้พิการทางสายตาในโครงงานนี้ ดำเนินการตามแนวคิดสถาปัตยกรรมแบบผสมผสาน (Hybrid Architecture) ที่ผสานการทำงานระหว่างระบบปัญญาประดิษฐ์ (AI-based Automation) และระบบมนุษย์ร่วมในวงรอบ (Human-in-the-Loop) เพื่อเพิ่มความแม่นยำ ความปลอดภัย และความน่าเชื่อถือของผลลัพธ์ โดยระบบถูกพัฒนาในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (Progressive Web Application: PWA) บนสมาร์ตโฟน

3.3.1 ภาพรวมสถาปัตยกรรมระบบ (System Architecture Overview)

ระบบถูกออกแบบในลักษณะ Client–Server Architecture โดยแบ่งองค์ประกอบออกเป็น 4 ส่วนหลัก ได้แก่

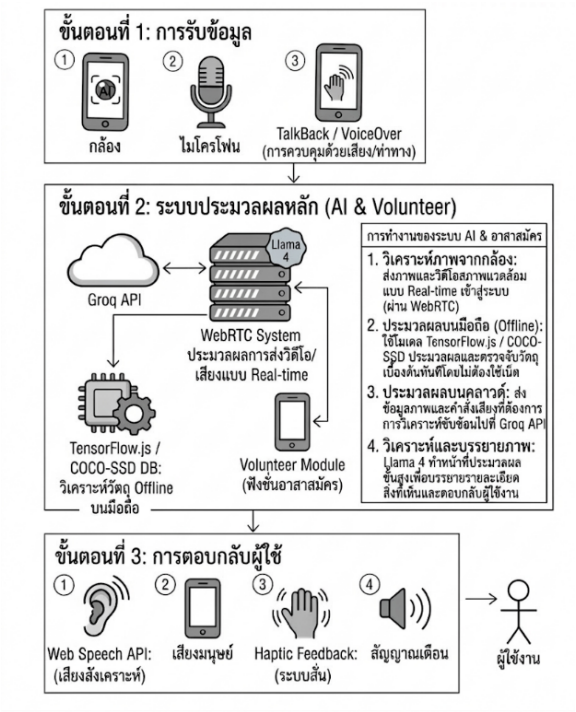
(1) ฝั่งผู้ใช้งาน (Client Side – PWA) — หน้าจอผู้พิการที่รับเสียงและกล้อง

(2) ระบบ AI นำทางบนมือถือ — TensorFlow.js ตรวจจับวัตถุและคำนวณทิศทาง

(3) ระบบ AI อัจฉริยะแบบคลาวด์ — Llama 4Vision ผ่าน Groq API

(4) ระบบเครือข่ายอาสาสมัคร — Pusher + PeerJS + Volunteer Dashboard

ภาพที่ 3.7 ภาพรวมสถาปัตยกรรมระบบ



จากภาพ 3.7 สามารถอธิบายกระบวนการทำงานของระบบโดยสังเขปได้ดังนี้

(1) กล้องสมาร์ทโฟนทำหน้าที่สตรีมภาพแบบ Real-time เพื่อนำข้อมูลภาพเข้าสู่ระบบสำหรับการประมวลผล

(2) ไมโครโฟนใช้สำหรับรับคำสั่งเสียงจากผู้ใช้ (Voice Input)

(3) TalkBack / VoiceOver ทำหน้าที่รองรับการควบคุมระบบด้วยเสียงและท่าทาง เพื่ออำนวยความสะดวกแก่ผู้ใช้ที่มีความบกพร่องทางการมองเห็น

(4) Offline Object Detection ทำหน้าที่ตรวจจับวัตถุจากภาพที่ได้จากกล้องแบบเรียลไทม์ โดยใช้โมเดล TensorFlow.js / COCO-SSD ในการวิเคราะห์ประเภทและตำแหน่งของวัตถุภายในภาพโดยไม่ต้องใช้อินเทอร์เน็ต

(5) Cloud Image Analysis ทำหน้าที่ส่งข้อมูลภาพและคำถามของผู้ใช้ไปยังระบบประมวลผลบนคลาวด์ผ่าน Groq API เพื่อให้โมเดล Llama 4 วิเคราะห์ภาพและสร้างคำอธิบายเพิ่มเติม

(6) Volunteer Module เป็นระบบเชื่อมต่อผู้ใช้งานกับอาสาสมัคร โดยผู้ใช้สามารถขอความช่วยเหลือและระบบจะจัดคิวอาสาสมัครเพื่อให้การช่วยเหลือแบบเรียลไทม์ผ่าน WebRTC

(7) ระบบตอบกลับผู้ใช้ (Output) โดยผลการวิเคราะห์จะถูกแปลงเป็นเสียงพูดผ่าน Web Speech API หรือเสียงมนุษย์จากอาสาสมัคร พร้อมทั้งการสั่นสะเทือน (Haptic Feedback) และสัญญาณเตือน เพื่อช่วยให้ผู้ใช้รับรู้ข้อมูลและสิ่งแวดล้อมรอบตัวได้

3.3.2 องค์ประกอบของระบบ

3.3.2.1 ส่วนติดต่อผู้ใช้สำหรับผู้ใช้ที่ตาบอด (Blind User Interface – PWA)

ระบบถูกพัฒนาในรูปแบบหน้าเว็บเบราว์เซอร์ล้ำหน้า Progressive Web Application แบบพร้อมให้บริการผ่านเซิฟเวอร์อัจฉริยะแบบ Next.js (SSR) สามารถแสดงผลและเข้าถึงอย่างเต็มรูปแบบ ออกแบบตามหลักการ Accessibility Design โดยเฉพาะ "Eyes-Free" เพื่อให้ทำงานร่วมกับโปรแกรมอ่านหน้าจอและอินเทอร์เฟซผู้ใช้แบบเรียลดรอฟท์ ได้อย่างมีประสิทธิภาพสูงสุด โดยมีฟังก์ชันหลัก ประกอบด้วย

(1) ระบบส่งคำสั่งแบบสัมผัสครั้งเดียวรวดเร็ว (One-Tap Calling)

(2) รับคำสั่งเสียงจากผู้ใช้ (Speech-to-Text) เชื่อมโยงบริบทแวดล้อมระบบคิวเพื่อถาม (Ask the Scene) แบบอัตโนมัติ

(3) ส่งเสียงตอบกลับด้วยตัวแปลข้อความเป็นเสียงตามข้อมูล (Text-to-Audible) ลำปักสัญญาณภาพและกรอบสถานะ

(4) การตอบโต้ด้วยการสั่นสะเทือนแบบสัมผัส (Haptic FeedbackAPI) เพื่อย้ำสถานะของการส่งและกระทำผ่าน Earcons

(5) การเชื่อมต่อวิดีโอคอลกับอาสาสมัครแบบ Real-Time Streaming ทันทีพร้อมป้องกันหน้าจอดับ (Wake Lock API)

3.3.2.2 ระบบเครือข่ายศูนย์อาสาสมัคร (Volunteer Dashboard Component)

โครงสร้างส่วนนี้ออกแบบมาเพื่อระบบประสานให้ความช่วยเหลือโดยบุคลากรภายนอก เมื่อเกิดข้อบกพร่องตามข้อตกลง โดยเชื่อมต่อผ่านเครือข่าย WebRTC (PeerJS) ทันทีบนมาตรฐานแบบตอบสนองเร็วสูงแบบ Peer-to-Peer บทบาทของอาสาสมัคร ได้แก่

(1) ตรวจสอบประเมินสถานภาพและวัตถุและรอบข้างผ่านกล้องหลังจากกดให้สิทธิใช้งาน (Rear Camera Feed)

(2) แจ้งให้คำแนะนำด้วยเสียงอย่างสะดวก

(3) ควบคุมอุปกรณ์ระยะไกล เช่น สั่งงานไฟฉายสมาร์ตโฟนผู้ใช้จากระยะไกลในบริเวณแสงน้อย (Flashlight Toggle Remote Control) รวมทั้งดูแลควบคุมเสียงตอบสนอง

3.3.2.3 ระบบวิเคราะห์ผลด้วยปัญญาประดิษฐ์จากส่วนกลาง (AI Visual Assistant Engine Mode)

ระบบวิเคราะห์ภาพด้วยปัญญาประดิษฐ์ถูกออกแบบให้ทำงานเป็น 2 ขั้นตอนต่อเนื่องกัน โดยเริ่มจากการนำทางกล้องด้วยการประมวลผลบนอุปกรณ์ (On-device) และต่อด้วยการวิเคราะห์เชิงลึกผ่านระบบคลาวด์ (Cloud-based) ดังนี้

ขั้นตอนที่ 1 การนำทางกล้องด้วย COCO-SSD (On-device Camera Guidance)

ระบบใช้โมเดล COCO-SSD ผ่าน TensorFlow.js ทำหน้าที่เป็นผู้ช่วยเล็งกล้องในขั้นตอนแรก โดยประมวลผลบนเบราว์เซอร์ของอุปกรณ์โดยตรงเพื่อลดความหน่วง ระบบรับภาพจากกล้องหลังผ่าน getUserMedia API และประมวลผลในอัตรา 10 เฟรมต่อวินาที เมื่อตรวจพบวัตถุแล้ว ระบบจะคำนวณระยะห่างระหว่างจุดกึ่งกลางของวัตถุกับจุดกึ่งกลางหน้าจอด้วยสูตร Euclidean Distance และส่งคำแนะนำทิศทางผ่านเสียงแบบเรียลไทม์จนกว่าวัตถุจะเข้าสู่ Tolerance Zone 20% ของจุดกึ่งกลางหน้าจอ เมื่อวัตถุอยู่ในตำแหน่งที่เหมาะสมแล้ว ระบบจะแจ้งเตือนผู้ใช้ว่าพร้อมสำหรับการวิเคราะห์เชิงลึกในขั้นตอนถัดไป

ขั้นตอนที่ 2 การวิเคราะห์เชิงลึกด้วย Llama 4 Vision (Cloud-based Scene Analysis)

เมื่อผู้ใช้สั่งงานด้วยเสียง ระบบจะดำเนินการดังนี้

(1) แปลงเสียงเป็นข้อความ (Speech-to-Text) ด้วย Web Speech API พร้อมรองรับพจนานุกรมภาษาไทย

(2) ดึงภาพจากวิดีโอในขณะนั้นมาวาดลงบน Canvas ชั่วคราว และแปลงเป็นข้อมูล Base64 รูปแบบ JPEG

(3) รวมภาพและคำถามของผู้ใช้เข้ากับ System Prompt ที่กำหนด Priority Framework ไว้ล่วงหน้า โดยเรียงลำดับความสำคัญดังนี้

ลำดับที่ 1 ตรวจสอบสิ่งบดบังหน้ากล้อง เช่น นิ้วมือบังเลนส์

ลำดับที่ 2 อ่านข้อความและฉลากสินค้าภาษาไทย (OCR)

ลำดับที่ 3 ตรวจสอบสภาพแวดล้อมด้านความปลอดภัย เช่น สิ่งกีดขวาง

ลำดับที่ 4 บรรยายบริบททั่วไปของวัตถุในภาพ

(4) ส่งข้อมูลภาพ คำถาม และประวัติการสนทนาสูงสุด 6 ข้อความล่าสุด (Contextual Memory) ไปยัง Llama 4 Vision ผ่าน Groq API เพื่อให้โมเดลสามารถตอบคำถามต่อเนื่องได้อย่างสอดคล้องกับบริบท

(5) เมื่อโมเดลประมวลผลเสร็จสิ้น ผลลัพธ์ที่ได้เป็นภาษาไทยจะถูกแสดงบนหน้าจอเพื่อให้โปรแกรมอ่านหน้าจอ (Screen Reader) อ่านออกเสียงให้ผู้ใช้รับฟังได้ทันที

การออกแบบระบบเป็น 2 ขั้นตอนนี้ช่วยให้การประมวลผลมีประสิทธิภาพสูงสุด กล่าวคือ COCO-SSD ทำหน้าที่เป็นตัวกรองเบื้องต้นบนอุปกรณ์เพื่อลดภาระการส่งข้อมูลไปยังคลาวด์ ในขณะที่ Llama 4 Vision ทำหน้าที่วิเคราะห์เชิงลึกเฉพาะเมื่อภาพอยู่ในตำแหน่งที่เหมาะสมแล้วเท่านั้น ส่งผลให้ระบบโดยรวมมีความหน่วงต่ำและประหยัดทรัพยากรได้อย่างมีประสิทธิภาพ

3.3.2.4 ระบบกลไกหลักเพื่อติดต่อเซิร์ฟเวอร์แบบต่อเนื่อง (Communication Layer & Presence Signaling) เพื่อการควบคุมที่เรียลไทม์ องค์ประกอบนี้ใช้ Pusher (WebSockets Signaling) ทำหน้าที่ตรวจสอบสถานะความพร้อมและการมีอยู่ระบบคิว (Presence)

(1) รับส่งสถานะ Online / Offline สู่ระบบโดยตรง

(2) สื่อคำร้องเพื่อแจ้งเตือนอาสาสมัครและเปิดรับการกระทำคำสั่งระยะไกลแบบ Event Broadcasting ทันทีเพื่อความปลอดภัยในผู้ใช้

3.3.3 ลำดับการทำงานของระบบ (System Workflow)

กระบวนการทำงานของระบบสามารถแบ่งออกเป็น 2 รูปแบบหลัก ตามที่ผู้ใช้เลือกใช้งานผ่าน PWA ดังนี้

รูปแบบที่ 1 : กระบวนการทำงานของระบบผู้ช่วยอัจฉริยะ (AI Visual Assistant Flow)

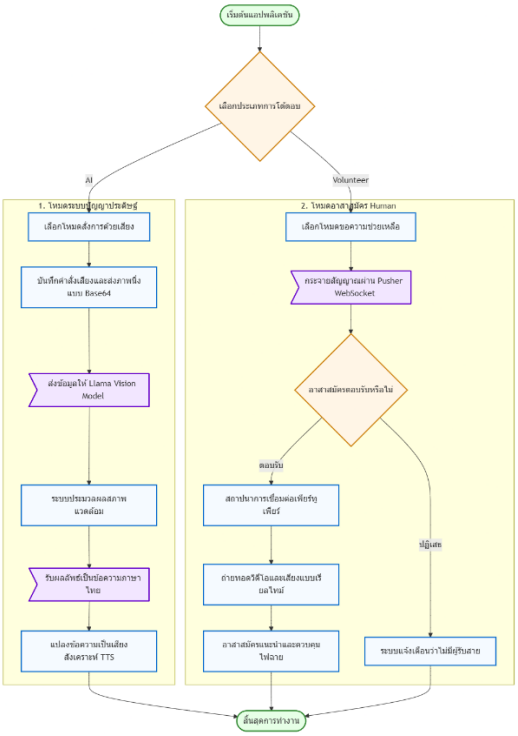
กระบวนการทำงานของระบบผู้ช่วยอัจฉริยะเริ่มต้นเมื่อผู้ใช้เปิดใช้งานเว็บแอปพลิเคชัน (PWA) บนสมาร์ตโฟนและเลือกเข้าสู่โหมดผู้ช่วย AI ผู้ใช้สามารถสั่งงานด้วยเสียงเพื่อสอบถามเกี่ยวกับวัตถุหรือสภาพแวดล้อมที่อยู่ตรงหน้า เช่น การใช้คำสั่ง “Ask the Scene” เพื่อให้ระบบอธิบายภาพหรือวิเคราะห์สิ่งของในบริเวณนั้นระบบจะดำเนินการเป็น 2 ขั้นตอนต่อเนื่องกัน โดยขั้นตอนแรกระบบรับภาพวิดีโอจากกล้องหลังผ่าน getUserMedia API และประมวลผลด้วยโมเดล COCO-SSD ผ่าน TensorFlow.js ในอัตรา 10 เฟรมต่อวินาที เพื่อนำทางกล้องแบบเรียลไทม์ด้วยเสียง เช่น ซ้าย ขวา บน หรือล่าง จนกว่าวัตถุจะเข้าสู่ Tolerance Zone 20% ของจุดกึ่งกลางหน้าจอ จึงแจ้งผู้ใช้ว่าพร้อมสำหรับขั้นตอนถัดไป จากนั้นขั้นตอนที่สองระบบจะแปลงเสียงคำสั่งเป็นข้อความด้วย Web Speech API และแปลงภาพเป็น Base64 JPEG ผ่าน Canvas แล้วส่งภาพ คำถาม และประวัติการสนทนาสูงสุด 6 ข้อความล่าสุด ไปยัง Llama 4 Vision ผ่าน Groq API เพื่อวิเคราะห์บริบทของภาพและประมวลผลคำตอบที่เหมาะสม หลังจากโมเดลประมวลผลเสร็จสิ้นระบบจะรับผลลัพธ์ในรูปแบบข้อความ และแปลงเป็นเสียงสังเคราะห์ด้วยเทคโนโลยี Text-to-Speech เพื่อส่งกลับให้

ผู้ใช้รับฟังแบบเรียลไทม์ กระบวนการทั้งหมดถูกออกแบบให้มีความหน่วงเวลาต่ำ (Low Latency) เพื่อให้ผู้ใช้งานได้รับข้อมูลอย่างรวดเร็วและต่อเนื่อง

รูปแบบที่ 2: กระบวนการทำงานของการขอความช่วยเหลือจากอาสาสมัคร (Volunteer Network Flow)

ในกรณีที่ผู้ใช้ต้องการความช่วยเหลือเพิ่มเติม หรือระบบปัญญาประดิษฐ์ไม่สามารถให้คำตอบที่มีความเชื่อมั่นเพียงพอ ผู้ใช้สามารถกดปุ่มขอความช่วยเหลือด้วยการสัมผัสเพียงครั้งเดียว (One-Tap Calling) ผ่านส่วนติดต่อผู้ใช้ของแอปพลิเคชัน เมื่อมีการร้องขอความช่วยเหลือ ระบบจะทำการส่งสัญญาณแจ้งเตือนไปยังศูนย์เครือข่ายอาสาสมัครที่กำลังออนไลน์อยู่ โดยใช้กลไกการส่งสัญญาณ (Signaling) ผ่านบริการ Pusher เพื่อจัดการสถานะการเชื่อมต่อแบบเรียลไทม์ เมื่ออาสาสมัครตอบรับคำร้องขอ ระบบจะทำการเชื่อมต่อวิดีโอคอลแบบเรียลไทม์โดยใช้เทคโนโลยี WebRTC ในลักษณะ Peer-to-Peer ระหว่างผู้ใช้และอาสาสมัครโดยตรง ภายหลังการเชื่อมต่อสำเร็จ อาสาสมัครสามารถรับชมภาพสดจากกล้องของผู้ใช้ และให้คำแนะนำผ่านเสียงพูดโดยตรง อีกทั้งระบบยังรองรับการสั่งงานเพิ่มเติม เช่น การเปิดไฟฉายของอุปกรณ์ระยะไกล เพื่อเพิ่มความชัดเจนของภาพในสภาพแวดล้อมที่มีแสงน้อย ผู้ใช้จะได้รับข้อมูลและคำแนะนำจากอาสาสมัครผ่านระบบเสียงแบบเรียลไทม์ จนกว่าภารกิจจะเสร็จสิ้น กระบวนการดังกล่าวสะท้อนแนวคิด Human-in-the-Loop ซึ่งผสานการทำงานระหว่างปัญญาประดิษฐ์และมนุษย์ เพื่อเพิ่มความน่าเชื่อถือ ความปลอดภัย และความแม่นยำของระบบโดยรวม

ภาพที่ 3.8 ลำดับการทำงานของระบบ



จากภาพ 3.8 สามารถอธิบายขั้นตอนการดำเนินงานโดยสังเขปได้ดังนี้

(1) การเริ่มต้นแอปพลิเคชันและการเลือกประเภทการโต้ตอบ (โหมด AI หรือ โหมดอาสาสมัคร)

(2) การนำทางกล้องด้วย COCO-SSD เพื่อเล็งวัตถุเป้าหมาย (On-device)

(3) การวิเคราะห์เชิงลึกด้วย Llama 4 Vision และแปลงเป็นเสียงสังเคราะห์ (Cloud-based)

(4) การทำงานในโหมดอาสาสมัคร โดยการกระจายสัญญาณขอความช่วยเหลือผ่านระบบเครือข่าย

5) การเชื่อมต่อแบบเพียร์ทูเพียร์ เพื่อถ่ายทอดวิดีโอและเสียงแบบเรียลไทม์เมื่ออาสาสมัครตอบรับ

(6) การสิ้นสุดการทำงานของระบบ (รวมถึงการแจ้งเตือนกรณีไม่มีอาสาสมัครตอบรับ)

3.4.4 หลักการออกแบบเพื่อความปลอดภัยและการเข้าถึง

การออกแบบระบบผู้ช่วยอัจฉริยะในโครงงานนี้ ไม่ได้มุ่งเน้นเพียงการทำงานได้ถูกต้องเท่านั้น แต่ยังคำนึงถึงความปลอดภัย ความน่าเชื่อถือ ประสิทธิภาพ และความสามารถในการเข้าถึงของผู้พิการทางสายตา โดยมีรายละเอียดดังนี้

(1) ความปลอดภัย (Safety) ระบบถูกออกแบบให้มี Confidence Threshold เพื่อลดความเสี่ยงจากการให้ข้อมูลผิดพลาด ซึ่งอาจส่งผลกระทบต่อความปลอดภัยของผู้พิการทางสายตา โดยเฉพาะในสถานการณ์ที่เกี่ยวข้องกับยา อุปกรณ์ไฟฟ้า หรือสิ่งกีดขวางภายในบ้าน

(2) ความน่าเชื่อถือและความต่อเนื่องของการทำงาน (Reliability) ระบบถูกออกแบบให้รองรับการทำงานแบบ Hybrid Architecture ซึ่งผสานการทำงานระหว่าง AI และมนุษย์ในลักษณะ Human-in-the-Loop เพื่อเพิ่มความน่าเชื่อถือของผลลัพธ์ ในกรณีที่ AI ไม่สามารถวิเคราะห์ได้อย่างมั่นใจ ระบบเครือข่ายอาสาสมัครจะทำหน้าที่เป็นกลไกสำรอง (Fallback Mechanism) ช่วยให้ผู้ใช้ยังคงได้รับความช่วยเหลืออย่างต่อเนื่องโดยไม่เกิดการหยุดชะงักของบริการ

(3) ประสิทธิภาพและความหน่วงเวลา (Performance & Latency Considerations) เนื่องจากระบบทำงานแบบเรียลไทม์ การออกแบบจึงมุ่งลดระยะเวลาหน่วง (Latency) ตั้งแต่ขั้นตอนการจับภาพ การประมวลผลด้วยโมเดล AI จนถึงการตอบกลับผ่านเสียงสังเคราะห์ การเลือกใช้เทคโนโลยีที่รองรับการประมวลผลรวดเร็ว เช่น การเรียกใช้โมเดลผ่าน API ที่มีประสิทธิภาพ และการเชื่อมต่อวิดีโอคอลแบบ Peer-to-Peer ด้วย WebRTC ช่วยลดภาระของเซิร์ฟเวอร์กลาง และเพิ่มความเร็วในการตอบสนองของระบบ

(4) การออกแบบเพื่อการเข้าถึง (Accessibility Design) ระบบถูกพัฒนาในรูปแบบ Progressive Web Application (PWA) และออกแบบตามหลักการ Accessibility Design เพื่อให้สามารถทำงานร่วมกับโปรแกรมอ่านหน้าจอ (TalkBack และ VoiceOver) ได้อย่างมีประสิทธิภาพ อินเทอร์เฟซถูกออกแบบให้มีปุ่มขนาดใหญ่ ใช้คำสั่งเสียงเป็นหลัก และลดขั้นตอนการโต้ตอบที่ซับซ้อน เพื่อให้เหมาะสมกับผู้พิการทางสายตาโดยเฉพาะ

(5) ความสามารถในการรองรับการขยายระบบ (Scalability) ในส่วนของระบบเครือข่ายอาสาสมัครได้มีการออกแบบกลไก Signaling Server แยกจากการสื่อสารแบบ Peer-to-Peer เพื่อรองรับการเชื่อมต่อหลายคู่พร้อมกันในอนาคต แนวทางดังกล่าวช่วยลดภาระของเซิร์ฟเวอร์กลาง และเพิ่มความสามารถในการขยายระบบเมื่อมีจำนวนผู้ใช้งานเพิ่มขึ้น

3.4 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในโครงงานนี้มุ่งเน้นการประเมินประสิทธิภาพของระบบในเชิงเทคนิค โดยผู้ดำเนินโครงงานเป็นผู้ทดสอบระบบด้วยตนเองในสภาพแวดล้อมภายในบ้าน (Indoor Environment) แบ่งการเก็บข้อมูลออกเป็น 3 ส่วน ดังนี้

3.4.1 การเก็บข้อมูลการจำแนกวัตถุ

ผู้ดำเนินโครงงานนำวัตถุพื้นฐานภายในบ้านจำนวน 10 รายการมาทดสอบกับระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization) โดยทดสอบรายการละ 10 ครั้ง ทั้งในสภาพแสงปกติและสภาพแสงน้อย รวมทั้งสิ้น 200 ครั้ง และบันทึกข้อมูลในแต่ละครั้ง ได้แก่ ชื่อวัตถุที่ใช้ทดสอบ ผลลัพธ์ที่ระบบแจ้ง ค่าความเชื่อมั่น (Confidence Score) และสภาวะแสงในขณะทดสอบ

3.4.2 การเก็บข้อมูลการระบุตำแหน่งและระยะเวลาค้นหาวัตถุ

ผู้ดำเนินโครงงานวางวัตถุทดสอบทั้ง 10 รายการ (วัตถุทดสอบใน 3.5.1) ในตำแหน่งที่แตกต่างกัน 5 ตำแหน่ง ได้แก่ ซ้าย ขวา บน ล่าง และกึ่งกลาง โดยทดสอบตำแหน่งละ 5 ครั้งต่อรายการ รวมทั้งสิ้น 250 ครั้ง และบันทึกว่าคำแนะนำทิศทางที่ระบบส่งออกมาสอดคล้องกับตำแหน่งจริงหรือไม่ พร้อมทั้งบันทึกระยะเวลาตั้งแต่เปิดกล้องจนระบบยืนยันว่าวัตถุเข้าสู่ Tolerance Zone ในแต่ละครั้ง

3.4.3 การเก็บข้อมูลระบบเครือข่ายอาสาสมัคร

ผู้ดำเนินโครงงานแบ่งบทบาทกันทดสอบ โดยคนหนึ่งทำหน้าที่เป็นผู้ใช้งานและอีกคนทำหน้าที่เป็นอาสาสมัคร จากนั้นบันทึกระยะเวลาตั้งแต่ผู้ใช้กดขอความช่วยเหลือจนกระทั่งอาสาสมัครตอบรับการเชื่อมต่อ ทดสอบทั้งหมด 10 ครั้ง

3.5 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในโครงงานครั้งนี้ดำเนินการโดยใช้สถิติเชิงพรรณนา (Descriptive Statistics) เป็นหลัก เพื่อสรุปและอธิบายผลการทดสอบประสิทธิภาพของระบบในแต่ละด้าน รวมถึงระดับความพึงพอใจของผู้พิการทางสายตาที่มีต่อการใช้งานระบบ โดยผู้ดำเนินโครงงานได้กำหนดแนวทางการวิเคราะห์ข้อมูลแยกตามประเภทของข้อมูลที่เก็บรวบรวมได้ ดังนี้

3.5.1 การวิเคราะห์ความแม่นยำในการจำแนกวัตถุ

การวิเคราะห์ประสิทธิภาพในการจำแนกวัตถุ (Object Classification Performance) ดำเนินการโดยนำข้อมูลจากแบบบันทึกผลการทดลองที่ได้จากการทดสอบระบบตรวจจับและระบุตำแหน่งวัตถุมาประเมินผ่านตารางความสับสน (Confusion Matrix) เพื่อคำนวณหาตัวชี้วัด 4 ประการ ได้แก่ ความแม่นยำ (Accuracy), ความเที่ยงตรง (Precision), ความระลึก (Recall) และค่าเอฟวัน (F1-Score) เกณฑ์การวิเคราะห์ ดังแสดงในตาราง 3.2

ตารางที่ 3.2

เกณฑ์การแปลผลค่าประสิทธิภาพของระบบ

ระดับประสิทธิภาพ	ค่า Accuracy	ค่า F1-Score	ความหมาย
ดีเยี่ยม	≥ 90%	≥ 0.90	ระบบมีความแม่นยำสูงมาก เหมาะสำหรับใช้งานจริง
ดี	80–89%	0.80–0.89	ระบบมีความแม่นยำอยู่ในเกณฑ์ที่ยอมรับได้
พอใช้	70–79%	0.70–0.79	ระบบต้องการการปรับปรุงเพิ่มเติม
ต้องปรับปรุง	< 70%	< 0.70	ระบบยังไม่เหมาะสำหรับใช้งานจริง

จากนั้นนำค่าความแม่นยำ (Accuracy) ที่ได้ไปเปรียบเทียบกับเกณฑ์สมมติฐานที่ตั้งไว้ว่าระบบต้องมีความแม่นยำไม่ต่ำกว่าร้อยละ 80 นอกจากนี้ผู้ดำเนินโครงงานยังนำค่า F1-Score มาวิเคราะห์ร่วมด้วยเพื่อประเมินความเสถียรของโมเดลและการลดข้อผิดพลาดเชิงลบ (เช่น การลดอัตรา False Positive เพื่อป้องกันการระบุชื่อวัตถุผิดพลาดซึ่งอาจเกิดอันตรายต่อผู้ใช้งาน) รวมทั้งวิเคราะห์ประสิทธิภาพแยกตามหมวดหมู่ของวัตถุ เพื่อระบุประเภทวัตถุที่ระบบมีข้อจำกัดในการจำแนก และนำไปสู่การอภิปรายถึงแนวทางการพัฒนาในบทต่อไป

3.5.2 การวิเคราะห์ความถูกต้องในการระบุตำแหน่งวัตถุ

การวิเคราะห์ความถูกต้องในการระบุตำแหน่งวัตถุ (Localization Accuracy) ดำเนินการโดยพิจารณาความสอดคล้องระหว่างทิศทางที่ระบบแนะนำกับตำแหน่งของวัตถุจริงในแต่ละครั้งที่ระบบส่งสัญญาณเสียง โดยบันทึกว่าคำแนะนำทิศทาง ไม่ว่าจะเป็นซ้าย ขวา บน หรือล่าง มีความสอดคล้องกับตำแหน่งวัตถุจริงในขณะนั้นหรือไม่ จากนั้นคำนวณเป็นค่าร้อยละของจำนวนครั้งที่ทิศทางถูกต้องต่อจำนวนครั้งที่ระบบแนะนำทิศทางทั้งหมด และนำผลที่ได้ไปประกอบการตัดสินว่าระบบสามารถนำทางผู้ใช้ให้เข้าถึงวัตถุได้อย่างแม่นยำในระดับที่ยอมรับได้หรือไม่

3.5.3 การวิเคราะห์ระยะเวลาเฉลี่ยในการค้นหาวัตถุ

ระยะเวลาเฉลี่ยในการค้นหาวัตถุ (Average Search Time) คำนวณจากข้อมูลที่บันทึกตั้งแต่เริ่มเปิดกล้องในแต่ละรายการจนกระทั่งระบบยืนยันว่าพบวัตถุเป้าหมายสำเร็จ โดยใช้ค่าเฉลี่ยเลขคณิต (Mean) ของระยะเวลาทั้งหมดจากการทดสอบทุกรายการ รวมถึงคำนวณค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เพื่อแสดงให้เห็นถึงความสม่ำเสมอของประสิทธิภาพระบบ นอกจากนี้ผู้ดำเนินโครงงานยังวิเคราะห์ระยะเวลาแยกตามประเภทของวัตถุ เพื่อตรวจสอบว่าวัตถุบางประเภทใช้เวลาในการค้นหานานกว่าค่าเฉลี่ยอย่างมีนัยสำคัญหรือไม่

3.5.4 การวิเคราะห์ประสิทธิภาพของระบบเครือข่ายอาสาสมัคร

สำหรับการวิเคราะห์ประสิทธิภาพของระบบเครือข่ายอาสาสมัคร ผู้ดำเนินโครงงานนำข้อมูลระยะเวลาที่บันทึกไว้จากการทดสอบ 10 ครั้งมาคำนวณค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของ

ระยะเวลาตั้งแต่ผู้ใช้กดขอความช่วยเหลือจนกระทั่งอาสาสมัครตอบรับการเชื่อมต่อ เพื่อประเมินว่าระบบสามารถทำหน้าที่เป็น Fallback System ที่มีประสิทธิภาพและตอบสนองได้รวดเร็วเพียงพอสำหรับการใช้งานจริงหรือไม่

3.5.5 การเปรียบเทียบผลกับสมมติฐาน

ในขั้นตอนสุดท้ายของการวิเคราะห์ข้อมูล ผู้ดำเนินโครงงานนำผลการวิเคราะห์จากทุกด้านมาเปรียบเทียบกับสมมติฐานของโครงงานทั้ง 2 ข้อ ได้แก่ ประการแรก ระบบต้องมีความแม่นยำในการจำแนกวัตถุไม่ต่ำกว่าร้อยละ 80 และประการที่สอง ระบบตรวจจับและระบุตำแหน่งวัตถุต้องสามารถให้คำแนะนำทิศทางแบบเรียลไทม์ได้อย่างถูกต้องในระดับที่ยอมรับได้ ผลการเปรียบเทียบดังกล่าวจะถูกนำไปใช้เป็นพื้นฐานในการอภิปรายผลและสรุปข้อค้นพบของโครงงานในบทที่ 4 และ 5 ต่อไป

3.6 นิยามปฏิบัติการของตัวแปร

นิยามปฏิบัติการของตัวแปรในการทำโครงงานครั้งนี้กำหนดขึ้นเพื่อให้เกิดความชัดเจนในการวัดและประเมินผลตัวแปรต้นและตัวแปรตามที่ใช้ในการศึกษา โดยอ้างอิงจากกรอบแนวคิดโครงงานที่กำหนดไว้ในบทที่ 1 ดังนี้

3.6.1 ตัวแปรต้น

ตัวแปรต้นในการทำโครงงานครั้งนี้ คือ ระบบผู้ช่วยอัจฉริยะสำหรับจำแนกและระบุตำแหน่งสิ่งของในรูปแบบเว็บแอปพลิเคชันแบบก้าวหน้า (PWA) ซึ่งประกอบด้วยองค์ประกอบหลัก 3 ส่วน ดังนี้

3.6.1.1 ระบบผู้ช่วยการมองเห็นอัจฉริยะ (AI Visual Assistant) หมายถึง ส่วนของระบบที่ใช้โมเดล Llama 4 Vision ผ่าน Groq API ในการประมวลผลภาพถ่ายที่ได้จากกล้องสมาร์ตโฟน เพื่อบรรยายรายละเอียดเชิงบริบทของวัตถุในภาพ อ่านข้อความภาษาไทยบนฉลากสินค้าหรือบรรจุภัณฑ์ และยืนยันความถูกต้องของวัตถุหลังจากที่ผู้ใช้หยิบจับขึ้นมา โดยรองรับการสั่งงานด้วยเสียงภาษาไทยผ่านเทคโนโลยี Web Speech API และส่งผลลัพธ์กลับมาในรูปแบบเสียงสังเคราะห์ผ่าน Text-to-Speech โดยระบบกำหนด Priority Framework ใน System Prompt เพื่อควบคุมลำดับการวิเคราะห์ภาพ ได้แก่ การตรวจสอบสิ่งบดบังหน้ากล้อง การอ่านฉลากสินค้าภาษาไทย (OCR) การตรวจสอบความปลอดภัยของสภาพแวดล้อม และการบรรยายบริบททั่วไปของวัตถุ นอกจากนี้ระบบยังรองรับ Contextual Memory โดยเก็บประวัติการสนทนาสูงสุด 6 ข้อความล่าสุด เพื่อให้โมเดลสามารถตอบคำถามต่อเนื่องได้อย่างสอดคล้องกับบริบท

3.6.1.2 ระบบตรวจจับและระบุตำแหน่งวัตถุ (Object Detection & Localization) หมายถึง ส่วนของระบบที่ใช้โมเดล COCO-SSD (Common Objects in Context - Single Shot MultiBox Detector) ประเภท Convolutional Neural Network ที่ทำงานบนเบราว์เซอร์ผ่าน TensorFlow.js ในการตรวจจับวัตถุจากภาพวิดีโอแบบเรียลไทม์ผ่านกล้องสมาร์ตโฟน โดยระบบจะสร้างกรอบสี่เหลี่ยม (Bounding Box) ล้อมรอบวัตถุที่ตรวจพบ และคำนวณพิกัดศูนย์กลางของกรอบดังกล่าวเพื่อเปรียบเทียบกับจุดศูนย์กลางของหน้าจอ จากนั้นแปลงความแตกต่างของพิกัดเป็นคำแนะนำทิศทาง ได้แก่ ซ้าย ขวา บน หรือล่าง ผ่านเสียงสังเคราะห์แบบเรียลไทม์ เพื่อช่วยให้ผู้พิการทางสายตาสามารถเล็งกล้องให้ตรงกับตำแหน่งของวัตถุเป้าหมายได้ โดยระบบรับภาพจากกล้องหลังผ่าน getUserMedia API และประมวลผลในอัตรา 10 เฟรมต่อวินาที เพื่อรักษาสมดุลระหว่างความเร็วในการตอบสนองและการประหยัดพลังงานแบตเตอรี่ ระบบใช้สูตร Euclidean Distance ในการคำนวณระยะห่างระหว่างจุดกึ่งกลางของวัตถุกับจุดกึ่งกลางหน้าจอ

และกำหนด Tolerance Zone ไว้ที่ 20% ของจุดกึ่งกลางหน้าจอ เพื่อใช้เป็นเกณฑ์ในการแจ้งเตือนผู้ใช้ว่าวัตถุอยู่ในตำแหน่งที่เหมาะสมแล้ว

3.6.1.3 ระบบเครือข่ายอาสาสมัคร (Volunteer Network System) หมายถึง ส่วนของระบบที่ผู้ดำเนินโครงงานพัฒนาขึ้นเพื่อทำหน้าที่เป็นกลไกสำรอง (Fallback Mechanism) ในกรณีที่ระบบปัญญาประดิษฐ์ไม่สามารถจำแนกหรือบรรยายวัตถุได้อย่างมั่นใจ โดยระบบจะเปิดให้ผู้พิการทางสายตาสามารถส่งคำร้องขอความช่วยเหลือผ่านการสัมผัสเพียงครั้งเดียว ซึ่งระบบจะส่งสัญญาณแจ้งเตือนไปยังอาสาสมัครที่ลงทะเบียนและออนไลน์อยู่บน Volunteer Dashboard ผ่าน Pusher และเมื่ออาสาสมัครตอบรับ ระบบจะเชื่อมต่อวิดีโอคอลแบบ Peer-to-Peer ผ่านเทคโนโลยี WebRTC โดยใช้ PeerJS เป็นตัวกลาง เพื่อให้อาสาสมัครสามารถมองเห็นภาพจากกล้องของผู้ใช้และให้คำแนะนำได้ทันที

3.6.2 ตัวแปรตาม

ตัวแปรตามในการทำโครงงานครั้งนี้ ได้แก่ ประสิทธิภาพของระบบและความพึงพอใจของผู้ใช้งาน ดังนี้

(1) ความแม่นยำในการจำแนกวัตถุ (Object Classification Accuracy) หมายถึง ค่าร้อยละของจำนวนครั้งที่ระบบสามารถระบุชนิดของวัตถุได้ถูกต้องตรงกับวัตถุจริงที่ผู้จัดทำโครงงานกำหนด เทียบกับจำนวนการทดสอบทั้งหมด โดยพิจารณาจากผลลัพธ์ที่ระบบแจ้งออกเสียงว่าตรงกับชื่อวัตถุที่กำหนดหรือไม่ วัดจากการทดสอบกับวัตถุ 10 รายการ โดยทดสอบรายการละ 10 ครั้ง

(2) ความถูกต้องในการระบุตำแหน่งวัตถุ (Localization Accuracy) หมายถึง ค่าร้อยละของจำนวนครั้งที่คำแนะนำทิศทางที่ระบบส่งออกมา ไม่ว่าจะเป็น ซ้าย ขวา บน หรือล่าง มีความสอดคล้องกับตำแหน่งของวัตถุจริงในภาพ ณ ขณะที่ระบบส่งสัญญาณเสียงนั้น โดยผู้จัดทำโครงงานทำหน้าที่ตรวจสอบและบันทึกความสอดคล้องดังกล่าวในแต่ละครั้งที่ระบบให้คำแนะนำ

(3) ระยะเวลาเฉลี่ยในการค้นหาวัตถุ (Average Search Time) หมายถึง ระยะเวลาเฉลี่ยที่ผู้จัดทำโครงงานได้ทดสอบใช้ ตั้งแต่เริ่มเปิดกล้องในโหมด Object Detection จนกระทั่งระบบยืนยันว่าพบวัตถุเป้าหมาย คำนวณจากค่าเฉลี่ยเลขคณิตของระยะเวลาทั้งหมดในการทดสอบแต่ละรายการ มีหน่วยเป็นวินาที

รายการอ้างอิง

กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ. (2550). *พระราชบัญญัติส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ พ.ศ. 2550*. https://dep.go.th

กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ. (2566). *สถิติคนพิการในประเทศไทย ประจำปี 2566*. https://dep.go.th/th/statistic

ราชกิจจานุเบกษา. (2550). *พระราชบัญญัติส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ พ.ศ. 2550*. สำนักเลขาธิการคณะรัฐมนตรี.

วรรณภา ศรีสุข และคณะ. (2565). การประยุกต์ใช้ Deep Learning เพื่อช่วยเหลือผู้พิการทางสายตาในการระบุสิ่งกีดขวางภายในอาคาร. *วารสารวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ*.

สุภาพร ทองดี และคณะ. (2566). การพัฒนาแอปพลิเคชันอ่านป้ายฉลากสินค้าด้วยเทคโนโลยี OCR สำหรับผู้พิการทางสายตา. *วารสารเทคโนโลยีสารสนเทศ*.

สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. (2561). *ยุทธศาสตร์ชาติ พ.ศ. 2561–2580*. https://www.nesdc.go.th/ewt_news.php?nid=9603

Ahmetovic, D., Gleason, C., Ruan, C., Kitani, K., Takagi, H., & Asakawa, C. (2020). NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. *ACM Transactions on Accessible Computing*, *13*(1), 1–29.

Apple. (2023). *Use VoiceOver on iPhone*. Apple Support. Retrieved March 20, 2026, from https://support.apple.com/guide/iphone/turn-on-and-practice-voiceover-iph3e2e415f/ios

Avila, M., Wolf, K., Brock, A., & Henze, N. (2016). Remote assistance for blind users in daily life: A survey about Be My Eyes. *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, Article 85, 1–2. https://doi.org/10.1145/2910674.2935839

Chaudary, B., Pohjolainen, S., Aziz, S., Arhippainen, L., & Pulli, P. (2023). Teleguidance-based remote navigation assistance for visually impaired and blind people—usability and user experience. *Virtual Reality*, *27*(1), 141–158. https://doi.org/10.1007/s10055-021-00536-z

Dandona, L., Dandona, R., Srinivas, M., Giridhar, P., McCarty, C. A., Rao, G. N., & Gilbert, C. E. (2002). Blindness in the Indian state of Andhra Pradesh. *Investigative Ophthalmology & Visual Science*, *43*(3), 755–761.

Google. (2023). *Get started on Android with TalkBack*. Google Accessibility Help. Retrieved March 20, 2026, from https://support.google.com/accessibility/android/answer/6283677

Groq, Inc. (2024). *Groq API documentation*. Retrieved March 20, 2026, from https://console.groq.com/docs

Islam, R. B., Akhter, S., Iqbal, F., Rahman, M. S. U., & Khan, R. (2023). Deep learning based object detection and surrounding environment description for visually impaired people. *Heliyon*, *9*(6), e16924. https://doi.org/10.1016/j.heliyon.2023.e16924

Jennings, C., Boström, H., & Bruaroey, J. (2021). *WebRTC 1.0: Real-time communication between browsers*. W3C Recommendation. https://www.w3.org/TR/webrtc/

Kacorri, H., Mascetti, S., Gerino, A., Picinali, L., Asakawa, C., & Kitani, K. (2017). Teachable object recognizer: A mobile application for blind people to create their own object recognizers. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 227–236.

Kumar, D. R., Thakkar, H. K., Merugu, S., Gunjan, V. K., & Gupta, S. K. (2022). Object detection system for visually impaired persons using smartphone. In A. Kumar, S. Senatore, & V. K. Gunjan (Eds.), *ICDSMLA 2020. Lecture Notes in Electrical Engineering* (Vol. 783). Springer. https://doi.org/10.1007/978-981-16-3690-5_154

Kuriakose, B., & Shajahan, P. A. (2020). Assistive tool for the visually impaired using computer vision and 3D audio feedback. *International Journal of Advanced Computer Science and Applications*, *11*(5), 1–8.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Massiceti, D., Sheratt, S., Sheratt, N., & Sheratt, H. (2021). ORBIT: A real-world few-shot dataset for teachable object recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10818–10827.

Meta AI. (2025). *The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation*. Retrieved March 20, 2026, from https://ai.meta.com/blog/llama-4-multimodal-intelligence/

PeerJS. (2024). *PeerJS documentation*. Retrieved March 20, 2026, from https://peerjs.com/docs/

Pusher Ltd. (2024). *Pusher channels documentation*. Retrieved March 20, 2026, from https://pusher.com/docs/channels

Roumeliotis, K. I., & Tselikas, N. D. (2022). Evaluating progressive web app accessibility for people with disabilities. *Network*, *2*(2), 350–369. https://doi.org/10.3390/network2020022

Shires, G., & Wennerstrom, H. (2023). *Web speech API specification*. W3C Community Group Final Report. Retrieved March 20, 2026, from https://wicg.github.io/speech-api/

Smilkov, D., Thorat, N., Assogba, Y., Yuan, A., Kreeger, N., Yu, P., Zhang, K., Cai, S., Nielsen, E., Soergel, D., Bileschi, S., Terry, M., Nicholson, C., Gupta, S. N., Sirajuddin, S., Sculley, D., Monga, R., Corrado, G., Viégas, F. B., & Wattenberg, M. (2019). TensorFlow.js: Machine learning for the web and beyond. *Proceedings of Machine Learning and Systems*, *1*, 309–321.

Tepsan, W., และคณะ. (2025). Application of generative AI for medication label reading for the elderly and visually impaired in Thailand. *Journal of Medical Informatics and Health Informatics*.

United Nations. (2006). *Convention on the rights of persons with disabilities*. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

Vercel. (2024). *Next.js documentation*. Retrieved March 20, 2026, from https://nextjs.org/docs

World Health Organization. (2019). *International classification of diseases 11th revision (ICD-11)*. https://icd.who.int/

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, 140–153. https://doi.org/10.1016/j.future.2022.05.014