



## รายงาน

### โครงการ Blind Blind

เทคโนโลยีช่วยเหลือผู้พิการทางสายตา

### จัดทำโดย

6510110092 นางสาวชไมพร แ้วไพรี

6510110311 นายพัทธดนย์ หนูดทอง

6510110330 นางสาวพริฮานา หมั่นเส้น

### เสนอ

ดร. อนันท์ ชกสุริวงศ์

โครงการนี้เป็นส่วนหนึ่งของ 241-202 MACHINE LEARNING 2

## บทนำ

การใช้ชีวิตโดยที่สายตามองไม่เห็นนับเป็นภาระที่มีความยากลำบากสำหรับผู้พิการทางสายตาหรือคนตาบอด ซึ่งต้องเผชิญกับความจำเป็นที่จะต้องพึ่งพาผู้อื่นหรือเทคโนโลยีต่าง ๆ เพื่อให้สามารถทำกิจกรรมประจำวันได้ตามปกติ ปัญหาที่เกิดขึ้นในชีวิตประจำวันของผู้พิการทางสายตาไม่เพียงแค่เรื่องการเดินทางหรือการทำงาน แต่ยังเกี่ยวข้องกับการสื่อสารและการเข้าถึงข้อมูลอย่างมีประสิทธิภาพ

โครงการ BLIND BLIND ถูกสร้างขึ้นเพื่อแก้ไขปัญหาดังกล่าวโดยการนำเสนอเทคโนโลยีประมวลผลสัญญาณเสียงและภาพ ร่วมกับปัญญาประดิษฐ์ (AI) เพื่อช่วยเหลือผู้พิการทางสายตาให้สามารถประสบความสำเร็จและมีคุณภาพชีวิตที่ดีขึ้น โดยโครงการนี้มุ่งเน้นการพัฒนาแอปพลิเคชันที่มีศักยภาพในการช่วยเหลือผู้พิการทางสายตาในการระบุวัตถุ อ่านข้อความ ระบุสี และตรวจจับความรู้สึกของผู้คนที่อยู่ตรงหน้า โดยใช้เทคโนโลยีต่าง ๆ ที่ทันสมัยและมีประสิทธิภาพ

โครงการนี้มุ่งหวังที่จะเป็นส่วนหนึ่งในการสร้างสังคมที่เปิดโอกาสและเป็นกันเองต่อผู้พิการทางสายตา และเชื่อว่าสามารถเป็นเครื่องมือที่มีประสิทธิภาพในการเพิ่มโอกาสและคุณภาพชีวิตให้ผู้พิการทางสายตา

## วัตถุประสงค์

1. เพื่อเสริมสร้างคุณภาพชีวิตของผู้พิการทางสายตาโดยการให้ความช่วยเหลือในการทำกิจกรรมที่ต้องการในชีวิตประจำวันอย่างรวดเร็วและสะดวกสบาย
2. เพื่อเพิ่มโอกาสในการเข้าถึงข้อมูลและการสื่อสารในสังคมของผู้พิการทางสายตา
3. เพื่อสร้างเครื่องมือและแอปพลิเคชันที่สามารถนำไปใช้ในสถานการณ์ต่าง ๆ เช่น การเดินทาง การอ่าน การศึกษา และกิจกรรมอื่น ๆ

## เครื่องมือที่ใช้ในโครงการ BLIND BLIND

### Object detection

- ANN : Artificial Neural Network หรือ โครงข่ายประสาทเทียม เป็นโมเดลทางคณิตศาสตร์ที่จำลองการทำงานของเซลล์ประสาทในสมองของมนุษย์ เพื่อใช้ในการแก้ปัญหาที่ต้องการความคล้ายคลึงกับการเรียนรู้และการปรับตัว โดยมักนำมาใช้ในการจำแนกประเภท การจำลองและการคาดการณ์ในหลายๆ ด้าน

### Emotion detection

- OpenCV : เป็นไลบรารีสำหรับคอมพิวเตอร์วิทัศน์ที่มีความสามารถในการประมวลผลภาพและวิดีโอ ในโครงการนี้ OpenCV ถูกใช้ในการตรวจจับวัตถุและการระบุสีของวัตถุในภาพ
- haarcascade : เป็น feature หนึ่งในไลบรารี OpenCV ที่ใช้ Detect แล้วก็ Recognize วัตถุต่างๆจาก Feature ที่มีอยู่ในรูปนั้นๆ
- CNN : โครงข่ายประสาทแบบคอนโวลูชัน เป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม bio-inspired โดยที่ CNN จะจำลองการมองเห็นของมนุษย์ที่มองพื้นที่เป็นที่ย่อย ๆ และนำกลุ่มของพื้นที่ย่อย ๆ มาผสมกันเพื่อดูว่าสิ่งที่เห็นอยู่คืออะไร

### Color detection

- OpenCV : เป็นไลบรารีสำหรับคอมพิวเตอร์วิทัศน์ที่มีความสามารถในการประมวลผลภาพและวิดีโอ ในโครงการนี้ OpenCV ถูกใช้ในการตรวจจับวัตถุและการระบุสีของวัตถุในภาพ

### Text detection

- Tesseract : เป็นเครื่องมือในการทำ OCR ซึ่งเป็นซอฟต์แวร์ฟรี ที่อยู่ภายใต้ใบอนุญาตของ Apache และได้รับการสนับสนุนจาก Google ตั้งแต่ปี 2006 โดย tesseract จะทำการดึงข้อความจากรูปภาพออกมาให้ในรูปแบบของ text ใช้ในการตรวจจับข้อความที่อ่านได้จากการแปลงภาพเป็น Gray Frame
- OCR : Optical Character Recognition ซึ่งเป็นการแปลงไฟล์เอกสารที่สแกนมา ให้เป็นไฟล์ข้อความโดยอัตโนมัติ

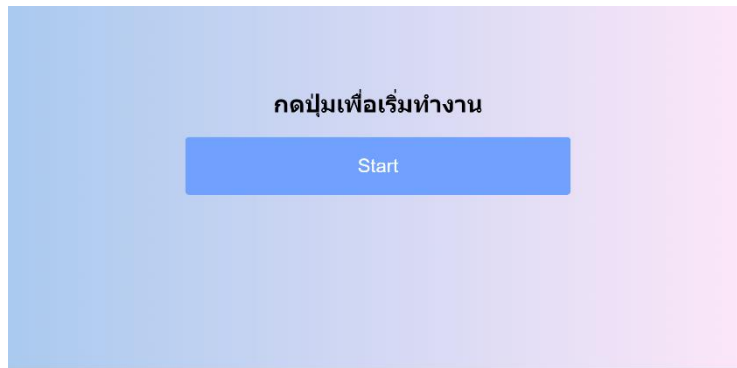
## Text-to-Speech

- gTTS (Google Text-to-Speech) : gTTS เป็นเครื่องมือที่ใช้ในการแปลงข้อความที่เป็นเสียง ในโครงการ BLIND BLIND คือ กระบวนการแปลงคำให้เป็นรูปแบบเสียงที่เปล่งออกมา โปรแกรมเครื่องมือหรือซอฟต์แวร์รับข้อความอินพุตจากผู้ใช้และการใช้วิธีการประมวลผลภาษาธรรมชาติเข้าใจภาษาศาสตร์ของภาษาที่ใช้และทำการอนุมานเชิงตรรกะกับข้อความ ข้อความที่ผ่านการประมวลผลนี้จะถูกส่งไปยังบล็อกถัดไปซึ่งจะดำเนินการประมวลผลสัญญาณดิจิทัลกับข้อความที่ประมวลผล การใช้อัลกอริธึมและการแปลงมากมายในที่สุดข้อความที่ประมวลผลนี้จะถูกแปลงเป็นรูปแบบเสียงพูด กระบวนการทั้งหมดนี้เกี่ยวข้องกับการสังเคราะห์เสียงพูด เครื่องมือนี้ถูกใช้ในการแจ้งเตือนผู้ช่วยเสียง

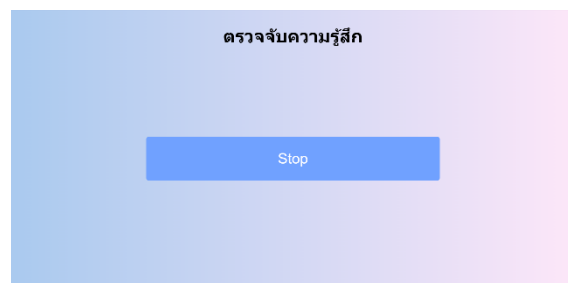
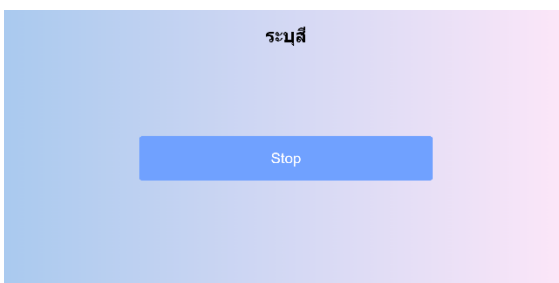
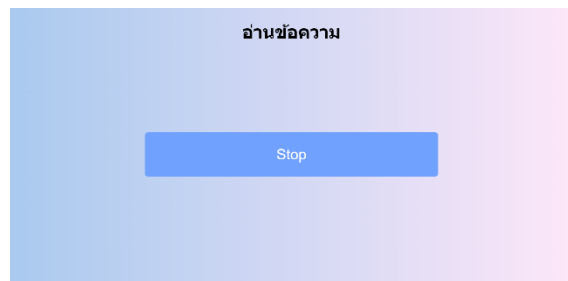
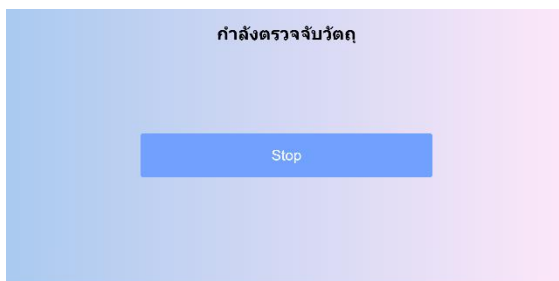
## Speech-to-Text

- Speech Recognition Library : Speech Recognition เป็นไลบรารีของ Python ที่ใช้ในการรับข้อมูลเสียงจากผู้ใช้และแปลงข้อมูลเสียงเป็นข้อความซึ่งใช้ในการรับคำสั่งเสียงจากผู้ใช้ในการเลือกหมวดการทำงาน โดยเราจะใช้ตัว Google Cloud Speech-to-Text ซึ่งเป็น Speech-to-Text ของ Google

## คู่มือการใช้งาน



- เมื่อรันโปรแกรม จะเข้าสู่หน้าการสั่งงานด้วยเสียง โดยผู้ใช้สามารถกดปุ่ม Start แล้วฟังเสียงแนะนำการใช้งาน จากนั้นหลังจากที่ได้ยินเสียงสัญญาณให้พูด สามารถสั่งการด้วยเสียงเพื่อเปิดใช้งานตัวเลือกที่ต้องการได้



- หลังจากสั่งการด้วยเสียงเพื่อเปิดการใช้งานตัวเลือกที่ต้องการแล้ว สามารถนำโทรศัพท์มือถือหรือกลับไปจ่อกับสิ่งที่ต้องการตรวจจับ แล้วจะเล่นเสียงสิ่งที่ตรวจจับได้ออกมา ไม่ว่าจะเป็น ข้อความ วัตถุ อารมณ์ของใบหน้า หรือสีที่ต้องการทราบ หากต้องการออกจากการทำงานนั้น สามารถกดปุ่ม Stop เพื่อกลับสู่หน้าแรกได้



หน้าแรกของการทำงานจะแสดงปุ่มให้เลือกการทำงานต่างๆที่ต้องการ

- การตรวจจับวัตถุ สามารถตรวจจับวัตถุต่างๆได้ แต่จะเป็นวัตถุทั่วไปในชีวิตประจำวัน เช่น โต๊ะ เก้าอี้ คน หน้าจอ โทรศัพท์ รถ เป็นต้น
- การอ่านข้อความ สามารถใช้เพื่ออ่านข้อความที่ต้องการได้ ไม่ว่าจะเป็นป้าย หรือเอกสารข้อความต่างๆ แต่ควรเป็นบทความที่มีขนาดความยาวไม่มากนักจึงจะทำงานได้ดี
- การระบุสี สามารถใช้ได้ทั้ง คนตาบอด และคนตาบอดสี เพื่อต้องการทราบสีต่างๆโดยสามารถใช้ในการเลือกสีของสิ่งของในชีวิตประจำวันได้ เช่น เสื้อผ้า
- ตรวจจับความรู้สึกบนใบหน้า สามารถตรวจจับความรู้สึกหน้าตาของคนตรงหน้าได้ ไม่ว่าจะเป็น ยิ้ม ร้องไห้ โกรธ เป็นต้น โดยฟีเจอร์นี้และที่กล่าวมาข้างต้นสามารถนำไปพัฒนาต่อให้ใช้ได้กับแว่นตาคนตาบอดแล้วเชื่อมต่อกับหูฟังในการใช้งานในอนาคต

## หลักการทำงาน

### Object Detection ตรวจจับวัตถุจากกล้อง

#### โมเดลและไลบรารีหลัก

- ใช้ โมเดล MobileNetSSD ในการตรวจจับวัตถุ

#### ฟังก์ชันการตรวจจับวัตถุ

- เรียกการทำงานของกล้องด้วย OpenCV แล้ววนลูปการอ่านภาพที่ได้จากกล้อง ปรับค่าที่ได้จากกล้อง
- สร้าง Blob จากภาพเพื่อนำเข้าเข้าสู่โมเดล MobileNet SSD แล้วทำการส่ง Blob เข้าไปในโมเดลเพื่อตรวจจับวัตถุ
- สร้าง Bounding Box สำหรับวัตถุที่ตรวจจับได้จากคลาสของโมเดล โดยใช้จุดกึ่งกลางของ Bounding Box ระบุตำแหน่งของวัตถุที่ตรวจจับได้ เพื่อระบุตำแหน่งที่ได้ให้ผู้ใช้ทราบ คือ ด้านหน้า ด้านซ้าย และด้านขวา แสดงผลลัพธ์ bounding box บนหน้าจอ

#### ฟังก์ชัน Text-to-Speech

- ใช้ Text-to-Speech (TTS) เพื่ออ่านข้อความของวัตถุที่ตรวจจับได้โดยเป็นภาษาไทย และจะอ่านทุกๆ 5 วินาที



## Text Detection อ่านข้อความจากกล้อง

### ไลบรารีหลัก

- Tesseract-OCR (pytesseract) ใช้ในการตรวจจับข้อความจากกล้อง
- gtts (Text-to-Speech) สำหรับเสียงอ่านข้อความ

### ฟังก์ชันประมวลผลข้อความที่ตรวจจับได้

- ทำการอ่าน Frame จากกล้อง จากนั้นแปลงภาพที่ได้เป็น Gray Frame เพื่อใช้ในการตรวจจับข้อความ
- อ่านข้อความจากภาพด้วย Tesseract-OCR (text) ภาษาไทย+อังกฤษ
- ถ้าตรวจจับข้อความได้จะทำการแสดงข้อความที่อ่านได้ และส่งไปยังฟังก์ชัน text-to-Speech

### ฟังก์ชัน Text-to-Speech

- ใช้ Text-to-Speech (TTS) เพื่ออ่านข้อความที่ตรวจจับได้โดยเป็นภาษาไทย และจะอ่านทุกๆ 10 วินาที เพื่อให้ข้อความก่อนหน้าอ่านจบก่อน

## Color Detection ตรวจจับสี

### ไลบรารีหลัก

- ใช้ OpenCV ในการแยกสี HSV

### ฟังก์ชันประมวลผลสี

- ทำการอ่าน Frame จากกล้องแล้วแปลงภาพที่ได้ให้เป็น HSV เพื่อใช้ในการแยกสี
- จาก Frame ที่ทำการอ่านมาทำการหาขนาดความกว้าง ความสูง เพื่อใช้ในการกำหนดจุดกึ่งกลางเพื่อใช้เป็นจุดสำหรับตรวจจับสี โดยกำหนดเป็นจุดกึ่งกลางของภาพ
- จากนั้นดึงค่า Hue ของพิกเซลที่จุดศูนย์กลาง แล้วกำหนดสีจากค่า Hue เพื่อแยกแยะแต่ละสี
- จากนั้นแยกค่าสี BGR ของพิกเซลที่จุดศูนย์กลางภาพออกมาเก็บไว้ในตัวแปร

### ฟังก์ชัน Text-to-Speech

- ใช้ Text-to-Speech (TTS) เพื่ออ่านข้อความที่ตรวจจับสีออกมาจากการตั้งเงื่อนไข แบ่งช่วงสีที่ตรวจจับได้ โดยเป็นภาษาไทย และจะอ่านทุกๆ 5 วินาที

## Emotion Detection ตรวจจับความรู้สึก

โมเดล

- Keras (keras.models)
- Emotion\_model.h5
- Emotion\_model.json

### ฟังก์ชันประมวลผลความรู้สึก

- เริ่มจากการเก็บ data ในการ test และ train โดย data จะเป็นรูปภาพหน้าคนในแต่ละอารมณ์ต่าง ๆ ประกอบด้วย angry,disgust,fear,happy,neutral,sad,surprise
- ทำการ test และ train model โดยการ train model ค่าเราสามารถปรับได้จะมี epoch,step\_per\_epoch,validation\_data,validation\_steps เมื่อเรา train เสร็จแล้วนั้น จะ save model เป็น model.json และ save ค่า weight ในไฟล์ .h5
- ต่อมาทำการ Evaluate model ของเรา และทำการ prediction
- จากนั้นก็ทำการ test emotion detection ของเราได้เลย

### ฟังก์ชัน Text-to-Speech

- ใช้ Text-to-Speech (TTS) เพื่ออ่านข้อความที่ตรวจจับสื่อกมาจากการตั้งเงื่อนไขแบ่งช่วงสีที่ตรวจจับได้ โดยเป็นภาษาไทย และจะอ่านทุกๆ 5 วินาที

## การนำไปใช้

1. สามารถนำมาใช้ในชีวิตประจำวันของผู้พิการทางสายตาได้ เนื่องจากแต่ละตัวเลือก ที่ต้องการใช้งาน สามารถสั่งการทำงานได้ด้วยเสียง อีกทั้งยังมีการตอบออกมาเป็นเสียงเช่นกัน (Text To Speech, Speech To Text)
2. สามารถนำไปพัฒนาต่อเป็น เครื่องมือเทคโนโลยี ที่ใช้ได้ในชีวิตประจำวันได้ เช่น แว่นตาสำหรับคนที่มีความพิการทางสายตา
3. สามารถนำไปพัฒนาในการทำ Application สำหรับคนพิการทางสายตาโดยเฉพาะ ไม่ว่าจะใช้ในการสื่อสาร หรือใช้ในชีวิตประจำวันโดยสามารถเชื่อมต่อกับเทคโนโลยีอื่นได้ เช่น โทรศัพท์ นาฬิกา แว่นตา โดยมีการเชื่อมต่อกับหูฟังเพื่อให้ใช้งานได้ง่ายสะดวก และเกิดประโยชน์มากที่สุด
4. สามารถนำไปพัฒนา GPS สำหรับผู้พิการทางสายตา ที่ใช้ในการนำทางได้ จากการ train model ของ Object Detection โดยมีการเชื่อมต่อกับ Google Map เพื่อใช้ในการเดินทางได้สะดวกยิ่งขึ้น

## การนำไปปรับใช้

- สามารถนำเอาทุก Option การทำงานรวมไว้ในการทำงานเดียว โดยเมื่อเปิดกล่องออกมาสามารถที่จะตรวจจับวัตถุ อ่านข้อความ ตรวจจับสีและตรวจจับความรู้สึกได้พร้อมกัน โดยจะใช้เสียงในการออกคำสั่งในทุกขั้นตอน และไม่จำเป็นต้องออกไปส่งคำสั่งใหม่ทุกครั้ง เพื่อเป็นการง่ายต่อการใช้งานของผู้ใช้
- สามารถทำเป็น Application เพื่อให้สะดวกต่อผู้ใช้งาน โดยเมื่อเปิด Application สามารถทำงานได้พร้อมกัน โดยไม่ต้องเลือกตัวเลือกใด ๆ

## อ้างอิง

CHAITron, สร้างAI ที่เข้าใจและโต้ตอบจากเสียงพูด | Speech Recognition with Python ,

[https://www.youtube.com/watch?v=OAM\\_K6HJzA&t=424s](https://www.youtube.com/watch?v=OAM_K6HJzA&t=424s) , code

[https://github.com/CHAITron/Youtube\\_Tutorials/blob/master/jarvis.py](https://github.com/CHAITron/Youtube_Tutorials/blob/master/jarvis.py)

Navapat tongpubet , pytesseract-ocr , <https://github.com/StrixzIV/pytesseract-ocr?tab=readme-ov-file>

[Python 3]ติดตั้งและใช้งาน Tesseract OCR สำหรับ window เพื่อสกัดข้อความจากภาพ ,

<https://medium.com/@navapat.tpb/python-3>

prasertcbs , สอนไพธอน: สร้างเครื่องเสียงด้วย google text-to-speech (gTTS package) ,

<https://www.youtube.com/watch?v=SmdVSqSg0B0>

TNN ONLINE , Envision Glasses แว่นตา AI อัจฉริยะ เสมือนเป็นดวงตาให้แก่ผู้พิการทางสายตา ,

<https://www.tnnthailand.com/news/tech/108131/>

SiamPhone , OrCam MyEye 2.0 แว่นติดกล้อง AI ช่วยคนตาบอดใช้ชีวิตได้ง่ายขึ้น,

<https://www.siamphone.com/contents/news-34375.html>

DROIDSANS, Google Lookout : แอปช่วยเหลือผู้พิการทางสายตา ใช้ AI บอกสิ่งต่างๆ รอบตัวได้ด้วยการเคาะ

โทรศัพท์ , โดย BSthun , <https://droidsans.com/google-lookout-app-for-the-blind/>