

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย).....	ก
บทคัดย่อ (ภาษาอังกฤษ)	ค
กิตติกรรมประกาศ	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ญ
สารบัญรูป	ฐ

บทที่

1 บทนำ	1
1.1 ความสำคัญและที่มาของปัญหาการวิจัย	1
1.2 วัตถุประสงค์ของการวิจัย	3
1.3 สมมติฐานการวิจัย.....	3
1.4 ขอบเขตของการวิจัย	3
1.5 ข้อตกลงเบื้องต้น	4
1.6 ประโยชน์ที่คาดว่าจะได้รับ	5
1.7 คำอธิบายศัพท์	5
2 ปรัชญาบรรณกรรมและงานวิจัยที่เกี่ยวข้อง	6
2.1 การค้นคืนรูปภาพ	7
2.1.1 การค้นคืนรูปภาพด้วยข้อความ	7
2.1.2 การค้นคืนรูปภาพจากเนื้อหา	8
2.1.3 การค้นคืนรูปภาพเชิงความหมาย.....	10
2.2 เทคนิคการค้นคืนรูปภาพเชิงความหมาย	12
2.2.1 เทคนิคการค้นคืนรูปภาพหลายระดับ.....	12
2.2.2 เทคนิคการใช้แม่แบบเชิงความหมาย	13
2.2.3 เทคนิคการใช้ข้อความเพื่อเรียนรู้ความหมายของรูปภาพ	14
2.2.4 เทคนิคการตอบกลับของผู้ใช้.....	15

สารบัญ (ต่อ)

หน้า

2.2.5	เทคนิคการใช้ออนโทโลยี	16
2.2.6	เทคนิคการจำแนกข้อมูลด้วยการเรียนรู้ของเครื่อง	18
2.3	โครงข่ายประสาทเทียมเชิงลึก.....	19
2.4	สถาปัตยกรรม ViT	28
2.5	การเรียนรู้ด้วยตนเอง.....	30
2.6	ตัวแบบ CLIP ที่ถูกฝึกฝนล่วงหน้า	32
2.7	การวัดประสิทธิภาพการค้นคืนรูปภาพแบบไบนารี	35
2.7.1	มาตรวัดแบบไม่สนใจลำดับของผลลัพธ์	35
2.7.2	มาตรวัดความเกี่ยวข้องแบบให้คะแนน.....	37
2.8	งานวิจัยที่เกี่ยวข้อง	41
3	วิธีดำเนินการวิจัย	52
3.1	วิธีวิจัย.....	52
3.1.1	การศึกษาข้อมูลเบื้องต้น	53
3.1.2	การออกแบบ	53
3.1.3	การเขียนโปรแกรม	77
3.1.4	การทดสอบ.....	77
3.1.5	การนำไปใช้งาน	78
3.2	เครื่องมือที่ใช้ในการวิจัย	78
3.3	การเก็บรวบรวมข้อมูล	79
3.4	การวิเคราะห์ข้อมูล	83
4	ผลการวิจัยและการอภิปรายผล	84
4.1	ผลการพัฒนาแบบจำลองการค้นคืนรูปภาพดิจิทัลเชิงความหมาย โดยประยุกต์โครงข่ายประสาทเทียมเชิงลึกที่ถูกฝึกฝนล่วงหน้า.....	84
4.1.1	ผลการพัฒนามอดูลการสร้างชุดคำอธิบายรูปภาพ	86
4.1.2	ผลการพัฒนามอดูลการประมวลผลข้อความค้นหา	95
4.1.3	ผลการพัฒนามอดูลการจับคู่เวกเตอร์	96

สารบัญ (ต่อ)

หน้า

4.2 ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพดิจิทัลเชิงความหมาย	98
4.2.1 ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพดิจิทัลเชิงความหมาย ด้วยค่าความแม่นยำที่ k อันดับ ค่าความครบถ้วนที่ k อันดับ และ ค่าประสิทธิภาพโดยรวมที่ k อันดับ	102
4.2.2 ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพดิจิทัลเชิงความหมายด้วย มาตรวัดความเกี่ยวข้องแบบให้คะแนนด้วยค่า NDCG ที่ k อันดับ	117
5 สรุปและข้อเสนอแนะ	126
5.1 สรุปผลการวิจัย.....	126
5.1.1 สรุปผลการพัฒนาแบบจำลองการค้นคืนรูปภาพดิจิทัลเชิงความหมาย โดยประยุกต์โครงข่ายประสาทเทียมเชิงลึกที่ผูกฝึกฝนล่วงหน้า	126
5.1.2 สรุปผลการประเมินประสิทธิภาพการค้นคืนรูปภาพดิจิทัลเชิงความหมาย ด้วยค่าความแม่นยำที่ k อันดับ ค่าความครบถ้วนที่ k อันดับ และค่าประสิทธิภาพโดยรวมที่ k อันดับ.....	127
5.1.3 สรุปผลการประเมินประสิทธิภาพการค้นคืนรูปภาพดิจิทัลเชิงความหมาย โดยจำแนกจากกลุ่มผู้เชี่ยวชาญ.....	130
5.1.4 สรุปผลการประเมินประสิทธิภาพการค้นคืนรูปภาพดิจิทัลเชิงความหมาย ด้วยค่า NDCG เปรียบเทียบระหว่างการค้นคืนรูปภาพด้วยตัวแบบ CLIP ดั้งเดิม กับการค้นคืนรูปภาพด้วยตัวแบบ CLIP ที่ผ่านการปรับค่าน้ำหนัก.....	132
5.2 การประยุกต์ผลการวิจัย.....	137
5.3 ข้อเสนอแนะในการวิจัยครั้งต่อไป.....	138
รายการอ้างอิง.....	140
ภาคผนวก	
ภาคผนวก ก ตัวอย่างรูปภาพต่อข้อความบรรยายความหมายเชิงคุณภาพ	149
ภาคผนวก ข ผลการพยากรณ์ค่าบรรยายรูปภาพจากตัวแบบเปรียบเทียบกับ ผลการประเมินจากผู้เชี่ยวชาญ	154

สารบัญ (ต่อ)

หน้า

ภาคผนวก ค	ตัวอย่างการคำนวณน้ำหนักโครงข่ายประสาทเทียม เมื่อกำหนดค่าน้ำหนักเริ่มต้นเท่ากับ 2	163
ภาคผนวก ง	คำศัพท์จากการสุ่มและตัวอย่างข้อความค้นหา ที่สร้างจากคำศัพท์ด้วย ChatGPT.....	172
ภาคผนวก จ	ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพ 20 อันดับ	178
ภาคผนวก ฉ	ตัวอย่างแบบประเมินความถูกต้องการค้นคืนรูปภาพโดยผู้เชี่ยวชาญ	209
ประวัติผู้เขียน		226

สารบัญตาราง

ตารางที่	หน้า
2.1	ส่วนประกอบระบบประสาทตามหลักชีววิทยาเปรียบเทียบกับโครงข่ายประสาทเทียม.....20
2.2	รายละเอียดตัวแบบ ViT แต่ละเวอร์ชัน29
2.3	การทดสอบประสิทธิภาพตัวแบบที่ถูกฝึกฝนล่วงหน้าบนชุดข้อมูล imagenet21k และปรับแต่งอย่างละเอียดบนชุดข้อมูล imagenet2012 30
2.4	ผลลัพธ์ 5 รูปภาพจากการค้นคืนรูปภาพดิจิทัลที่ผ่านการประเมินความถูกต้อง.....35
2.5	การคำนวณค่าประสิทธิภาพโดยรวม จากค่าความแม่นยำและค่าความครบถ้วน 37
2.6	ตัวอย่างผลลัพธ์จำนวน 5 รูปภาพจากข้อความค้นหา “a tabby cat on the floor” ที่ผ่านการประเมินคะแนนความเกี่ยวข้อง38
2.7	การปฐมนิเทศงานวิจัยที่เกี่ยวข้องกับเทคนิคการค้นคืนรูปภาพเชิงความหมาย45
3.1	ตัวอย่างรูปภาพที่สุ่มจากชุดข้อมูล Flickr30k และชุดข้อมูลที่ผู้วิจัยรวบรวมเอง55
3.2	ตัวอย่างรูปภาพที่สุ่มจากชุดข้อมูล Flickr30k และชุดข้อมูลที่ผู้วิจัยรวบรวมเอง ต่อข้อความบรรยายรูปภาพที่อธิบายความหมายเชิงคุณภาพของรูปภาพ.....57
3.3	ผลการพยากรณ์ความน่าจะเป็นของป้ายกำกับภาษาธรรมชาติ63
3.4	ตัวอย่างการเปรียบเทียบผลการประเมินความหมายของรูปภาพจากผู้เชี่ยวชาญ ท่านที่ 1 กับผลการพยากรณ์ความน่าจะเป็นของป้ายกำกับภาษาธรรมชาติ65
3.5	ตัวอย่างรูปภาพจากชุดข้อมูลที่ผู้วิจัยรวบรวมเอง 70
3.6	ตัวอย่างข้อความค้นหาภายใต้ 3 เงื่อนไข ที่ถูกสร้างจาก ChatGPT81
3.7	คำศัพท์จากการสุ่มและตัวอย่างข้อความค้นหาที่สร้างจากคำศัพท์ด้วย ChatGPT82
3.8	การตีความผลการวิเคราะห์ข้อมูลในรูปของร้อยละ ตามระดับคะแนน 5 ระดับ83
4.1	การคำนวณค่าถ่วงน้ำหนักเฉลี่ยจากผลการประเมินความรูปภาพโดยผู้เชี่ยวชาญ จำนวน 10 ท่านสำหรับนำไปปรับค่าน้ำหนักแบบย้อนกลับ91
4.2	ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพ 20 อันดับ ที่ผ่านการประเมินความถูกต้อง.....100
4.3	ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพแบบไม่สนใจลำดับของผลลัพธ์..... 103
4.4	การเปรียบเทียบผลการวิจัยกับงานวิจัยอื่น ๆ ที่มุ่งค้นคืนรูปภาพเชิงความหมาย โดยใช้รูปภาพค้นหา105

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.5 ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพด้วยชื่อรูปภาพและหมวดหมู่ ในลักษณะ ป้ายกำกับของรูปภาพที่ผ่านการประเมินความถูกต้องโดยผู้เชี่ยวชาญท่านที่ 1	106
4.6 ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพด้วยชื่อรูปภาพและหมวดหมู่ ในลักษณะป้ายกำกับของรูปภาพโดยจำแนกจากกลุ่มผู้เชี่ยวชาญ	108
4.7 ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพด้วยข้อความค้นหาสั้น ๆ เกี่ยวกับ แนวคิดระดับสูงของรูปภาพ โดยผู้เชี่ยวชาญท่านที่ 1	109
4.8 ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพด้วยข้อความค้นหาสั้น ๆ เกี่ยวกับ แนวคิดระดับสูงของรูปภาพโดยจำแนกจากกลุ่มผู้เชี่ยวชาญ	111
4.9 ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพด้วยข้อความค้นหาที่อธิบายความหมาย เชิงคุณภาพของรูปภาพที่ผ่านการประเมินความถูกต้องโดยผู้เชี่ยวชาญท่านที่ 1	113
4.10 ผลการประเมินประสิทธิภาพการค้นคืนรูปภาพด้วยข้อความค้นหา ที่อธิบายความหมายเชิงคุณภาพของรูปภาพโดยจำแนกจากกลุ่มผู้เชี่ยวชาญ	115
4.11 ตัวอย่างผลการประเมินความถูกต้องของผลลัพธ์การค้นคืนรูปภาพด้วยแนวคิด ระดับสูงโดยผู้เชี่ยวชาญ	116
4.12 การเปรียบเทียบสภาพแวดล้อมระหว่างการค้นคืนรูปภาพด้วยตัวแบบ CLIP ดั้งเดิม กับการค้นคืนรูปภาพด้วยตัวแบบ CLIP ที่ผ่านการปรับค่าน้ำหนัก	117
4.13 ตัวอย่างผลลัพธ์การค้นคืนรูปภาพด้วยตัวแบบ CLIP ดั้งเดิมกับการค้นคืนรูปภาพ ด้วยตัวแบบ CLIP ที่ผ่านการปรับค่าน้ำหนัก บนชุดข้อมูล Flickr30k	119
4.14 ค่า NDCG ของการค้นคืนลำดับที่ 1, 3 และ 5 ระหว่างการค้นคืนรูปภาพด้วยตัวแบบ CLIP ดั้งเดิมกับการค้นคืนรูปภาพด้วยตัวแบบ CLIP ที่ผ่านการปรับค่าน้ำหนัก	122
4.15 ตัวอย่างผลการเปรียบเทียบผลลัพธ์จากข้อความค้นหาจำนวน 5 รูปภาพ	123
4.16 ตัวอย่างผลการเปรียบเทียบผลลัพธ์จากข้อความค้นหาแบบจำแนกรายละเอียดเชิงลึก ...	124
4.17 ตัวอย่างผลการเปรียบเทียบผลลัพธ์จากข้อความค้นหาแบบปฏิเสธ	125
5.1 ประโยชน์จากการพัฒนาแบบจำลองการค้นคืนรูปภาพดิจิทัลเชิงความหมาย โดยใช้โครงข่ายประสาทเทียมเชิงลึกที่ถูกฝึกฝนล่วงหน้า	133

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
ก.1 ตัวอย่างรูปภาพที่สุ่มจากชุดข้อมูล Flickr30k และชุดข้อมูลที่ผู้วิจัยรวบรวมเอง ต่อข้อความบรรยายรูปภาพที่อธิบายความหมายเชิงคุณภาพ	150
ข.1 ผลการพยากรณ์คำบรรยายรูปภาพจากตัวแบบเปรียบเทียบกับ ผลการประเมินจากผู้เชี่ยวชาญ จำนวน 10 ท่าน	155
ค.1 ตัวอย่างการคำนวณน้ำหนักโครงข่ายประสาทเทียม เมื่อกำหนดค่าน้ำหนักเริ่มต้นเท่ากับ 2	164
ง.1 คำศัพท์จากการสุ่มและตัวอย่างข้อความค้นหาที่สร้างจากคำศัพท์ด้วย ChatGPT	173
จ.1 ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพ 20 อันดับ ภายใต้เงื่อนไขข้อความค้นหา ด้วยชื่อรูปภาพและหมวดหมู่ ในลักษณะป้ายกำกับของรูปภาพ	179
จ.2 ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพ 20 อันดับ ภายใต้เงื่อนไขข้อความค้นหาสั้น ๆ เกี่ยวกับแนวคิดระดับสูงของรูปภาพ	189
จ.3 ตัวอย่างผลลัพธ์จากการค้นคืนรูปภาพ 20 อันดับ ภายใต้เงื่อนไข ข้อความค้นหา ที่อธิบายความหมายเชิงคุณภาพของรูปภาพ	199

สารบัญรูป

รูปที่	หน้า
2.1 การค้นคืนรูปภาพด้วยข้อความ	7
2.2 การค้นคืนรูปภาพจากเนื้อหา	8
2.3 ผลลัพธ์การค้นคืนรูปภาพจากเนื้อหาด้วยข้อความค้นหา “move forward” บนเว็บไซต์ google.com	9
2.4 ความสัมพันธ์ระหว่างแนวคิดระดับสูงที่เป็นรูปธรรมและที่เป็นนามธรรม.....	10
2.5 การค้นคืนรูปภาพเชิงความหมาย.....	11
2.6 ข้อมูลรูปภาพ 3 ระดับชั้น	12
2.7 การค้นคืนรูปภาพโดยใช้แม่แบบความหมายร่วมกับ WordNet	13
2.8 ผลลัพธ์เริ่มต้นต่อผู้ใช้สำหรับข้อความค้นหาเกี่ยวกับจักรยาน.....	15
2.9 ผลลัพธ์หลังจากปรับปรุงความถูกต้องจากการตอบกลับของผู้ใช้.....	16
2.10 ชุดออนโทโลยี แสดงคุณลักษณะระดับต่ำที่อธิบายสี ตำแหน่ง ขนาด และรูปร่าง.....	17
2.11 แสดงระบบประสาทตามหลักชีววิทยา (ซ้าย) และโครงข่ายประสาทเทียม (ขวา)	19
2.12 เพอร์เซปตรอนแบบไม่กำหนดไบแอส (ซ้าย) และแบบกำหนดไบแอส (ขวา)	20
2.13 โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า	23
2.14 ฟังก์ชันซอฟต์แวร์แมกซ์	24
2.15 การทำงานของฟังก์ชันสูญเสียในโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า.....	25
2.16 โครงข่ายประสาทเทียมแบบย้อนกลับ.....	26
2.17 การทำงานของสถาปัตยกรรม ViT	28
2.18 แนวคิดการเรียนรู้ด้วยตนเอง	31
2.19 กรอบการทำงานของตัวแบบ CLIP	33
3.1 วงจรการพัฒนาระบบแบบนำตกแบบย้อนกลับได้.....	52
3.2 วงล้ออารมณ์	54
3.3 แบบจำลองการค้นคืนรูปภาพเชิงความหมาย โดยฝึกฝนล่วงหน้าแบบคอนทราสต์.....	61
3.4 โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าที่ถูกกำหนดอินพุต คำนำหน้า และไบแอส.....	67

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.5 ตัวอย่างรูปภาพที่ผ่านการเรียนรู้ความคล้ายคลึงเชิงความหมาย	71
3.6 กระบวนการประมวลผลข้อความค้นหาด้วยตัวแบบภาษา DistilBERT	73
3.7 การพิจารณาระยะทางระหว่างเวกเตอร์รูปภาพและเวกเตอร์ข้อความค้นหา	76
3.8 การจับคู่เวกเตอร์ ระหว่างเวกเตอร์รูปภาพ และเวกเตอร์ข้อความค้นหา	76
3.9 แบบประเมินความถูกต้องการค้นคืนรูปภาพโดยผู้เชี่ยวชาญผ่าน Jotform	79
4.1 การออกแบบและพัฒนาแบบจำลองการค้นคืนรูปภาพดิจิทัลเชิงความหมาย โดยประยุกต์โครงข่ายประสาทเทียมเชิงลึกที่ถูกฝึกฝนล่วงหน้า	85
4.2 ผลการสร้างข้อความบรรยายรูปภาพในรูปแบบภาษาอังกฤษจากโมเดล LLaVA	86
4.3 ตัวอย่างการสร้างรูปภาพใหม่จากการดัดแปลงรูปภาพต้นฉบับ	87
4.4 การผสมผสานกันการสร้างรูปภาพใหม่ เพื่อให้ตัวเข้ารหัสเรียนรู้คุณลักษณะรูปภาพ	88
4.5 คำบรรยายรูปภาพที่ผ่านการทำความสะอาดข้อความ และการกำหนดมาตรฐานข้อความ	88
4.6 การเรียนรู้แบบคอนทราสต์ระหว่างเวกเตอร์คุณลักษณะข้อความ และเวกเตอร์คุณลักษณะรูปภาพบนพื้นที่การฝังหลายรูปแบบ	90
4.7 ความน่าจะเป็นของป้ายกำกับภาษาธรรมชาติ ที่กำหนดเป็นค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก	90
4.8 การปรับแต่งค่าน้ำหนักเองตามแนวคิดการแพร่กระจายย้อนกลับ	92
4.9 การเรียนรู้ข้ามชุดข้อมูลที่กำหนดเอง	93
4.10 รายละเอียดชุดคำอธิบายรูปภาพ	94
4.11 ผลการแปลงข้อความค้นหาเป็นเวกเตอร์	95
4.12 ผลลัพธ์การค้นคืนรูปภาพที่เกี่ยวข้องกับข้อความค้นหา “move forward” มากที่สุด	97
4.13 ภาพรวมแบบจำลองการค้นคืนรูปภาพดิจิทัลเชิงความหมายโดยใช้ โครงข่ายประสาทเทียมเชิงลึกที่ถูกฝึกฝนล่วงหน้า	98
4.14 ตัวอย่างการประเมินความถูกต้องของรูปภาพผ่านแบบประเมินออนไลน์ Jotform	99
4.15 ข้อความค้นหา ระหว่าง “photo of sunset” กับ “photo of sunrise”	109
4.16 ตัวอย่างรูปภาพที่มีความแปรปรวนที่พบในการจำแนกข้อมูล	112

สารบัญรูป (ต่อ)

รูปที่	หน้า
5.1 ประสิทธิภาพการค้นคืนรูปภาพดิจิทัล เมื่อค่าความแม่นยำที่ k อันดับ เท่ากับ 3, 5, 10, 15 และ 20 ตามลำดับ	128
5.2 ประสิทธิภาพการค้นคืนรูปภาพดิจิทัล เมื่อค่าความครบถ้วนที่ k อันดับ เท่ากับ 3, 5, 10, 15 และ 20 ตามลำดับ	129
5.3 ประสิทธิภาพการค้นคืนรูปภาพดิจิทัล เมื่อค่าประสิทธิภาพโดยรวมที่ k อันดับ เท่ากับ 3, 5, 10, 15 และ 20 ตามลำดับ	130