สรุปงานวิจัยที่เกี่ยวข้อง

การสร้างคำอธิบายรูปภาพและกราฟในเอกสารวิชาการ
และ Retrieval Augmented Generation (RAG)

Figure & Chart Captioning · RAG: Naïve to Agentic

ภาพรวม (Overview)

งานวิจัยที่รวบรวมไว้ครอบคลุมพัฒนาการของการสร้างคำอธิบายรูปภาพและกราฟในเอกสารวิชาการ ตั้งแต่การสร้างชุดข้อมูล (2021) จนถึงการใช้ Multimodal LLM แบบร่วมมือ (2025) รวมถึงเทคนิค Retrieval Augmented Generation (RAG) ที่เสริม LLM ด้วยความรู้จากแหล่งข้อมูลภายนอก โดยสามารถจัดกลุ่มได้เป็น 4 แนวทางหลัก: **(1)** การสร้างชุดข้อมูลขนาดใหญ่, **(2)** การแปลง plot เป็นข้อความเพื่อ reasoning, **(3)** การใช้ LLM แบบหลายตัวร่วมกัน, และ **(4)** RAG แบบ Naive จนถึง Agentic

---

Paper 1 — SciCap (2021)

ชื่อ: Generating Captions for Scientific Figures
Hsu, Giles, Huang · Pennsylvania State University · arXiv:2110.11624

งานนี้เสนอ **SciCap** ซึ่งเป็นชุดข้อมูลขนาดใหญ่สำหรับการสร้างคำอธิบาย รูปภาพในบทความวิชาการ โดยดึงข้อมูลจากบทความ arXiv สาขา Computer Science ที่ตีพิมพ์ระหว่างปี 2010–2020 รวมกว่า 290,000 บทความ ผ่านกระบวนการ pre-processing ได้แก่ การจำแนกประเภทรูปภาพ, การระบุ sub-figure, การ normalize ข้อความ และการเลือก caption text ทำให้ได้รูปภาพกว่า **2 ล้านรูป**

จุดเด่น:

- ชุดข้อมูลแรกในระดับ large-scale สำหรับ scientific figure captioning
- มีระบบจำแนก figure type — graph plot เป็น type หลัก (19.2%)
- สร้าง baseline model สำหรับ graph plot captioning
- เปิดเผยทั้งโอกาสและความท้าทายของงานนี้

Paper 2 — SciCap+ (2023)

ชื่อ: A Knowledge Augmented Dataset to Study the Challenges of Scientific Figure Captioning
arXiv:2306.03491

SciCap+ ต่อยอดจาก SciCap โดยเพิ่ม **knowledge augmentation** เข้าไปในชุดข้อมูล เพื่อศึกษาความท้าทายของการสร้างคำอธิบาย รูปภาพวิทยาศาสตร์ให้ลึกขึ้น งานนี้ขยายขอบเขตการเข้าใจเอกสารวิชาการ ออกไปเกินกว่าการวิเคราะห์ข้อความล้วน โดยรวมบริบทจากเนื้อหาของบทความเข้ามาด้วย

จุดเด่น:

- เพิ่ม knowledge context จากเนื้อหาบทความให้แต่ละรูปภาพ
- ศึกษาความท้าทายที่ชุดข้อมูลดั้งเดิม (SciCap) ยังไม่ครอบคลุม
- เหมาะสำหรับการพัฒนา context-aware captioning

Paper 3 — DePlot (2023)

ชื่อ: One-shot Visual Language Reasoning by Plot-to-Table Translation
Liu et al. · Google DeepMind · arXiv:2212.10505

DePlot นำเสนอแนวทาง few-shot สำหรับ visual language reasoning โดยแยกปัญหาออกเป็น 2 ขั้นตอน: **(1)** แปลง plot/chart เป็นตารางข้อความ (plot-to-table translation) และ **(2)** ใช้ LLM reasoning บนข้อความที่ได้ โมดูล DePlot ทำหน้าที่แปลง modality โดยเฉพาะ แล้วส่งผลลัพธ์ไปยัง LLM ใด ๆ แบบ plug-and-play

ผลลัพธ์:
- DePlot + LLM (one-shot) ดีกว่า SOTA ที่ fine-tune บนข้อมูลหลายพัน samples
- **+29.4%** บน ChartQA (human-written queries)
- ไม่ต้องการ fine-tuning บน downstream task — ใช้ได้ทันที

Paper 4 — MLBCAP (2025)

ชื่อ: Multi-LLM Collaborative Caption Generation in Scientific Documents
Kim, Lee, Choi, Hsu et al. · Teamreboott / Penn State / Adobe Research · arXiv:2501.02552

MLBCAP เสนอ framework ที่ใช้ LLM หลายตัวทำงานร่วมกัน เพื่อสร้าง caption สำหรับรูปภาพในเอกสารวิชาการ แก้ปัญหาของวิธีการเดิมที่ใช้ข้อมูลไม่ครบถ้วน และมองงาน captioning แบบ isolated โดยไม่คำนึงถึงบริบทรอบข้าง

จุดเด่น:
- ใช้ LLM หลายตัวร่วมมือกัน (collaborative multi-LLM)
- รวมข้อมูลที่หลากหลายเข้าด้วยกัน ไม่ใช่แค่ภาพเดียว
- ต่อยอดจากงานของ SciCap series (Hsu et al.)

Paper 5 — Survey: Chart Understanding with MLLMs (2025)

ชื่อ: Multimodal Information Fusion for Chart Understanding: A Survey of MLLMs
arXiv:2602.10138

งาน survey นี้รวบรวมพัฒนาการของ Multimodal LLM (MLLM) สำหรับ chart understanding โดยครอบคลุมการผสานข้อมูลหลาย modality (กราฟิกและข้อความ) เพื่อดึงความหมาย พร้อมวิเคราะห์ข้อจำกัด และแนวทางการปรับปรุงด้วย cognitive enhancement

ประเด็นที่ครอบคลุม:
- วิวัฒนาการของ MLLM สำหรับ chart understanding
- ข้อจำกัดของโมเดลปัจจุบัน
- แนวทาง cognitive enhancement เพื่อเพิ่มความสามารถในการ reasoning

---

Paper 6 — RAG: Naïve to Agentic (2026)

ชื่อ: RAG: Naïve to Agentic
Sirapakit Lim (Tony) · IPU, NECTEC · Knowledge Sharing #3 · เมษายน 2026

เอกสารนี้เป็น slide deck สำหรับ knowledge sharing ที่ครอบคลุมเทคนิค **Retrieval Augmented Generation (RAG)** ตั้งแต่รูปแบบพื้นฐาน จนถึงระดับ Agentic โดยมีเนื้อหา 4 ส่วนหลัก: **(1)** ข้อจำกัดของ LLM ที่ RAG แก้ไข (knowledge cutoff, hallucination, private data), **(2)** Basic RAG — retriever + knowledge base + augmented prompt, **(3)** RAG Improvement — chunking strategies, query parsing (HyDE, NER, rewriting), reranking (bi-encoders, cross-encoders, ColBERT), และ **(4)** Agentic RAG ที่ LLM ทำหน้าที่ตัดสินใจเลือกเครื่องมือเอง มีการกล่าวถึงการรองรับภาษาไทยด้วยการใช้ `pythainlp: attacut` สำหรับ tokenization ในขั้นตอน chunking

จุดเด่น:

- ครอบคลุม RAG ทั้งสาย: Basic → Improvement → Agentic
- เปรียบเทียบ Keyword Search (TF-IDF, BM25) กับ Semantic Search (Vector)
- กล่าวถึง Thai NLP — ใช้ `pythainlp: attacut` tokenize ก่อน chunk
- แนะนำ RAGAS library สำหรับ evaluation และ LangGraph สำหรับ Agentic RAG
- เหมาะเป็น foundation สำหรับ context-aware scientific figure retrieval

---

ตารางเปรียบเทียบ

Paper	ปี	แนวทาง	จุดเด่น
SciCap	2021	Dataset	2M+ figures จาก 290K papers
SciCap+	2023	Dataset+	Knowledge augmentation
DePlot	2023	Plot→Table→LLM	+29.4% บน ChartQA (one-shot)
MLBCAP	2025	Multi-LLM	Collaborative captioning
Survey	2025	Survey	MLLM evolution & limitations
RAG (Tony)	2026	RAG Tutorial	Naïve→Agentic, Thai NLP support

สรุปโดย Claude Sonnet 4.6 · May 23, 2026