การวิเคราะห์ช่องว่างงานวิจัย

Personalized Scientific Figure Caption Generation
with RAG and Multimodal LLMs

Research Gap Analysis · 17 Papers · สรุปครบชุด

จัดทำ 7 มิถุนายน 2569

ภาพรวมงานวิจัยที่เกี่ยวข้อง

งานวิจัยทั้ง 17 ชิ้นสามารถจัดกลุ่มได้เป็น **5 กลุ่มหลัก** ซึ่งล้วนมีส่วนเกี่ยวข้องกับโจทย์วิทยานิพนธ์เรื่อง *Personalized Scientific Figure Caption Generation using RAG + Multimodal LLMs*

กลุ่ม	ชื่อกลุ่ม	Papers	จำนวน
A	Scientific Figure Captioning (Core)	1, 2, 4, 8, 9, 10	6
B	Chart Understanding / Vision-Language Models	3, 5, 14, 15, 16, 17	6
C	RAG & Retrieval	6_{RAG}, 11	2
D	Evaluation Methods	7	1
E	Adjacent / Background	12, 13	2

สมการวิทยานิพนธ์: $\underbrace{\text{Chart VLM}}_{\text{Group B}} + \underbrace{\text{RAG context}}_{\text{Group C}} + \underbrace{\text{Personalization}}_{\text{Group A}} \xrightarrow{\text{งานเรา}}$ Personalized Scientific Figure Captio

ข้อสังเกตสำคัญ: ยังไม่มีงานใดที่รวมทั้ง 3 องค์ประกอบนี้เข้าด้วยกันในบริบท scientific figures — นี่คือ research gap หลักของวิทยานิพนธ์

รายละเอียดแต่ละ Paper

กลุ่ม A — Scientific Figure Captioning (Core)

Paper #1 — SciCap (2021) Hsu, Giles, Huang · Penn State · arXiv:2110.11624 · EMNLP Findings 2021

ปัญหาที่แก้: ขาดชุดข้อมูลขนาดใหญ่สำหรับ scientific figure captioning
แนวทาง: ดึงรูปภาพและ caption จาก 290,000+ arXiv CS papers (2010–2020) ผ่าน LaTeX source
ผลลัพธ์: ชุดข้อมูล 2+ ล้านรูป; graph plot เป็น type หลัก (19.2%); baseline model ด้วย vision encoder + text decoder
ข้อจำกัด: Caption สั้น (เฉลี่ย 51 คำ), ไม่มีบริบทจากเนื้อหาบทความ, ไม่รองรับ personalization

Paper #2 — SciCap+ (2023) arXiv:2306.03491

ปัญหาที่แก้: SciCap เดิมขาดบริบทจากเนื้อหาบทความ
แนวทาง: เพิ่ม knowledge augmentation — figure-mentioning paragraphs เข้าไปในชุดข้อมูล
ผลลัพธ์: ศึกษา context-aware captioning; แสดงว่าบริบทช่วยได้แต่ยังมีความท้าทาย
ข้อจำกัด: ยังไม่มี personalization, RAG หรือ retrieval; ใช้เฉพาะ text context ไม่ใช่ visual context

Paper #4 — MLBCAP (2025) Kim et al. · Teamreboott/PSU/Adobe · arXiv:2501.02552 · AAAI 2025 Workshop

ปัญหาที่แก้: single-LLM approach ใช้ข้อมูลไม่ครบและมองงานแบบ isolated
แนวทาง: Multi-LLM collaborative framework — หลายโมเดลทำงานร่วมกัน ใช้ข้อมูลหลาย modality
ผลลัพธ์: ชนะ SciCap Challenge; performance ดีกว่า single-LLM baselines
ข้อจำกัด: ไม่มี author-specific personalization; ไม่ใช้ retrieval; computational cost สูง

Paper #8 — LaMPCap (2025) Ng, Hsu, Ramakrishnan et al. · PSU/Adobe · arXiv:2506.06561

ปัญหาที่แก้: Author ต้องแก้ AI-generated caption ให้ match writing style ตัวเอง
แนวทาง: Dataset ใหม่: target figure + profile (□3 figures จากบทความเดียวกัน พร้อม image, caption, paragraphs)
ผลลัพธ์: Profile ช่วยสร้าง caption ใกล้เคียง author-written มากขึ้น; image ใน profile สำคัญกว่า text paragraphs
ข้อจำกัด: Profile จำกัดอยู่ในบทความเดียวกัน; ไม่ใช้ retrieved knowledge จากแหล่งภายนอก; **307,903 figures จาก 110,828 papers**

Paper #9 — AuthorContextSciCap (2025) Timklaypachara et al. · Mahidol University · arXiv:2510.07993

ปัญหาที่แก้: Generic captions ไม่ match writing style ของ author
แนวทาง: Two-stage pipeline: (1) DSPy MIPROv2/SIMBA optimize prompts per figure category + context filtering; (2) few-shot profile refinement
ผลลัพธ์: Category-specific prompts: ROUGE-1 recall +8.3%; Profile refinement: BLEU +40–48%, ROUGE +25–27%
ข้อจำกัด: ใช้ profile แค่ within-paper; ไม่มี external knowledge retrieval; ไม่มี visual understanding ของ scientific figures

Paper #10 — PersonalizedSciFigCap (2025) Kim, Lee, Choi, Jang · arXiv:2509.25817

ปัญหาที่แก้: Author-specific writing style transfer ใน scientific figure captioning
แนวทาง: Empirical study: ทดสอบหลาย strategy สำหรับ style transfer จาก author writing profile
ผลลัพธ์: Profile/context ช่วยให้ caption เหมือน author style มากขึ้น แต่ optimize เพื่อ style overlap อาจลด informativeness
ข้อจำกัด: Tradeoff ระหว่าง style fidelity กับ caption quality; ไม่รวม external retrieval

กลุ่ม B — Chart Understanding / Vision-Language Models

Paper #17 — Pix2Struct (2022) Lee, Joshi, Turc et al. · Google Research · arXiv:2210.03347 · ICML 2023

ปัญหาที่แก้: ไม่มี unified VLM ที่ handle visually-situated language ได้; fixed-resolution ทำให้ text ใน charts อ่านไม่ออก
แนวทาง: Pretrain ViT encoder + T5 decoder ด้วย screenshot-to-HTML; variable-resolution patches (□2048 patches, คง aspect ratio); prompt render บน image
ผลลัพธ์: SoTA บน 6/9 tasks — DocVQA 88.0 ANLS, AI2D 89.5%, ChartQA 56.0%
ข้อจำกัด: Web pretraining □ scientific charts; ไม่เข้าใจ math; image-only (ไม่รับ text context)

Paper #16 — MatCha (2022) Liu et al. · Google DeepMind · arXiv:2212.09662 · ACL 2023

ปัญหาที่แก้: Pix2Struct ไม่เข้าใจ mathematical/numerical relationships ใน charts
แนวทาง: เพิ่ม 2 objectives ใน finetuning บน Pix2Struct: (1) Math reasoning (arithmetic, counting); (2) Chart derendering (chart→table)
ผลลัพธ์: PlotQA +20%, ChartQA human +2.4%, ChartQA augmented +4.2% over prior SoTA
ข้อจำกัด: ยังเป็น image-only; ไม่มี chart summarization/captioning objective; scientific figures ยังไม่ครอบคลุม

Paper #15 — UniChart (2023) Masry et al. · George Mason University · arXiv:2305.14761 · EMNLP 2023

ปัญหาที่แก้: ขาดโมเดล chart-specific ที่ครอบคลุมทั้ง low-level และ high-level tasks
แนวทาง: Pretrain บน 611K charts ด้วย 4 objectives: (1) Chart-to-Table; (2) Open-ended QA; (3) Chart Summarization; (4) Chart-to-Caption (Chart2Text)
ผลลัพธ์: SoTA บน 4 benchmarks; 11× เร็วกว่า MatCha; ดีกว่า Pix2Struct/DePlot ทุก task
ข้อจำกัด: Corpus เป็น general charts ไม่ใช่ scientific figures; ไม่มี RAG; ไม่มี personalization

Paper #14 — ChartQA (2022) Masry et al. · arXiv:2203.10244 · ACL Findings 2022

ปัญหาที่แก้: Benchmarks เดิม (FigureQA, PlotQA) ใช้ synthetic charts และ template questions
แนวทาง: 9,608 คำถามจากมนุษย์จริง + 23,111 คำถาม augmented บน real-world charts จาก 4 แหล่ง; เพิ่ม visual + arithmetic reasoning
ผลลัพธ์: Baseline: VisionTaPas 45.5%, T5 22.4%; Human accuracy 69.8%; gap ยังกว้างมาก
ข้อจำกัด: QA-only (ไม่ใช่ captioning); charts จาก news/business ไม่ใช่ scientific papers

Paper #3 — DePlot (2023) Liu et al. · Google DeepMind · arXiv:2212.10505 · ACL Findings 2023

ปัญหาที่แก้: Visual reasoning บน charts ต้องการ fine-tuning บน data จำนวนมาก
แนวทาง: แยกเป็น 2 module: DePlot (chart→table, fine-tuned) + LLM (table→answer, one-shot)
ผลลัพธ์: +29.4% บน ChartQA human-written; ไม่ต้องการ downstream fine-tuning
ข้อจำกัด: Plot-to-table translation อาจสูญเสียข้อมูล visual; ไม่รวม caption generation

Paper #5 — Survey: MLLM for Chart Understanding (2025) arXiv:2602.10138

เนื้อหา: Survey วิวัฒนาการ MLLM สำหรับ chart understanding; ครอบคลุม multimodal information fusion, cognitive enhancement, ข้อจำกัดปัจจุบัน
ประโยชน์: ภาพรวมสถานะปัจจุบันของ chart understanding field; ระบุช่องว่างที่ MLLM ยังทำไม่ได้

กลุ่ม C — RAG & Retrieval

Paper #6$_{RAG}$ — RAGBeyondTheBuzz (2024) Zhao, Yang et al. · Microsoft Research Asia · arXiv:2409.14924

ปัญหาที่แก้: ขาด framework สำหรับเลือกเทคนิค RAG ให้เหมาะกับงาน
แนวทาง: จัดระดับ query เป็น 4 ประเภท: explicit fact, implicit fact, interpretable rationale, hidden rationale; แนะนำเทคนิคสำหรับแต่ละระดับ
ผลลัพธ์: Framework สำหรับออกแบบระบบ RAG; ครอบคลุม context/small-model/fine-tuning

approaches
ข้อจำกัด: เน้น text-only RAG; ไม่ครอบคลุม multimodal retrieval (image queries)

Paper #11 — MIND-RAG (2025) arXiv preprint · ICCVW 2025 (MRR Workshop)

ปัญหาที่แก้: RAG ทั่วไปไม่เข้าใจ multimodal context และ user intent ใน educational publications
แนวทาง: (1) Context-aware image summarization (ผสาน visual + text context); (2) Intent-aware reranking
ผลลัพธ์: 84.29% accuracy บน MEED-QA benchmark
ข้อจำกัด: ใช้ educational domain; ไม่ได้ทดสอบ scientific figure generation; ไม่มี personalization

กลุ่ม D — Evaluation Methods

Paper #7 — LLM-as-a-Judge (2023) Zheng et al. · UC Berkeley et al. · arXiv:2306.05685 · NeurIPS 2023

ปัญหาที่แก้: Human evaluation แพงและช้า; benchmarks เดิมไม่วัด open-ended quality
แนวทาง: MT-Bench (80 multi-turn questions, 8 categories) + Chatbot Arena (crowdsourced battles) ให้ GPT-4 เป็น judge
ผลลัพธ์: Strong LLM judge (GPT-4) ตรงกับ human preference >80%; เทียบเท่า inter-human agreement
ข้อจำกัด: Position bias, verbosity bias; ไม่ได้ออกแบบสำหรับ scientific figure evaluation

กลุ่ม E — Adjacent / Background

Paper #12 — NLIDBSensitive (2025) ITNG 2025

เนื้อหา: NL interface สำหรับ query database ที่มีข้อมูล sensitive ด้วย OCR + Neo4j + LLM Cypher translation
ความเกี่ยวข้อง: แสดง pattern ของการใช้ LLM เป็น intermediary ระหว่าง unstructured query กับ structured data — คล้ายกับ RAG สำหรับ scientific figures

Paper #13 — WebSearchViz (2010) IEEE Symposium on Web Society 2010

เนื้อหา: Scatter plot visualization สำหรับแสดงผลลัพธ์ web search ตาม modification date และ document type
ความเกี่ยวข้อง: Background สำหรับ information visualization; แนวคิด temporal + type-based result display

---

ตารางเปรียบเทียบ (17 Papers)

#	Paper	ปี	แนวทาง	จุดเด่น	ข้อจำกัด	SciFig	RAG	Personal
1	SciCap	2021	Large-scale dataset	2M+ figs, arXiv CS	ไม่มีบริบท, ไม่มี personalization	✓	—	—
2	SciCap+	2023	Knowledge augmentation	Context-aware dataset	Text-only context, ไม่มี retrieval	✓	—	—
3	DePlot	2023	Plot→Table→LLM	+29.4% ChartQA one-shot	QA-only, ไม่ใช่ captioning	—	—	—
4	MLBCAP	2025	Multi-LLM collaborative	ชนะ SciCap Challenge	ไม่มี personalization, ไม่มี retrieval	✓	—	—
5	MLLM Survey	2025	Survey	Chart MLLM evolution	—	—	—	—
6	RAGBeyondTheBuzz	2024	RAG Survey	4-level query taxonomy	Text-only, ไม่ใช่ multimodal	—	✓	—
7	LLM-as-Judge	2023	Evaluation framework	>80% human agreement	ไม่เฉพาะ scientific figs	—	—	—
8	LaMPCap	2025	Profile-based generation	307K figs, multimodal profile	Profile เฉพาะ within-paper	✓	—	✓
9	AuthorContextSciCa	2025	DSPy+profile refinement	BLEU +40–48%	ไม่มี external retrieval	✓	—	✓
10	PersonalizedSciFigC	2025	Style transfer study	Style fidelity analysis	Tradeoff: style vs. quality	✓	—	✓
11	MIND-RAG	2025	Multimodal RAG	84.29% MEED-QA	Educational domain, ไม่ใช่ sci-fig gen	—	✓	—
12	NLIDBSensitive	2025	NL-to-Cypher	Sensitive data safety	ไม่เกี่ยวกับ captioning	—	—	—
13	WebSearchViz	2010	Search visualization	Temporal scatter plot UI	เก่า, ไม่เกี่ยวกับ ML	—	—	—
14	ChartQA	2022	QA Benchmark	9.6K human questions	QA-only, ไม่ใช่ scientific figs	—	—	—
15	UniChart	2023	Universal chart VLM	4 objectives, 611K charts, 11× faster	General charts, ไม่มี RAG	—	—	—
16	MatCha	2022	Math+chart pretraining	+20% PlotQA vs. prior SoTA	Image-only, ไม่มี captioning obj.	—	—	—
17	Pix2Struct	2022	Screenshot pretraining	Variable-res, SoTA 6/9 tasks	Web domain → scientific charts	—	—	—
	งานของเรา	2025	**RAG+VLM+Profile**	**ครบทั้ง 3 dimension**	—	✓	✓	✓

SciF = Scientific Figure domain, **RAG** = ใช้ retrieval/external knowledge, **Personal** = มี personalization/author-specific

การวิเคราะห์ช่องว่างงานวิจัย (Gap Analysis)

Gap #1 — ไม่มีงานที่รวม Chart VLM + RAG + Personalization ในบริบท Scientific Figures

งานด้าน chart VLM (Pix2Struct, MatCha, UniChart) เน้น QA บน general/business charts งานด้าน scientific captioning (LaMPCap, MLBCAP) ไม่ใช้ retrieved external knowledge งานด้าน RAG (MIND-RAG, RAGBeyondTheBuzz) ไม่ได้ออกแบบสำหรับ figure generation → **โอกาสวิจัย: ระบบที่รับ scientific figure image + retrieved paper context → สร้าง caption ที่ match author style**

Gap #2 — Profile-based Personalization ยังจำกัดอยู่ใน Single Paper

LaMPCap (#8), AuthorContextSciCap (#9), PersonalizedSciFigCap (#10) ทั้งหมดใช้ profile เฉพาะ รูปภาพอื่นในบทความเดียวกัน (within-paper profile) ยังไม่มีงานใดที่ retrieve author's previous papers หรือ domain-specific figure patterns จาก external corpus **โอกาสวิจัย: Cross-paper author profile ด้วย RAG ที่ retrieve ผลงานที่ผ่านมาของ author คนเดียวกัน**

Gap #3 — Scientific Figure General Chart: ขาด Domain-Specific Pretraining

Pix2Struct pretrain บน web screenshots; MatCha/UniChart ใช้ general charts (Statista, Pew Research etc.) Scientific figures มีลักษณะต่างออกไป: axis labels ซับซ้อน, error bars, multi-panel, specialized notation ยังไม่มี VLM ที่ pretrain บน scientific figure corpus (arXiv/PubMed figures) โดยเฉพาะ **โอกาสวิจัย: Fine-tune chart VLM ด้วย scientific figure data หรือใช้ domain adaptation**

Gap #4 — ขาด Evaluation Framework สำหรับ Personalized Scientific Figure Captions

Metrics ปัจจุบัน (BLEU, ROUGE, BERTScore) วัด surface-level similarity ไม่ใช่ style fidelity หรือ scientific accuracy LLM-as-Judge (#7) ยังไม่ถูกนำมาใช้ใน scientific figure captioning evaluation PersonalizedSciFigCap (#10) ชี้ถึง tradeoff ระหว่าง style overlap กับ caption quality แต่ยังขาด formal evaluation framework **โอกาสวิจัย: Multi-dimensional evaluation: (1) scientific accuracy, (2) style fidelity, (3) informativeness**

Gap #5 — Multimodal RAG สำหรับ Scientific Documents ยังอยู่ในระยะเริ่มต้น

MIND-RAG (#11) แสดงให้เห็นว่า multimodal context-aware RAG ทำงานได้บน educational publications แต่ใช้ text QA เป็น downstream task ไม่ใช่ generation RAGBeyondTheBuzz (#6) ยังเป็น text-only framework; ไม่ครอบคลุม image-as-query scenario **โอกาสวิจัย: Multimodal RAG ที่รับ figure image เป็น query และ retrieve paper sections/related figures**

Gap Matrix — มิติที่ยังขาด

มิติ	LaMPCap	MLBCAP	UniChart	MIND-RAG	งานเรา
Scientific figure domain	✓	✓	—	—	✓
Chart visual understanding	—	partial	✓	—	✓
External RAG retrieval	—	—	—	✓	✓
Author personalization	✓	—	—	—	✓
Cross-paper profile	—	—	—	—	✓
Caption generation task	✓	✓	partial	—	✓
Scientific accuracy eval	—	—	—	—	✓

สรุป: ไม่มีงานใดในชุดนี้ที่ครอบคลุมทุกมิติ — งานวิทยานิพนธ์จะเป็นงานแรกที่รวม *scientific figure domain* + *chart visual understanding* + *external RAG* + *author personalization* ไว้ในระบบเดียวกัน

ตำแหน่งงานวิทยานิพนธ์

ด้าน	รายละเอียด
สิ่งที่ยืมมา	Variable-resolution VLM จาก Pix2Struct (#17); Chart pretraining objectives จาก MatCha (#16) หรือ UniChart (#15); Profile-based personalization framework จาก LaMPCap (#8); RAG taxonomy และ techniques จาก RAGBeyondTheBuzz (#6); LLM-as-Judge evaluation จาก LLMJudge (#7)
สิ่งที่แตกต่าง	เป้าหมายคือ *scientific figure captioning* (ไม่ใช่ QA); ใช้ RAG retrieve external paper context (ข้ามบทความ); รวม author profile + domain knowledge + visual understanding ในระบบเดียว
Novel contribution	(1) Multimodal RAG pipeline สำหรับ scientific figure captioning; (2) Cross-paper author style profile; (3) Evaluation framework สำหรับ personalized scientific captions

Evolution chain ของสาย Chart VLM:

$$\underbrace{\text{Pix2Struct}}_{\#17\ 2022} \rightarrow \underbrace{\text{MatCha}}_{\#16\ 2022} \rightarrow \underbrace{\text{UniChart}}_{\#15\ 2023} \rightarrow \underbrace{\text{DePlot}}_{\#3\ 2023} \rightarrow \underbrace{\textbf{งานเรา}}_{\text{+ RAG + Profile}}$$

Evolution chain ของสาย Personalized Captioning:

$$\underbrace{\text{SciCap}}_{\#1\ 2021} \rightarrow \underbrace{\text{SciCap+}}_{\#2\ 2023} \rightarrow \underbrace{\text{MLBCAP}}_{\#4\ 2025} \rightarrow \underbrace{\text{LaMPCap}}_{\#8\ 2025} \rightarrow \underbrace{\textbf{งานเรา}}_{\text{+ RAG + VLM}}$$

สรุปโดย Claude Sonnet 4.6 · 7 มิถุนายน 2569 · 17 papers ครบชุด