อวสานคอขวดงานวิจัย: สถาปัตยกรรม PaperOrchestra

ระบบ Multi-Agent อัจฉริยะเพื่อการเขียนเปเปอร์ AI ระดับตีพิมพ์อัตโนมัติ

NotebookLM

ช่องว่างมหาศาลระหว่าง "ผลการทดลอง" และ "เอกสารพร้อมตีพิมพ์"

ข้อมูลนำเข้า (The Input)

ไอเดีย
(Idea Summaries)

บันทึกการทดลอง
(Experimental Logs)

ข้อมูลดิบ
(Raw Metrics)

ปัญหา: การทำงานแบบเดี่ยว (Monolithic LLMs) นำไปสู่การอ้างอิงปลอม (Hallucinations), การสังเคราะห์ข้อมูลที่ตื้นเขิน, และไม่สามารถจัดรูปแบบ LaTeX ได้เอง

เป้าหมาย (The Output)

บทความพร้อมตีพิมพ์
(Submission-Ready)

ทบทวนวรรณกรรมเชิงลึก
(Rigorous Lit Review)

รูปแบบ LaTeX
(LaTeX Formatting)

Key Insight: โมเดลภาษาขนาดใหญ่ (LLMs) ส่วนมากถูกผูกติดกับวงจรการทดลองเฉพาะทาง ทำให้ไม่สามารถทำงานเป็น “ผู้ช่วยเขียนอิสระ” (Standalone Writer) ที่แปลงข้อมูลดิบให้เป็นเปเปอร์วิจัยที่สมบูรณ์ได้

NotebookLM

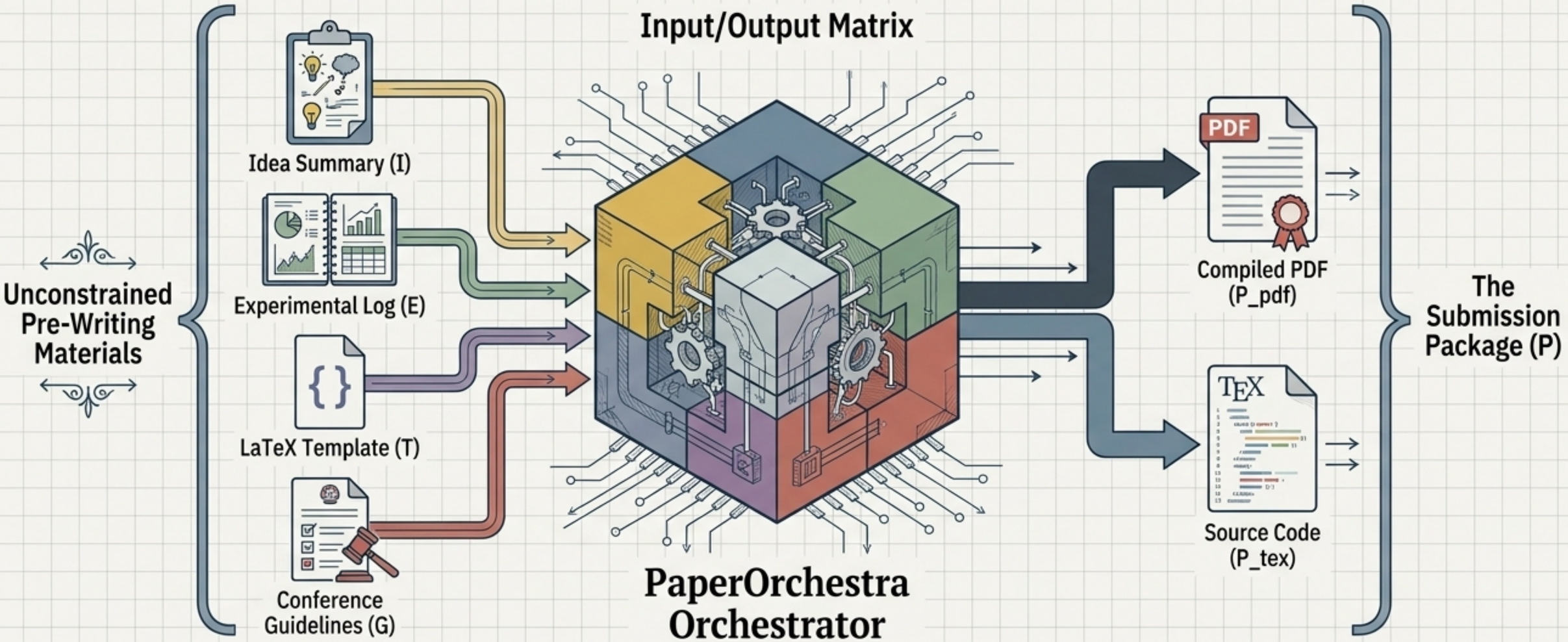
ข้อจำกัดของระบบ AI Scientist ในปัจจุบัน

ระบบ (System)	End-to-End LaTeX	Standalone Writer	รับข้อมูลไม่มีโครงสร้าง (Unstructured Input)	เจาะลึก Literature Review	สร้างแผนภาพแนวคิด (Conceptual Diagrams)
PaperRobot	○	○	○	○	○
data-to-paper	✓	✓	○	○	○
AI-Researcher	○	○	✓	○	○
Cycle Researcher	○	○	○	○	○
AI Scientist-v2	○	✓	✓	○	○
PaperOrchestra (Ours)	✓	✓	✓	✓	✓

PaperOrchestra เป็นเฟรมเวิร์กเดียวที่ผ่านทุกเงื่อนไข
โดยไม่ต้องพึ่งพาข้อมูลตั้งต้นที่มีโครงสร้างสมบูรณ์ หรือ Reference List ที่เตรียมไว้ล่วงหน้า

NotebookLM

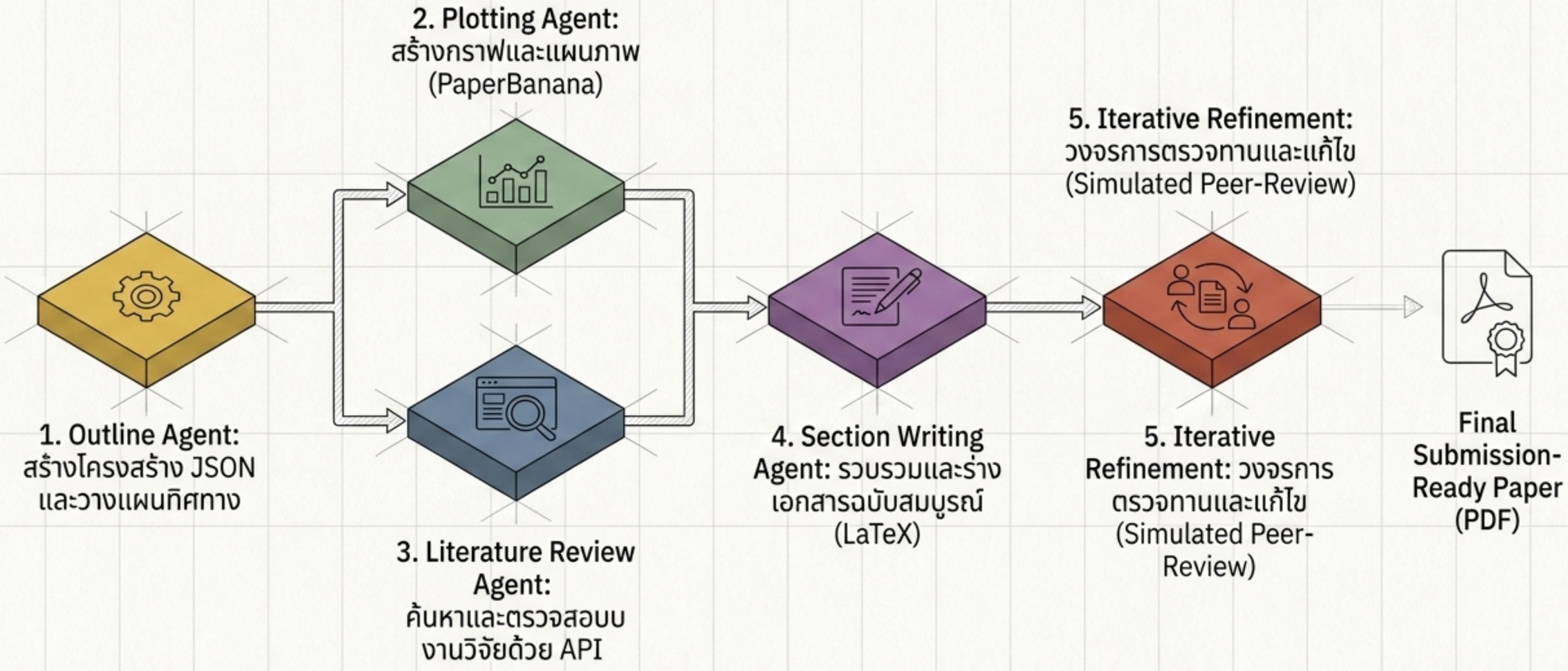
กระบวนทัศน์ใหม่: สถาปัตยกรรมที่แยกส่วนการทำงาน (Decoupled Multi-Agent)



Key Insight: การทำงานแบบ Decoupled ช่วยให้ระบบสามารถรับเอกสาร "ร่างหยาบ" จากมนุษย์ แล้วแปลงเป็นเอกสาร LaTeX ที่สมบูรณ์ได้ 100%

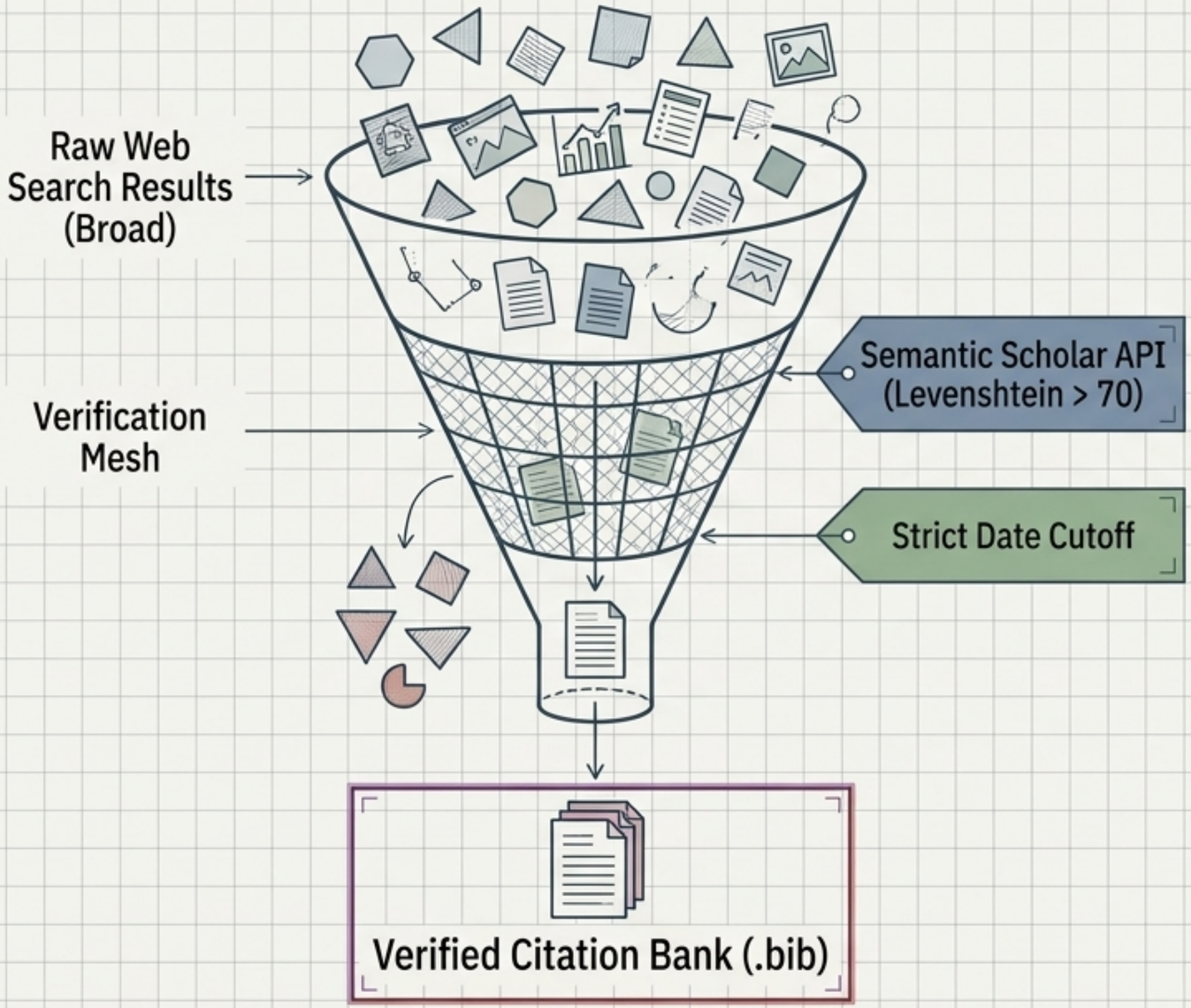
NotebookLM

สถาปัตยกรรมแบบ 5 ขั้นตอน (The 5-Step Agentic Symphony)



NotebookLM

เจาะลึก 1: ระบบกรอง Hallucination สำหรับ Literature Review



Data Scorecard

ประสิทธิภาพการอ้างอิง (Recall Gains)

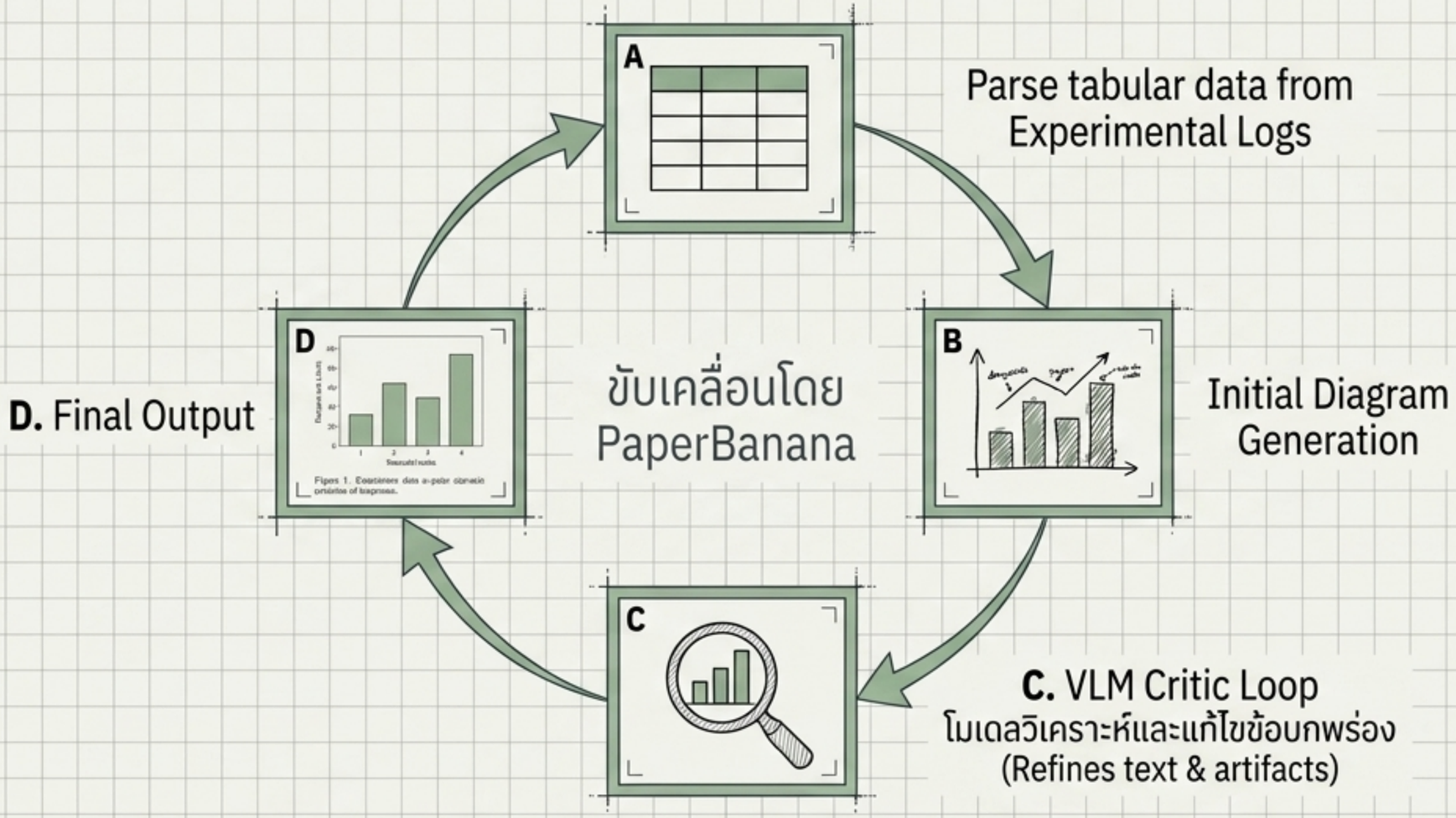
P0 (Must-Cite):
+2.13% to +6.07% เหนือคู่แข่ง

P1 (Good-to-Cite):
+12.59% to +13.75% เหนือคู่แข่ง

เฟรมเวิร์กสามารถสร้างอ้างอิงเฉลี่ย **45-47 รายการ** (ใกล้เคียงมนุษย์ที่ 59 รายการ) เทียบกับคู่แข่งที่สร้างได้เพียง 9-14 รายการ

NotebookLM

เจาะลึก 2: การสร้างภาพประกอบและกราฟโดยอัตโนมัติ



แม้ไม่ใช้รูปภาพจริงจากมนุษย์ (PlotOn) ระบบก็ยังสามารถ**ชนะหรือเสมอ**ในการประเมินเปรียบเทียบ (SxS) ได้ถึง **51% - 66%**

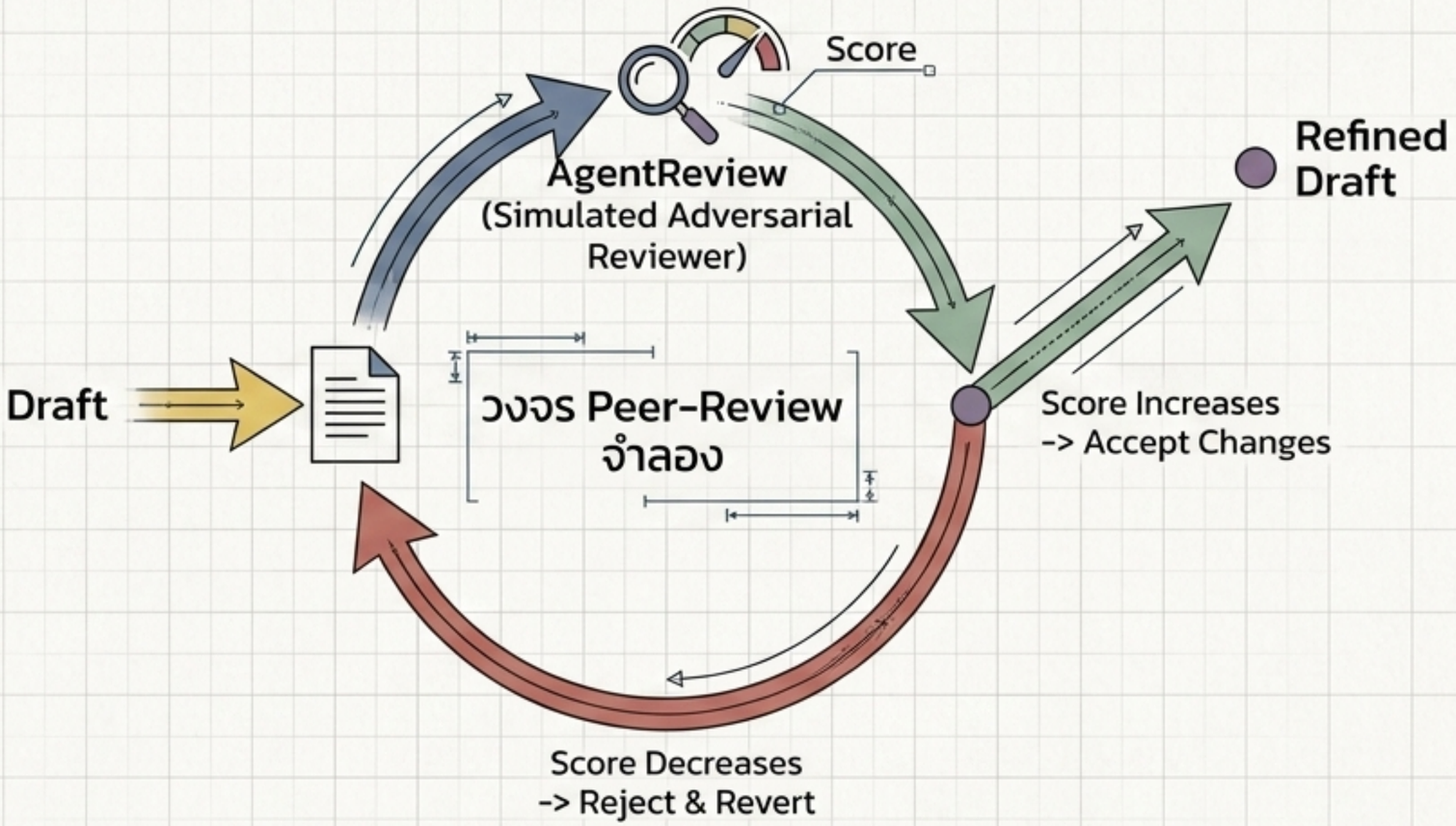
NotebookLM

เจาะลึก 3: วงจรพัฒนาเนื้อหาด้วยแบบจำลอง Peer-Review

Impact Callout

+19%
Acceptance Rate gain
(CVPR Simulation)

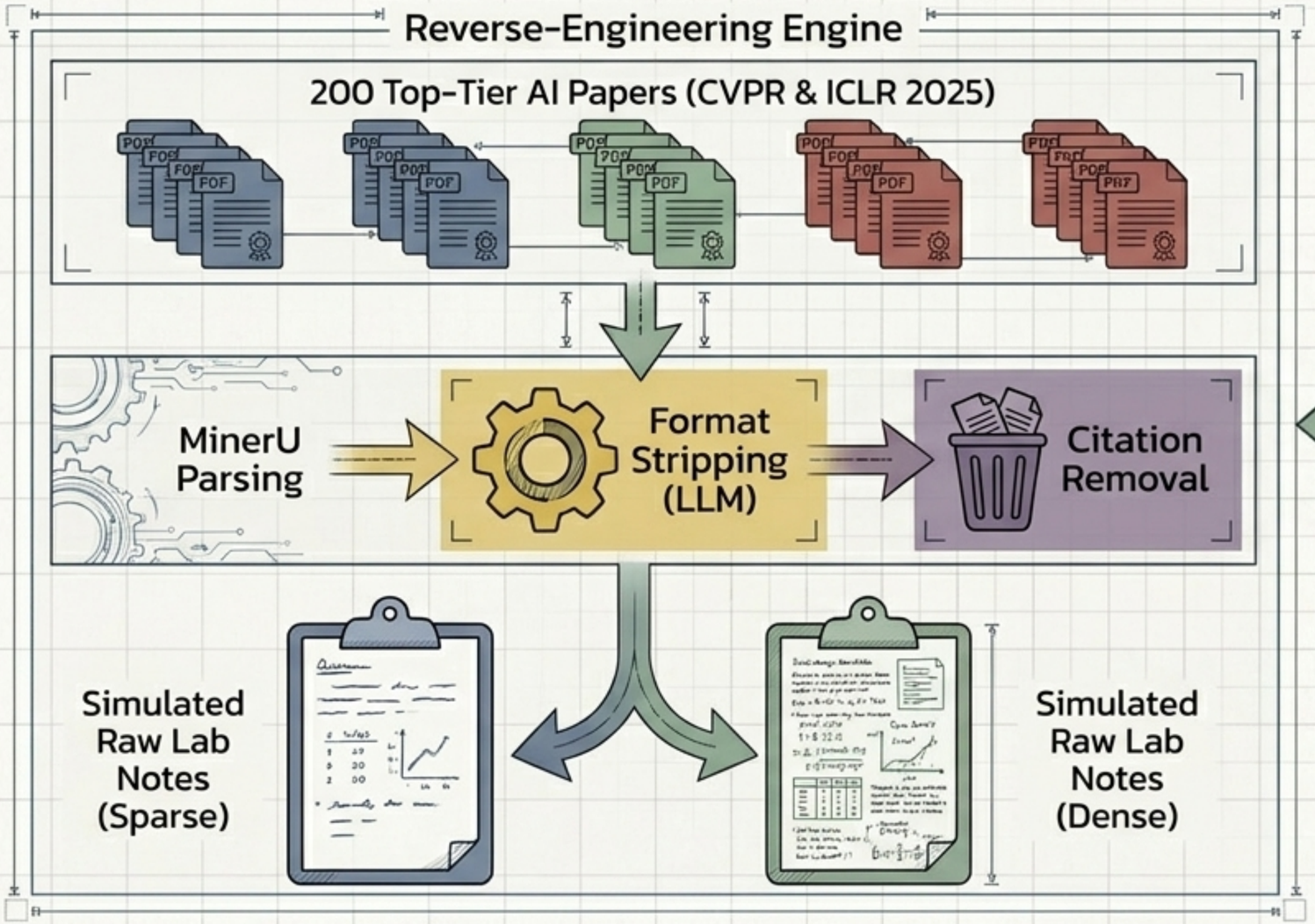
+22%
Acceptance Rate gain
(ICLR Simulation)



วงจรนี้รับประกันความถูกต้องทางเทคนิคและความลื่นไหลของการนำเสนอ ก่อนการคอมไพล์ขั้นสุดท้าย

NotebookLM

การทดสอบขีดจำกัดด้วย PaperWritingBench



การสกัดข้อมูลย้อนกลับ (Reverse-engineering) ช่วยสร้างสภาพแวดล้อมที่ปราศจากการชี้นำ ทดสอบว่าระบบ AI สามารถสร้างโครงสร้างทางวิชาการขึ้นใหม่จากศูนย์ได้ดีเพียงใด

NotebookLM

สถิติการประเมินจากผู้เชี่ยวชาญ (Human Evaluation Scorecard)

Scorecard 1: Literature Review Quality

+50% to +68% ↗

Absolute Win Rate Margin vs. AI Scientist-v2

Top Winning Axes (vs. AI Scientist-v2)

- Problem Framing ↗
- Prior Work Coverage ↗
- Organization & Synthesis ↗

Scorecard 2: Overall Paper Quality

+14% to +38% ↗

Absolute Win Rate Margin vs. AI Scientist-v2

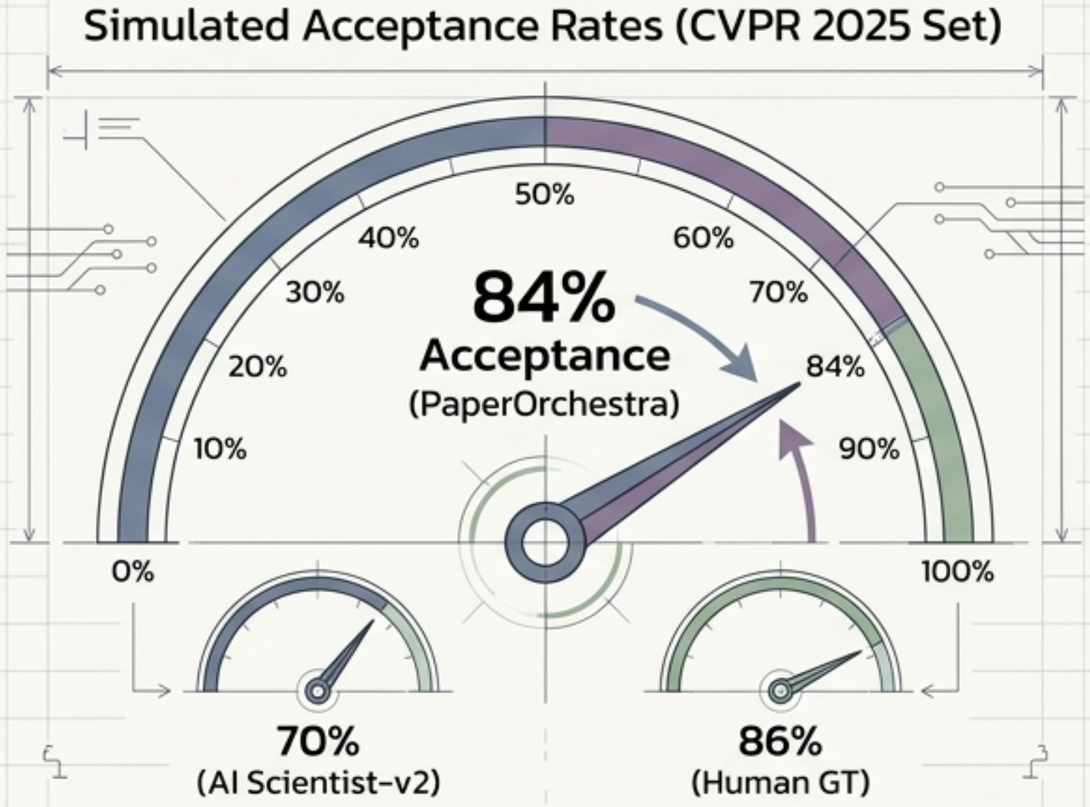
Top Winning Axes (vs. AI Scientist-v2)

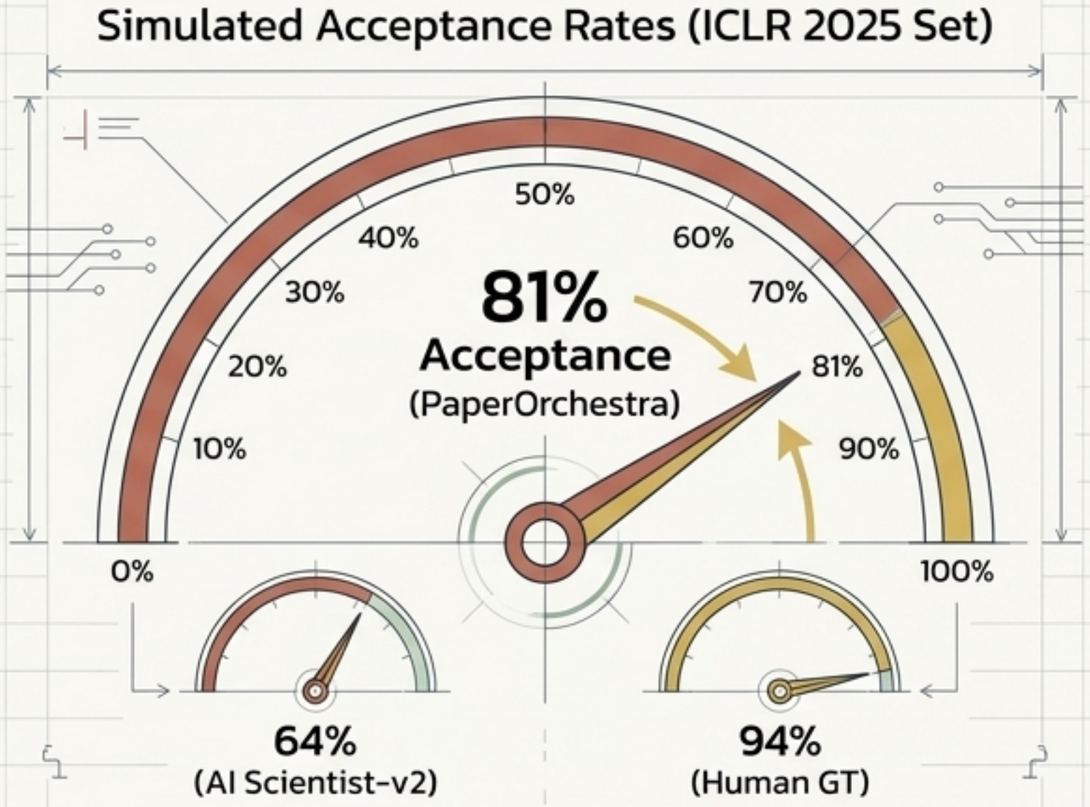
- Scientific Depth ↗
- Logical Flow ↗
- Presentation of Evidence ↗

Key Insight: ผู้เชี่ยวชาญด้าน AI ยืนยันว่า PaperOrchestra ให้ผลลัพธ์ที่เหนือกว่าโมเดลพื้นฐานในทุกมิติของการประเมิน

NotebookLM

ผลประเมินคุณภาพทางเทคนิคระดับ ScholarPeer



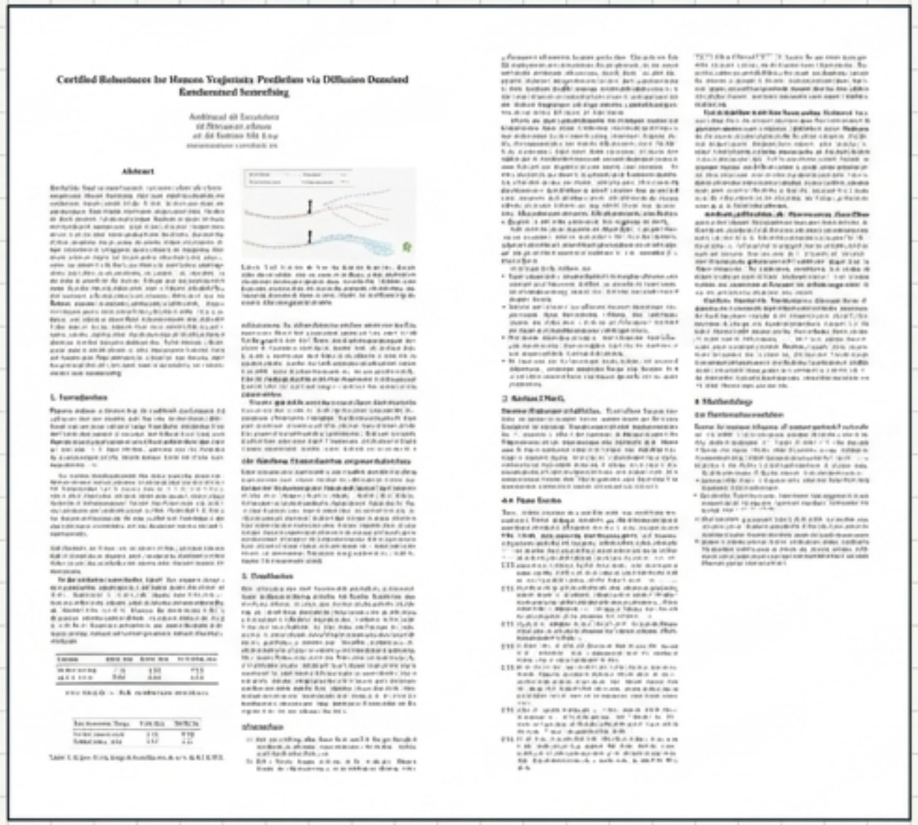


สถิติอัตราการตอบรับ (Acceptance Rate) จำลองที่พุ่งสูงถึง 84% พิสูจน์ให้เห็นว่าโครงสร้าง Multi-Agent ผลิตงานวิจัยที่มีน้ำหนักและความน่าเชื่อถือทางวิชาการเทียบชั้นมนุษย์

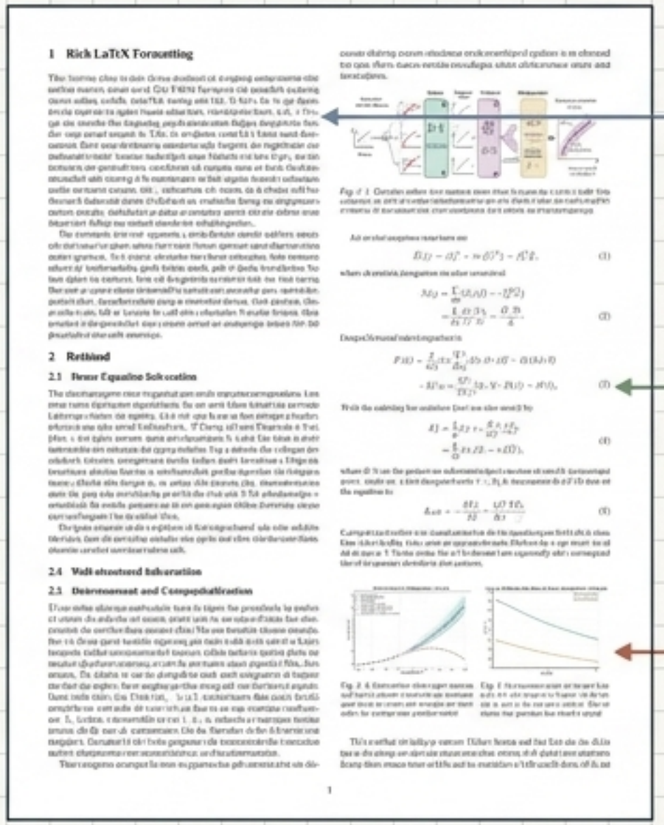
NotebookLM

โครงสร้างเอกสารผลลัพธ์ (Generated Visual Proof)

Baseline (AI Scientist-v2)



PaperOrchestra (Ours)



Rich LaTeX Formatting

Dense Equation Integration

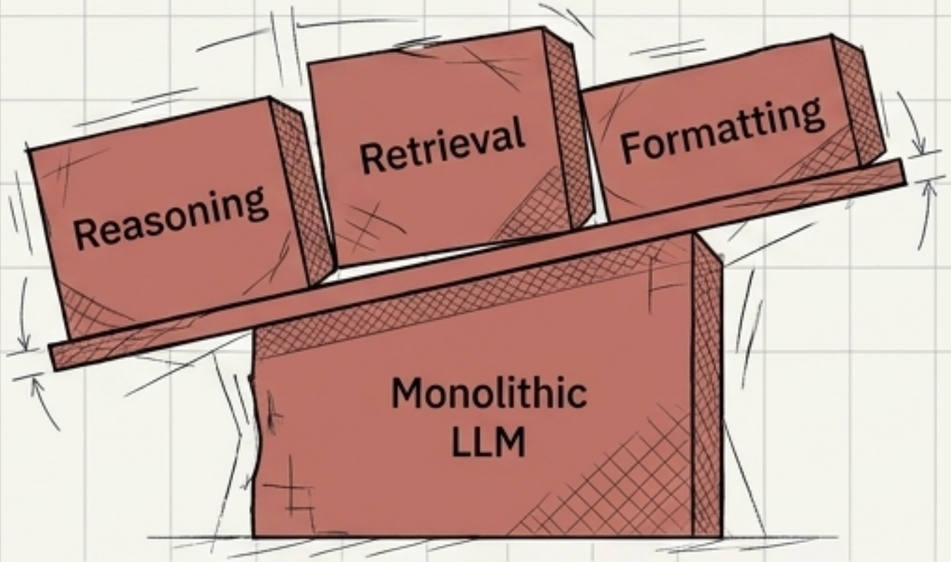
Seamless Plot Generation

Key Insight: สถาปัตยกรรมของเรารักษาความลึกซึ้งทางวิทยาศาสตร์ (Scientific Depth) ควบคู่ไปกับการจัดรูปแบบหน้ากระดาษที่สวยงามและพร้อมสำหรับการส่งตีพิมพ์

NotebookLM

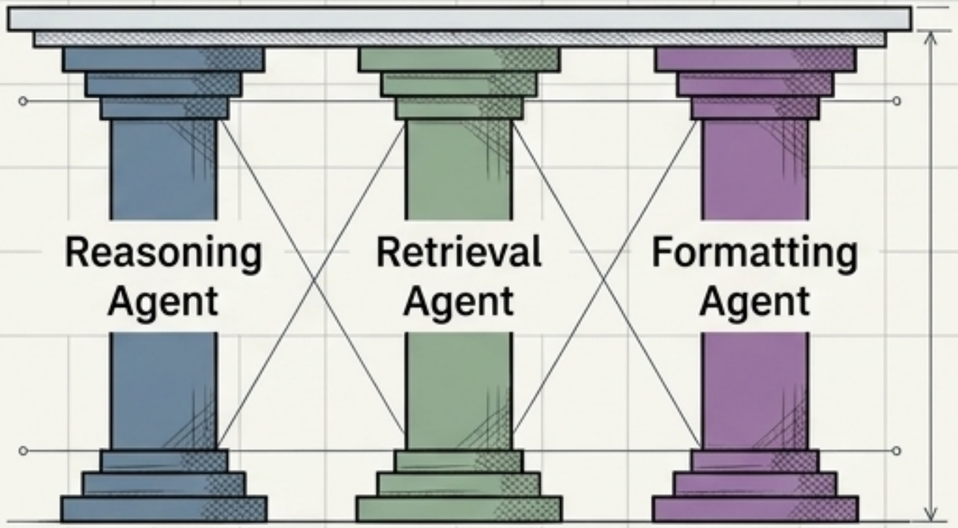
บทสรุป: ข้อได้เปรียบของระบบ Multi-Agent

Monolithic LLM
(Unbalanced)



Conflates tasks -> Hallucinated citations, shallow logic, formatting errors

Multi-Agent Orchestration
(Balanced)



Strictly decouples tasks -> Parallel API calls for rapid discovery + rigorous sequential verification

ความสำเร็จของ PaperOrchestra ไม่ได้เกิดจากการขยายขนาดโมเดล (Scaling) แต่เกิดจากการแยกส่วนความรับผิดชอบ (Decoupling) อย่างเป็นระบบ

NotebookLM

อนาคตของการวิจัยที่ขับเคลื่อนด้วยมนุษย์และ AI

1. Empowerment, Not Replacement:
ระบบทำหน้าที่เป็นผู้ช่วยขั้นสูง (Advanced Assistive Tool) เพื่อลดภาระงานเขียนที่ใช้เวลานาน

2. Human in the Loop:
มนุษย์รับผิดชอบความคิดริเริ่มและความถูกต้อง ในขณะที่ AI ดูแลการจัดโครงสร้างและสังเคราะห์วรรณกรรม

3. Next Steps:
เตรียมพร้อมก้าวสู่สภาพแวดล้อมการเขียนเชิงโต้ตอบ (Interactive Editing) แบบเรียลไทม์

เปลี่ยนไอเดียที่ไร้รูปแบบ ให้เป็นวิทยาศาสตร์ที่พร้อมตีพิมพ์
(Transforming unconstrained ideas into submission-ready science).

NotebookLM