

# AI Chat with PDF (RAG)

idea-hub · ai-chat-with-pdf-rag

## 1. The opportunity

ตลาดไทยมีเอกสารราชการ สัญญากฎหมาย คู่มือนโยบาย และรายงานวิจัยที่หนาและซับซ้อนมหาศาล แต่เครื่องมือ Chat-with-PDF ที่ดีส่วนใหญ่เป็นภาษาอังกฤษ ไม่รองรับ Thai encoding ได้ดี และไม่มี citation แบบอ้างอิงเลขหน้าที่ตรวจสอบได้จริง ผู้ใช้กลุ่มหลัก เช่น นักกฎหมาย นักวิจัย และ compliance officer ต้องการความแม่นยำระดับ "อ้างอิงได้" ไม่ใช่แค่สรุปแบบ hallucinate ช่วง 2024-2025 LLM ราคาถูกลงมาก (GPT-4o mini, Gemini Flash) ทำให้ต้นทุน per-query ต่ำพอที่จะทำ SaaS ราคา SME ไทยจ่ายได้ ขนาดตลาด RAG tools ระดับโลกอยู่ที่ประมาณ \$1.5B ในปี 2024 และโตเร็ว ส่วนแบ่งตลาดไทยที่ localized ยังแทบว่าง

## 2. Who would pay

**Persona หลัก: นักกฎหมาย/ที่ปรึกษากฎหมายในสำนักงานกลางขนาดเล็ก-กลาง (5-50 คน)**

- Job title: Associate Lawyer, Legal Counsel, Compliance Officer
- Industry: Law firm, ธนาคาร, บริษัทประกัน, รัฐวิสาหกิจ
- Company size: 10-200 คน
- ปัจจุบันทำอะไรแทน: Ctrl+F ใน PDF, จ้าง paralegal ค้นเอกสาร, ใช้ Adobe Acrobat Pro (\$23/เดือน) ที่ค้นได้แค่ keyword ไม่เข้าใจความหมาย หรือใช้ ChatGPT แบบ paste ข้อความซึ่งเสี่ยง data leak และไม่มี citation
- Willingness to pay: ฿500-1,500/เดือน/คน ถ้าระบบอ้างอิงเลขหน้าได้และเชื่อถือได้ในศาล

**Persona รอง: นักศึกษาปริญญาโท/นักวิจัยมหาวิทยาลัยไทย** ที่ต้องอ่าน paper/thesis จำนวนมาก — ยอมจ่าย ฿99-299/เดือน

## 3. Competitor landscape

ชื่อ	URL	จุดแข็ง	จุดอ่อนสำหรับตลาดไทย
ChatPDF	chatpdf.com	ง่าย, ฟรี tier ดี	ภาษาไทยฟัง บ่อย, ไม่มี citation เลขหน้าจริง, เก็บข้อมูลต่างประเทศ
Adobe Acrobat AI	adobe.com	Integrate กับ PDF workflow	ราคาแพง (\$23+/เดือน), AI ตอบภาษาไทยได้ไม่ดี
NotebookLM	notebooklm.google.com	ฟรี, citation ดีมาก	ไม่มี Thai UI, ข้อมูลขึ้น Google server, ไม่มี enterprise privacy

# AI Chat with PDF (RAG)

idea-hub · ai-chat-with-pdf-rag

<b>AskYourPDF</b>	askyourpdf.com	API พร้อม, integrations	UI ภาษาอังกฤษ, support ภาษาไทยอ่อน
<b>Humata</b>	humata.ai	เน้น research/academic	ราคา \$20+/เดือน, ไม่รู้จัก context กฎหมายไทย

**ทำไมตลาดไทยยัง "ว่าง":** ไม่ใช่ไม่มี demand แต่ hard to crack เพราะ Thai NLP tokenization ซับซ้อน (ไม่มี whitespace), character encoding บน PDF ภาษาไทยแตกต่างกัน (TH Sarabun vs embedded font), และเจ้าตลาดต่างชาติไม่คุ้มที่จะ optimize ตลาดเล็ก

## 4. Wrapper risk reality check

**ความเสี่ยง:** สูง ถ้า build แค่ "upload PDF 'ถาม GPT-4" — ChatGPT Plus มี feature นี้อยู่แล้ว และ NotebookLM ฟรี

**Moat จริงๆ ต้องมาจาก 3 เรื่อง:**

- \*\*Thai PDF parsing ที่ดีกว่า\*\*** — ต้องลงทุนใน pre-processing pipeline ที่รองรับ font embedding แปลกๆ ของเอกสารราชการไทย (กรมสรรพากร, สำนักงาน ก.อ.ต.) ซึ่งพวก generic tool parse ผิดบ่อยมาก
- \*\*Citation ที่ audit-ready\*\*** — แสดง highlight บน PDF ต้นฉบับได้จริง ไม่ใช่แค่บอกเลขหน้า หมายจะจ่ายแพงขึ้น 2-3x ถ้าคลิกแล้วดึงไปหน้าจริงได้
- \*\*Data residency / On-premise option\*\*** — สำนักงานกฎหมายและธนาคารไทยมีข้อกำหนด PDPA ที่ห้ามส่งข้อมูลลูกค้าออกนอกประเทศ ถ้า offer self-hosted หรือ private cloud บน AWS Bangkok region นี่คือ moat ที่ ChatGPT ทำให้ไม่ได้

## 5. Go-to-market angle

**10 ลูกค้าแรกหาจากไหน:**

เริ่มจาก **สมาคมทนายความแห่งประเทศไทย** และกลุ่ม Facebook/LINE "กฎหมายสำหรับนักกฎหมาย" (สมาชิก 30,000+) ทำ demo video สั้น 60 วินาที แสดงการอัปเดตสัญญาเช่า 'ถามว่า "ผู้เช่าผิดสัญญาข้อไหน" 'highlight หน้า 3 ข้อ 7.2 กันที คนดูจะ share เอง

**SEO keywords ที่ยังไม่แข่งมาก (ภาษาไทย):**

- "ถามเอกสาร PDF ภาษาไทย AI"
- "วิเคราะห์สัญญาด้วย AI"
- "AI อ่าน PDF กฎหมาย"
- "RAG ภาษาไทย"

# AI Chat with PDF (RAG)

idea-hub · ai-chat-with-pdf-rag

**Direct outreach:** LinkedIn ทา "Legal Counsel" + "Thailand" + "FinTech/Insurance" มีประมาณ 2,000+ profile ส่ง cold message พร้อม loom demo 2 นาที อัตรา reply ในกลุ่มนี้ดีกว่า cold email มาก

**Community:** กลุ่ม Pantip ห้องสิบลร, กลุ่ม Startup Thailand ใน Facebook, และ **TLCA (Thai Listed Companies Association)** ที่ compliance officer รวมตัวกัน

**Partnership:** บริษัท e-signature ไทย เช่น DigSign หรือ CoSign ที่ลูกค้าเหมือนกันทุกคน — เสนอ white-label หรือ referral deal

## 6. Pricing thesis

Tier	ราคา/เดือน	สิ่งที่ได้
Free	฿0	3 PDF/เดือน, max 20 หน้า/ไฟล์
Pro	฿499	Unlimited PDF, highlight citation, 200MB storage
Team	฿1,499/คน (min 3 คน)	Shared workspace, audit log, PDPA-compliant storage
Enterprise	฿15,000+/เดือน	On-premise/private cloud, SSO, SLA

**คำโฆษณา: free alternatives:** NotebookLM ไม่มี audit trail, ไม่ compliance-ready, ไม่ highlight ต้นฉบับได้ ChatGPT ต้อง paste เอง เสี่ยง data leak ราคา ฿499 เทียบกับ paralegal รายชั่วโมง ฿500-800/ชม. คู่มนั่นนอนถ้าประหยัดได้ 1 ชั่วโมง/เดือน

## 7. Build complexity

**Time to MVP:** 6-8 สัปดาห์ (คน 1-2 คน)

**Core tech stack:**

- **PDF parsing:** PyMuPDF + pdfplumber (รองรับ Thai font ได้ดีกว่า pdfminer)
- **Chunking + Embedding:** LangChain หรือ LlamaIndex + OpenAI `text-embedding-3-small`
- **Vector DB:** Qdrant (self-host ได้) หรือ Supabase pgvector (ถูกกว่าสำหรับ early stage)
- **LLM:** GPT-4o mini (ถูก, เร็ว, ภาษาไทยดี) หรือ Gemini 1.5 Flash
- **Frontend:** Next.js + PDF.js สำหรับ highlight annotation

## RESEARCH BUNDLE

# AI Chat with PDF (RAG)

idea-hub · ai-chat-with-pdf-rag

- **Auth + billing:** Clerk + Stripe (หรือ Omise สำหรับคนไทย)

**\*\***