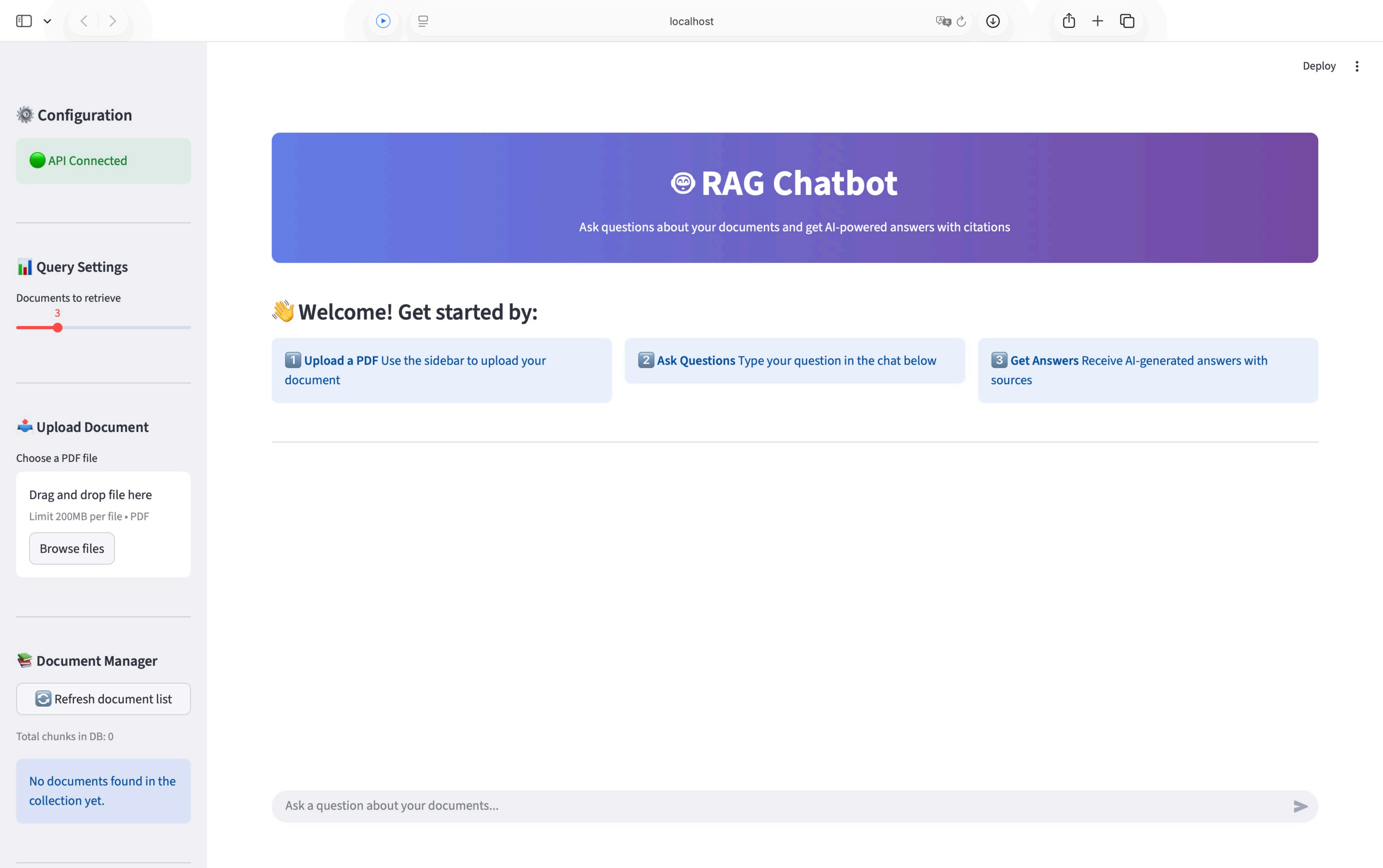


⚙️ Configuration

🟢 API Connected

📊 Query Settings

Documents to retrieve: 3

📤 Upload Document

Choose a PDF file

Drag and drop file here

Limit 200MB per file • PDF

📚 Document Manager

Total chunks in DB: 0

No documents found in the collection yet.

🤖 RAG Chatbot

Ask questions about your documents and get AI-powered answers with citations

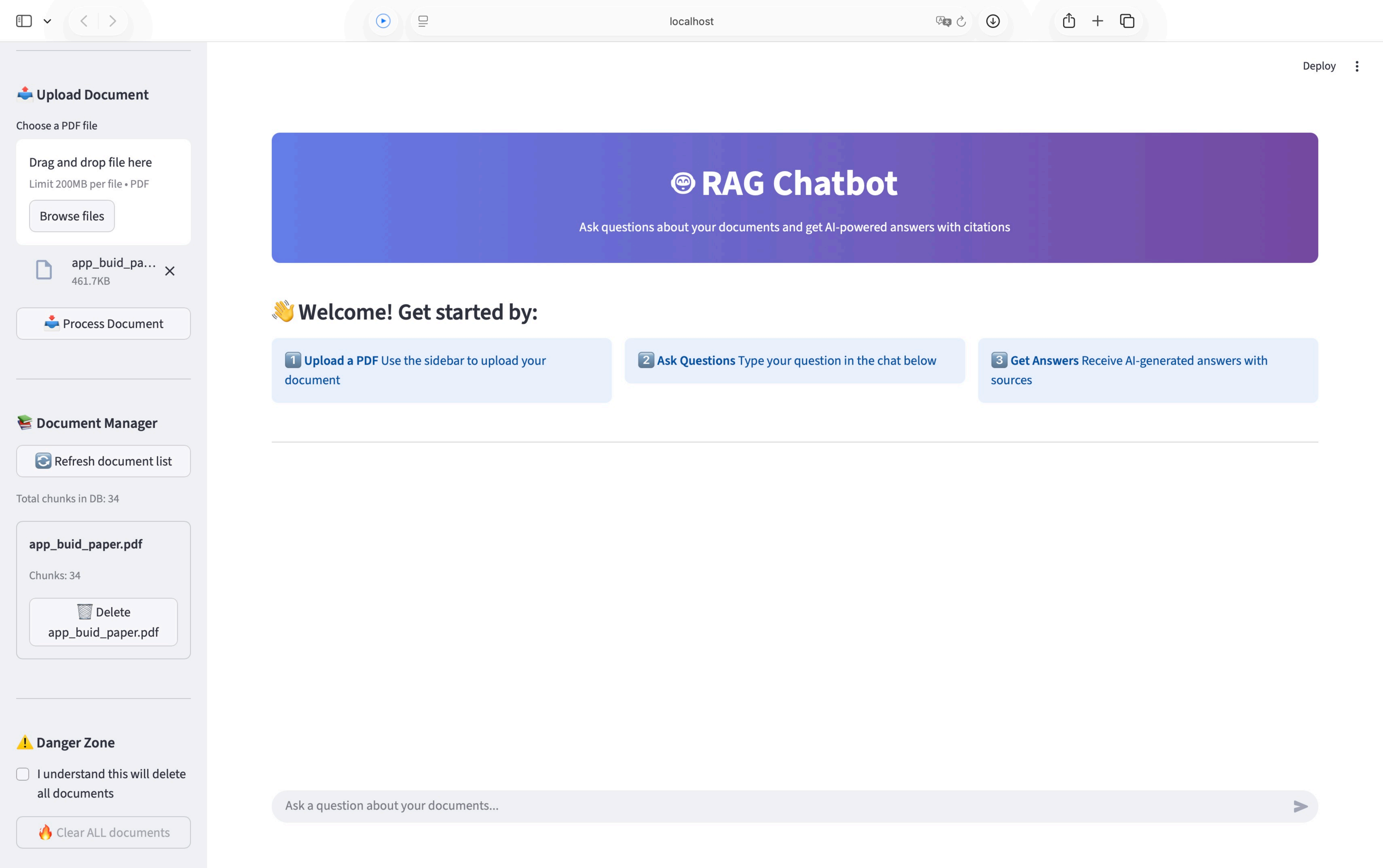
👋 Welcome! Get started by:

1️⃣ **Upload a PDF** Use the sidebar to upload your document

2️⃣ **Ask Questions** Type your question in the chat below

3️⃣ **Get Answers** Receive AI-generated answers with sources

Ask a question about your documents...



📥 **Upload Document**

Choose a PDF file

Drag and drop file here

Limit 200MB per file • PDF

app_buid_pa…
461.7KB

📚 **Document Manager**

Total chunks in DB: 34

app_buid_paper.pdf

Chunks: 34

⚠️ **Danger Zone**

☐ I understand this will delete all documents

🤖 RAG Chatbot

Ask questions about your documents and get AI-powered answers with citations

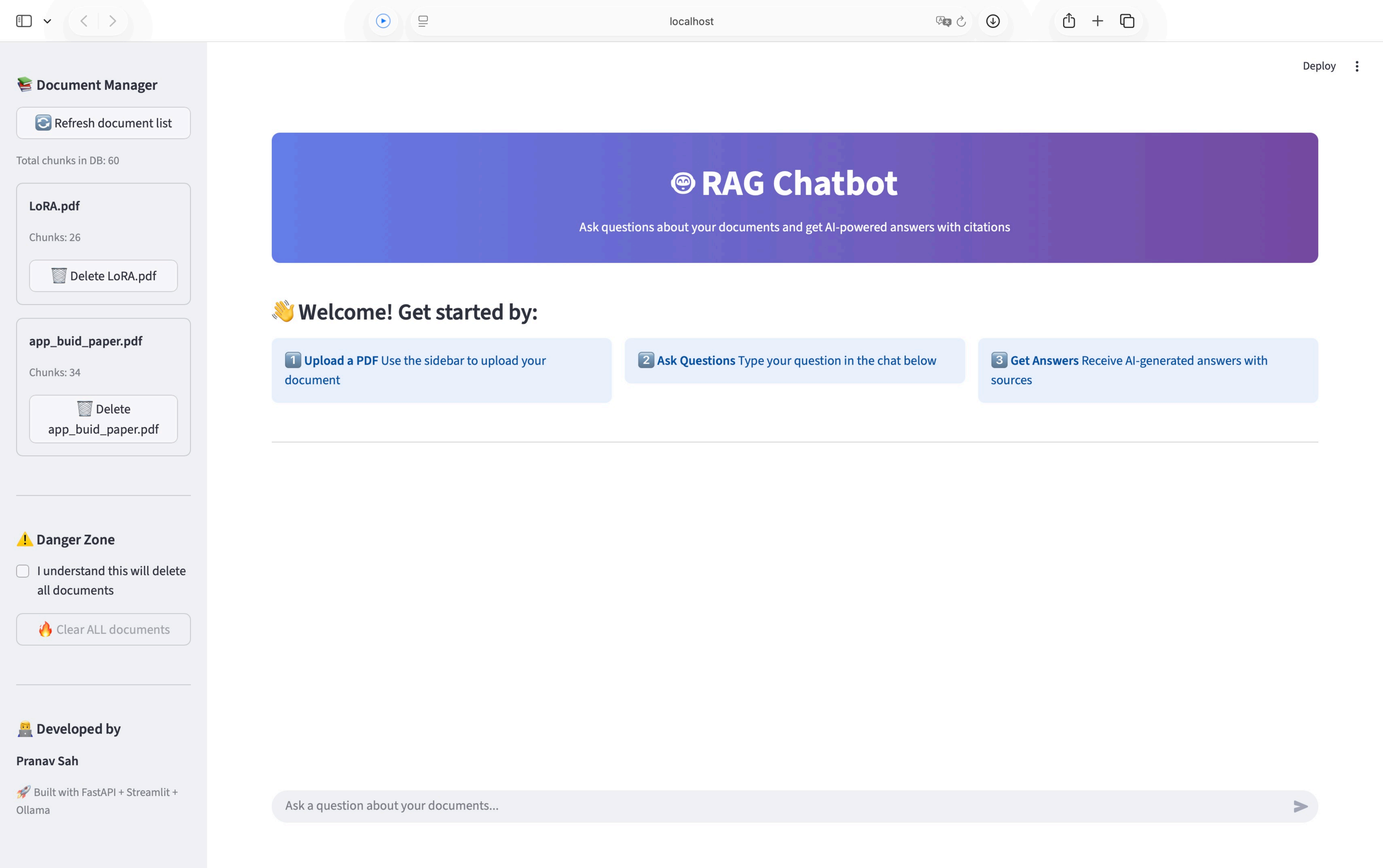
👋 Welcome! Get started by:

1️⃣ **Upload a PDF** Use the sidebar to upload your document

2️⃣ **Ask Questions** Type your question in the chat below

3️⃣ **Get Answers** Receive AI-generated answers with sources

Ask a question about your documents...



📚 **Document Manager**

Total chunks in DB: 60

LoRA.pdf

Chunks: 26

app_buid_paper.pdf

Chunks: 34

⚠️ **Danger Zone**

☐ I understand this will delete all documents

🧑‍💻 **Developed by**

Pranav Sah

🚀 Built with FastAPI + Streamlit + Ollama

🤖 **RAG Chatbot**

Ask questions about your documents and get AI-powered answers with citations

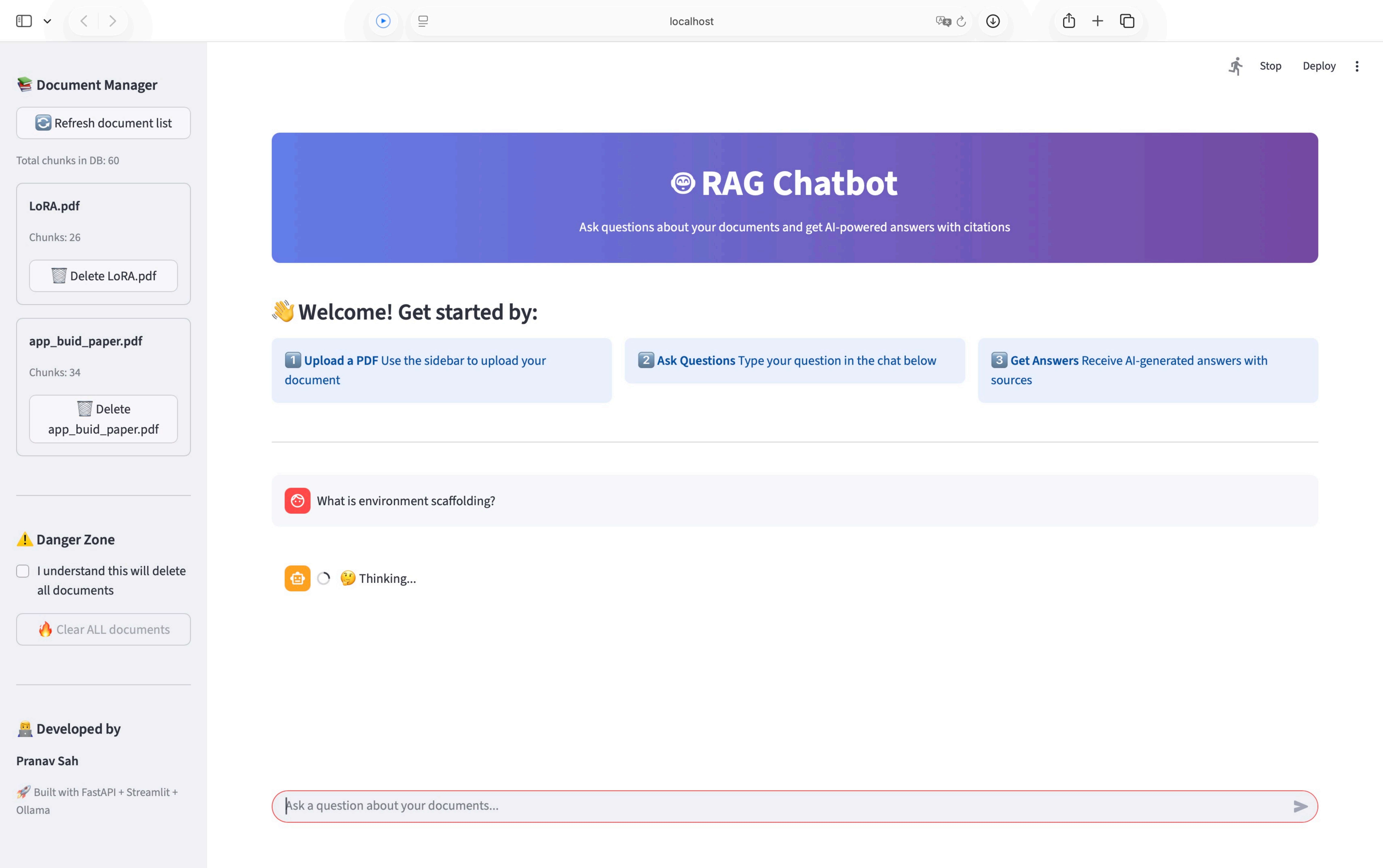
👋 **Welcome! Get started by:**

1️⃣ **Upload a PDF** Use the sidebar to upload your document

2️⃣ **Ask Questions** Type your question in the chat below

3️⃣ **Get Answers** Receive AI-generated answers with sources

Ask a question about your documents...



📚 Document Manager

Total chunks in DB: 60

LoRA.pdf

Chunks: 26

app_buid_paper.pdf

Chunks: 34

⚠️ Danger Zone

☐ I understand this will delete all documents

👨‍💻 Developed by

Pranav Sah

🚀 Built with FastAPI + Streamlit + Ollama

🤖 RAG Chatbot

Ask questions about your documents and get AI-powered answers with citations

👋 Welcome! Get started by:

1️⃣ **Upload a PDF** Use the sidebar to upload your document

2️⃣ **Ask Questions** Type your question in the chat below

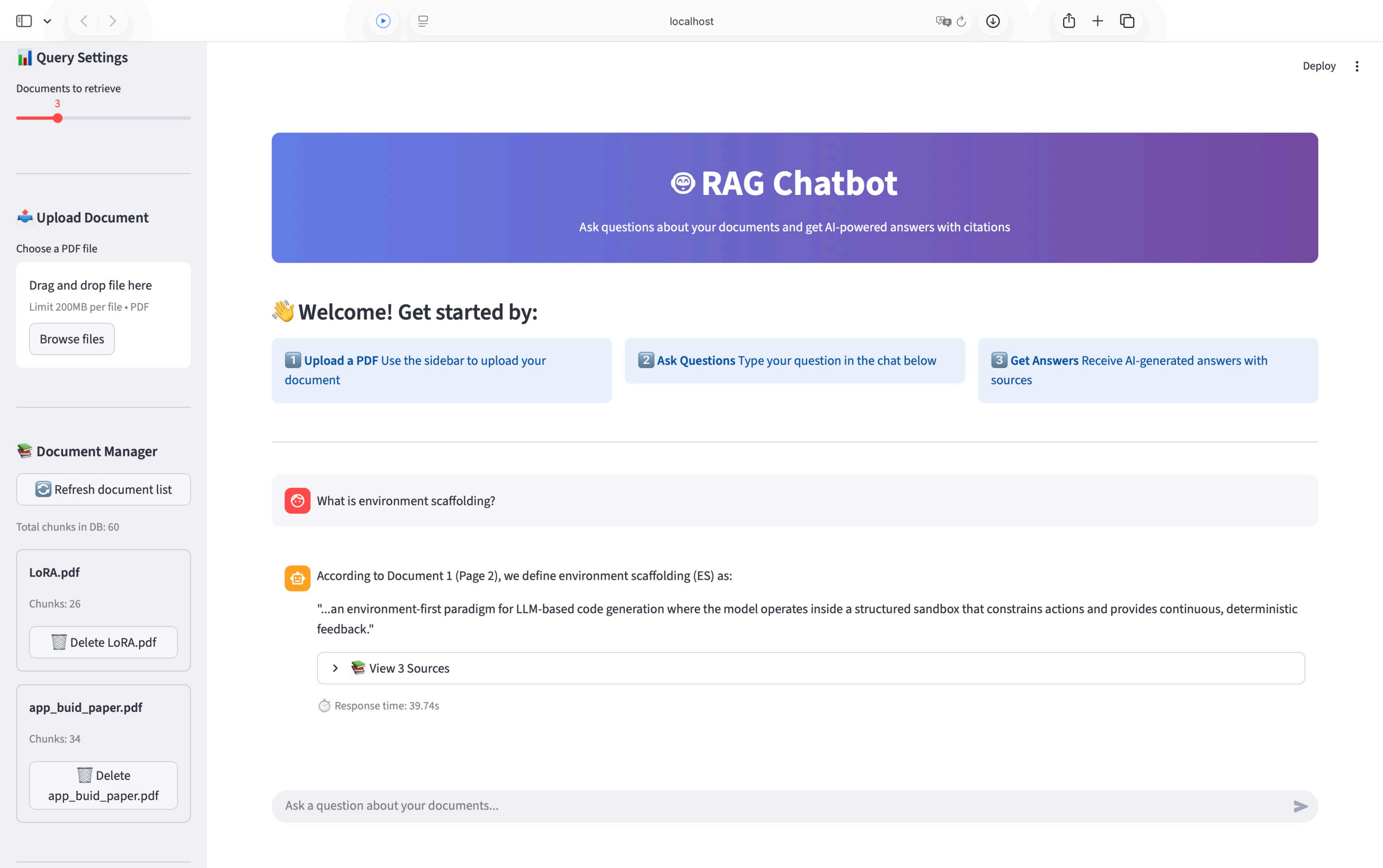
3️⃣ **Get Answers** Receive AI-generated answers with sources

---

What is environment scaffolding?

🤔 Thinking...

Ask a question about your documents...



Documents to retrieve: 3

Total chunks in DB: 60

LoRA.pdf

Chunks: 26

app_buid_paper.pdf

Chunks: 34

😊 RAG Chatbot

Ask questions about your documents and get AI-powered answers with citations

👋 Welcome! Get started by:

1️⃣ **Upload a PDF** Use the sidebar to upload your document

2️⃣ **Ask Questions** Type your question in the chat below

3️⃣ **Get Answers** Receive AI-generated answers with sources

---

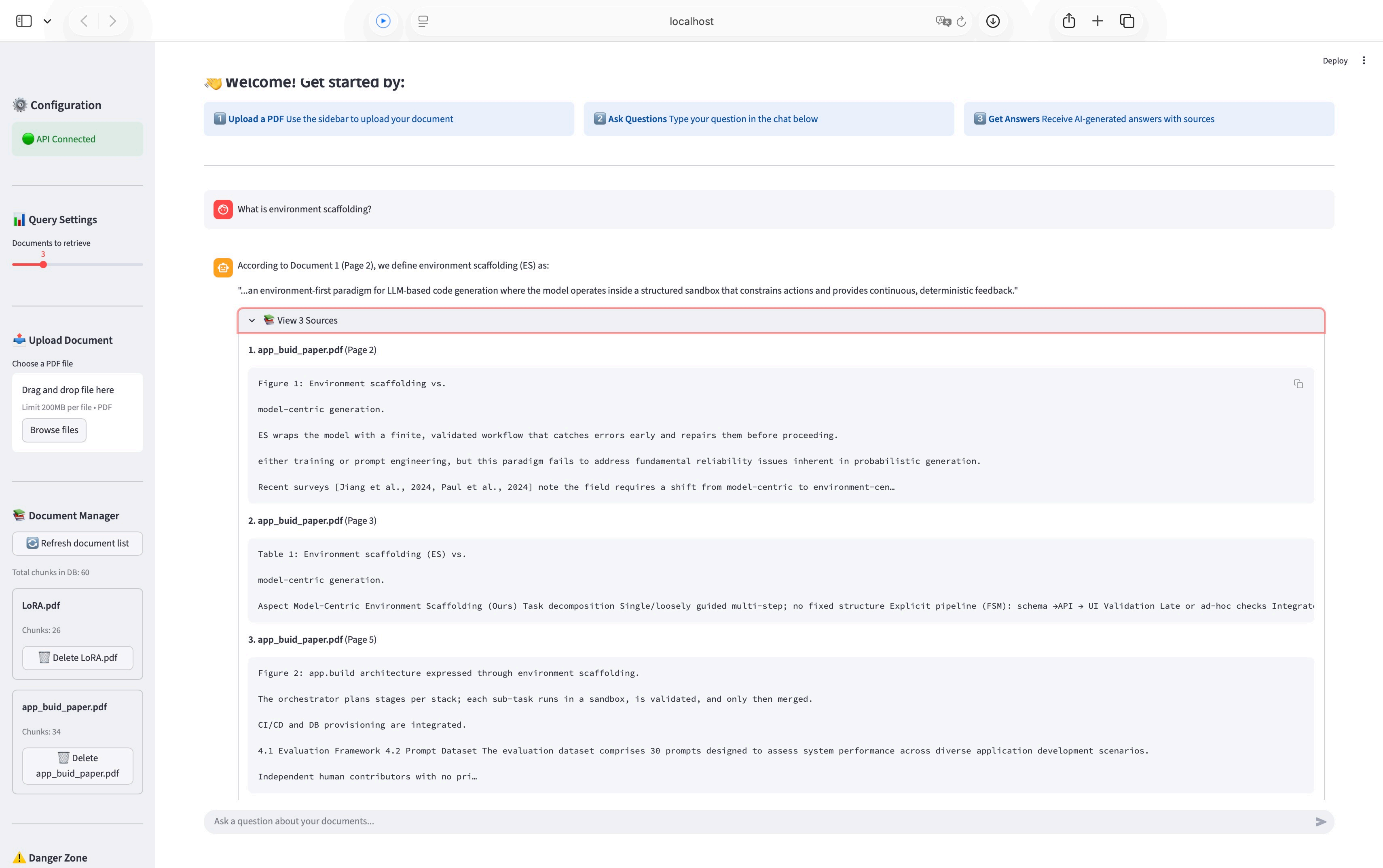
What is environment scaffolding?

According to Document 1 (Page 2), we define environment scaffolding (ES) as:

"...an environment-first paradigm for LLM-based code generation where the model operates inside a structured sandbox that constrains actions and provides continuous, deterministic feedback."

⏱️ Response time: 39.74s

Ask a question about your documents...



Configuration

API Connected

Query Settings

Documents to retrieve: 3

Upload Document

Choose a PDF file

Drag and drop file here

Limit 200MB per file • PDF

Document Manager

Total chunks in DB: 60

LoRA.pdf

Chunks: 26

app_buid_paper.pdf

Chunks: 34

Danger Zone

Welcome! Get started by:

1 **Upload a PDF** Use the sidebar to upload your document

2 **Ask Questions** Type your question in the chat below

3 **Get Answers** Receive AI-generated answers with sources

What is environment scaffolding?

According to Document 1 (Page 2), we define environment scaffolding (ES) as:

"...an environment-first paradigm for LLM-based code generation where the model operates inside a structured sandbox that constrains actions and provides continuous, deterministic feedback."

View 3 Sources

1. **app_buid_paper.pdf** (Page 2)

```
Figure 1: Environment scaffolding vs.

model-centric generation.

ES wraps the model with a finite, validated workflow that catches errors early and repairs them before proceeding.

either training or prompt engineering, but this paradigm fails to address fundamental reliability issues inherent in probabilistic generation.

Recent surveys [Jiang et al., 2024, Paul et al., 2024] note the field requires a shift from model-centric to environment-cen…
```

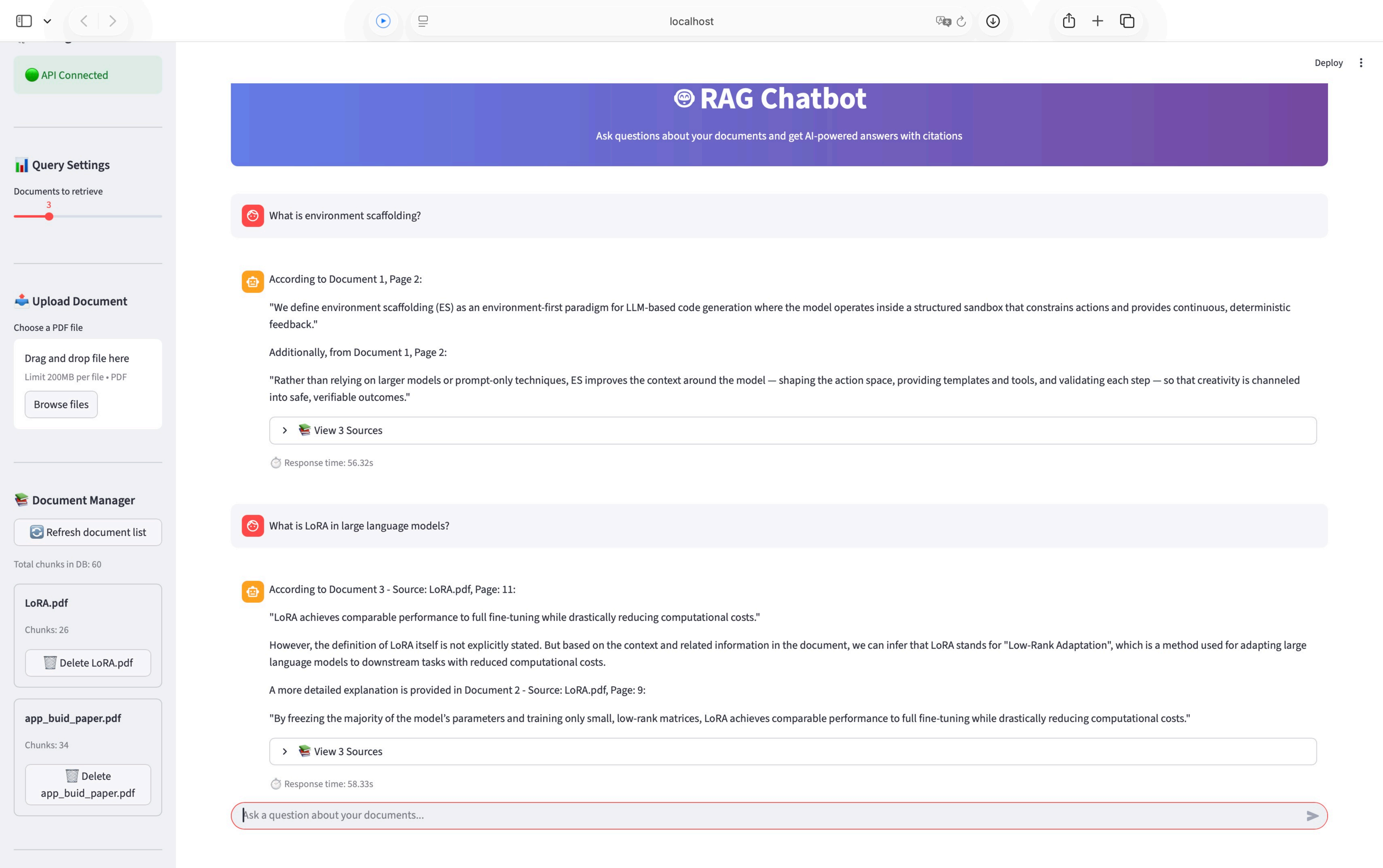
2. **app_buid_paper.pdf** (Page 3)

```
Table 1: Environment scaffolding (ES) vs.

model-centric generation.

Aspect Model-Centric Environment Scaffolding (Ours) Task decomposition Single/loosely guided multi-step; no fixed structure Explicit pipeline (FSM): schema →API → UI Validation Late or ad-hoc checks Integrat
```

3. **app_buid_paper.pdf** (Page 5)

```
Figure 2: app.build architecture expressed through environment scaffolding.

The orchestrator plans stages per stack; each sub-task runs in a sandbox, is validated, and only then merged.

CI/CD and DB provisioning are integrated.

4.1 Evaluation Framework 4.2 Prompt Dataset The evaluation dataset comprises 30 prompts designed to assess system performance across diverse application development scenarios.

Independent human contributors with no pri…
```

Ask a question about your documents...



Documents to retrieve: 3

Total chunks in DB: 60

LoRA.pdf

Chunks: 26

app_buid_paper.pdf

Chunks: 34

RAG Chatbot

Ask questions about your documents and get AI-powered answers with citations

What is environment scaffolding?

According to Document 1, Page 2:

"We define environment scaffolding (ES) as an environment-first paradigm for LLM-based code generation where the model operates inside a structured sandbox that constrains actions and provides continuous, deterministic feedback."

Additionally, from Document 1, Page 2:

"Rather than relying on larger models or prompt-only techniques, ES improves the context around the model — shaping the action space, providing templates and tools, and validating each step — so that creativity is channeled into safe, verifiable outcomes."

📚 View 3 Sources

Response time: 56.32s

What is LoRA in large language models?

According to Document 3 - Source: LoRA.pdf, Page: 11:

"LoRA achieves comparable performance to full fine-tuning while drastically reducing computational costs."

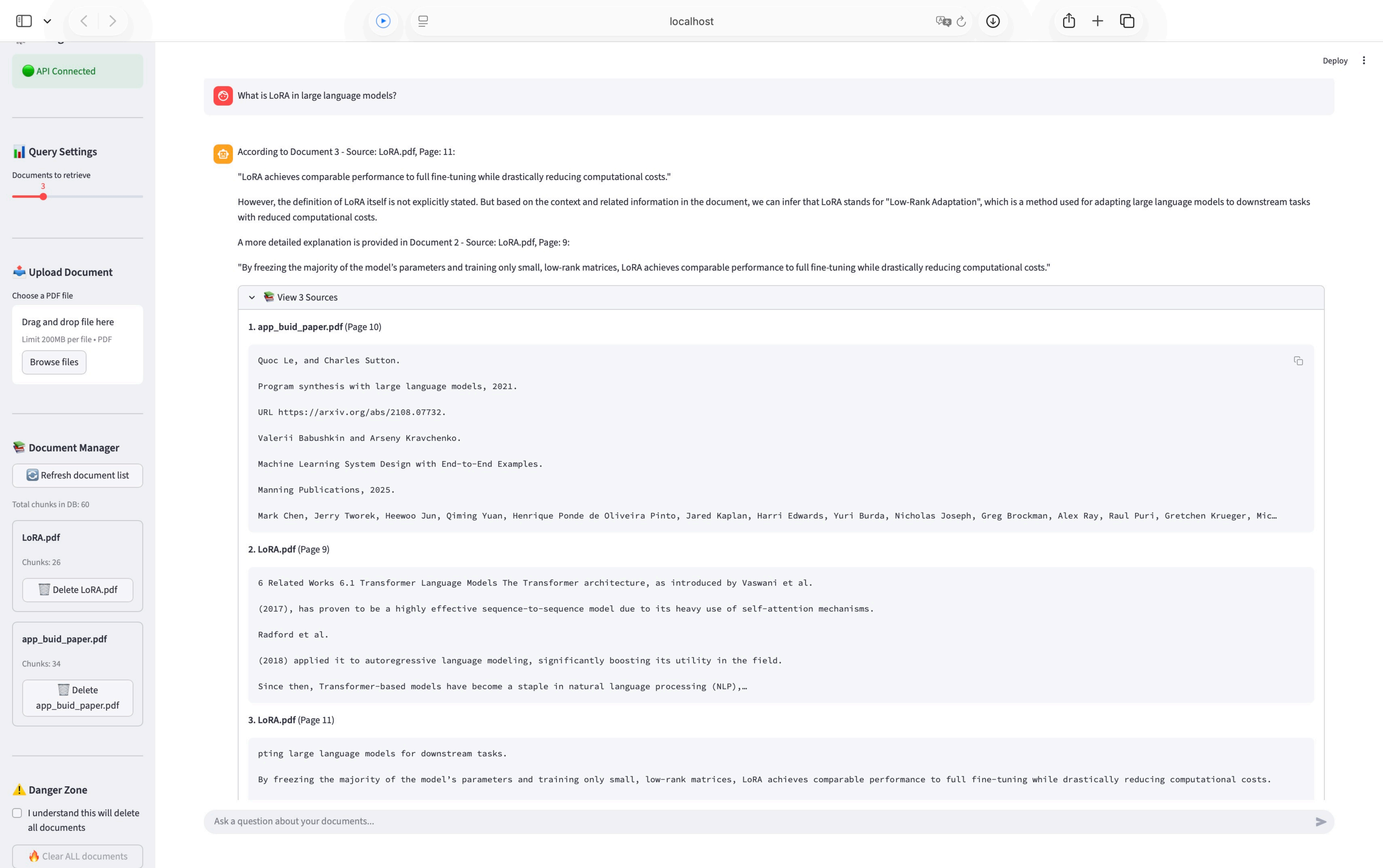
However, the definition of LoRA itself is not explicitly stated. But based on the context and related information in the document, we can infer that LoRA stands for "Low-Rank Adaptation", which is a method used for adapting large language models to downstream tasks with reduced computational costs.

A more detailed explanation is provided in Document 2 - Source: LoRA.pdf, Page: 9:

"By freezing the majority of the model's parameters and training only small, low-rank matrices, LoRA achieves comparable performance to full fine-tuning while drastically reducing computational costs."

📚 View 3 Sources

Response time: 58.33s



Query Settings

Documents to retrieve: 3

Upload Document

Choose a PDF file

Document Manager

Total chunks in DB: 60

LoRA.pdf

Chunks: 26

app_buid_paper.pdf

Chunks: 34

Danger Zone

☐ I understand this will delete all documents

What is LoRA in large language models?

According to Document 3 - Source: LoRA.pdf, Page: 11:

"LoRA achieves comparable performance to full fine-tuning while drastically reducing computational costs."

However, the definition of LoRA itself is not explicitly stated. But based on the context and related information in the document, we can infer that LoRA stands for "Low-Rank Adaptation", which is a method used for adapting large language models to downstream tasks with reduced computational costs.

A more detailed explanation is provided in Document 2 - Source: LoRA.pdf, Page: 9:

"By freezing the majority of the model's parameters and training only small, low-rank matrices, LoRA achieves comparable performance to full fine-tuning while drastically reducing computational costs."

View 3 Sources

1. app_buid_paper.pdf (Page 10)

```
Quoc Le, and Charles Sutton.

Program synthesis with large language models, 2021.

URL https://arxiv.org/abs/2108.07732.

Valerii Babushkin and Arseny Kravchenko.

Machine Learning System Design with End-to-End Examples.

Manning Publications, 2025.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Mic…
```

2. LoRA.pdf (Page 9)

```
6 Related Works 6.1 Transformer Language Models The Transformer architecture, as introduced by Vaswani et al.

(2017), has proven to be a highly effective sequence-to-sequence model due to its heavy use of self-attention mechanisms.

Radford et al.

(2018) applied it to autoregressive language modeling, significantly boosting its utility in the field.

Since then, Transformer-based models have become a staple in natural language processing (NLP),…
```

3. LoRA.pdf (Page 11)

```
pting large language models for downstream tasks.

By freezing the majority of the model’s parameters and training only small, low-rank matrices, LoRA achieves comparable performance to full fine-tuning while drastically reducing computational costs.
```