Welcome to the “Connecting the Dots” Challenge

Rethink Reading. Rediscover Knowledge

What if every time you opened a PDF, it didn’t just *sit there*—it *spoke to you*, *connected ideas*, and *narrated meaning* across your entire library?

That’s the future we’re building — and we want **you** to help shape it.

In the **Connecting the Dots Challenge**, your mission is to reimagine the humble PDF as an intelligent, interactive experience—one that **understands structure**, **surfaces insights**, and responds **to you** like a trusted research companion.

The Journey Ahead

- **Round 1:**

  Kick things off by building the brains — extract structured outlines from raw PDFs with blazing speed and pinpoint accuracy. Then, power it up with on-device intelligence that understands sections and links related ideas together.

- **Round 2:**

  It’s showtime! Build a beautiful, intuitive reading webapp using Adobe’s PDF Embed API. You will be using your Round 1 work to design a futuristic webapp.

Why This Matters

In a world flooded with documents, what wins is not more content — it’s **context**. You’re not just building tools — you’re building the future of how we **read, learn, and connect**. No matter your background — ML hacker, UI builder, or insight whisperer — this is your stage.

Are you in?

It’s time to read between the lines. Connect the dots. And build a PDF experience that feels like **magic**. Let’s go.

Round 1A: Understand Your Document

Challenge Theme: Connecting the Dots Through Docs

Your Mission

You're handed a PDF — but instead of simply reading it, you're tasked with making sense of it like a machine would. Your job is to extract a structured outline of the document — essentially the **Title**, and headings like **H1**, **H2**, and **H3** — in a clean, hierarchical format.

This outline will be the **foundation** for the rest of your hackathon journey.

Why This Matters

PDFs are everywhere — but machines don't naturally understand their structure. By building an outline extractor, you're enabling smarter document experiences, like semantic search, recommendation systems, and insight generation.

What You Need to Build

You must build a solution that:

- Accepts a **PDF file** (up to 50 pages)
- Extracts:
  - **Title**
  - **Headings:** H1, H2, H3 (with level and page number)
- Outputs a valid JSON file in the format below:

```
{
  "title": "Understanding AI",
  "outline": [
    { "level": "H1", "text": "Introduction", "page": 1 },
    { "level": "H2", "text": "What is AI?", "page": 2 },
    { "level": "H3", "text": "History of AI", "page": 3 }
  ]
}
```

You Will Be Provided

1. A **sample input PDF** (e.g., sample.pdf)
2. A **sample ground truth output** (sample.json) for format clarity