# Value AI Labs – Take-Home Assignment

Congratulations on being shortlisted for the next stage. This take-home task is designed to simulate the kind of problems we work on — real-world, ambiguous, and open to creative interpretation.

Please **choose one** of the two challenges below and submit your solution by the timeline agreed in the meeting.

Please submit the scope of work by agreed "scope of work timeline". The scope of work should include:

- Wireframe
- What will be in scope
- What will not be developed

Feel free to use any stack, platform, or GenAI tool of your choice.

## Option A: Reconstruct the Jumbled PDF

### Problem

You've been handed a multi-page scanned PDF where the pages are **randomly shuffled** — a surprisingly common issue when working with documents in the mortgage and lending industry.

Your task is to build a tool that can **analyze and reorder** the pages to restore the document to its likely original sequence.

### What You Should Build

An app that:

- Accepts a jumbled PDF as input
- Analyzes page content using any methods (OCR, LLM, embeddings, etc.)
- Outputs a reordered PDF
- Optionally provides logs, reasoning, or natural language notes explaining the page order

### Bonus Points

- Add a **Table of Contents** based on detected sections
- Identify **missing or duplicate pages**

# Option B: Redact PII from Text and PDFs

## Problem

Responsible AI often starts with protecting sensitive data. Your task is to build a mini tool that **detects and redacts Personally Identifiable Information (PII)** such as Email addresses, Phone numbers,Names,Addresses from either **plain text** or **PDF files**.

## What You Should Build

An app that:

- Accepts text or PDF input
- Detects and redacts PII (use regex, LLMs, hybrid logic)
- Displays both original and redacted outputs
- Offers clear documentation or reasoning on what was removed

## Bonus Points

- Allow toggling redaction types (e.g., redact emails but not names)
- Show redacted tokens as `[EMAIL_1]`, black boxes, or user-defined labels
- Export the redacted output (PDF/text) via download

# Submission Instructions

- Add your test files: We have provided two input files for you to test your solution - Loan-Agreement-Shuffled-Scanned.pdf, Please add more input test files as appropriate.
- Share a **GitHub repo** and a short `README.md` explaining:
  - What you built and why you chose that approach
  - Any assumptions, limitations, or trade-offs
  - What you'd improve with more time
- Include a short **Loom video (2–3 mins)** walking through your solution
- **Due Dates:**
  - **Scope of work: 24 hrs**
  - **Final submission: 3 days from Scope of Work**

# Evaluation Rubric

| Criterion | What we are looking for |
|---|---|

| Clarity of thought | Can you frame and structure a messy problem clearly? |
|---|---|
| Creativity | Do you go beyond basic implementations? Are your design choices thoughtful? |
| Usage of Gen AI coding agents | Are you using them wisely and taking ownership for the code being generated? |
| Presentation | Is your tool usable? Does it look and feel like a mini-product? |
| Effort and ownership | Did you explore the problem deeply? |

We're not looking for a polished product — we're evaluating how you think, explore, and build with curiosity.