Technical Assesment

AI/ML Engineering
Technical Assessment

Company Context

You are working with a mid-market private equity firm managing approximately £10 billion in assets across multiple funds. The firm evaluates thousands of potential investment opportunities annually, with a team of analysts responsible for screening incoming deals against the firm's investment criteria.

When a new opportunity arrives, analysts receive a Confidential Information Memorandum (CIM) – a document prepared by investment bankers that describes the target company, its financials, market position, and growth prospects. These documents are typically 50-100 pages and arrive as PDFs.

The Problem

The firm has been investing for over two decades and has accumulated a substantial history of evaluated deals – both those they invested in and those they passed on. This institutional knowledge currently exists across:

- A CRM system containing deal metadata (sector, size, outcome, key metrics)

- Document repositories holding historical CIMs, investment committee memos, and analyst notes

- The memories of senior partners (largely inaccessible and inconsistent)

When a new CIM arrives, analysts often ask: 'Have we seen anything like this before? What happened? What did we learn?'

Currently, answering this question requires manual searches through the CRM, asking colleagues who might remember similar deals, and keyword searches through document repositories (largely ineffective).

The firm wants to build an AI-powered system that can automatically surface relevant historical deals when a new opportunity is being evaluated.

The Challenge

Design and describe a system that can identify 'similar' historical deals to a new incoming opportunity.

The core technical challenge is deceptively simple to state but difficult to solve:
What makes two deals 'similar'?

Consider:

- Two companies might have identical financial profiles but operate in completely different markets

- Two companies in the same sector might have fundamentally different business models

- A deal that was attractive in 2019 might be unattractive in 2024 due to market conditions

- The reasons the firm passed on a deal often contain critical insights not captured in structured data

What We Know

The firm has access to:

Structured CRM data for historical deals including:

- Deal outcome (Invested, Passed, Lost to competitor)
- Basic financials (Revenue, EBITDA, growth rates, margins)
- Sector/sub-sector classification
- Deal size and type
- Notes fields (unstructured, inconsistent quality)

Historical documents including:

- Previous CIMs (PDFs, varying quality)
- Investment committee memos
- First visit forms (initial analyst assessments)

New deal inputs:

- The incoming CIM (PDF)
- Potentially some banker correspondence

Your Task

Prepare a technical proposal that addresses:

Part 1: System Architecture (30%)

Describe at a high level how you would architect a system to solve this problem. Consider:

- How would you represent deals in a way that enables similarity comparison?
- What embedding or representation strategies might you employ?
- How would you handle the multi-modal nature of the data (structured metrics + unstructured documents)?

Part 2: The Similarity Problem (40%)

This is the core intellectual challenge. Address:

- How would you define and measure 'similarity' between deals?
- How would you handle the fact that similarity is context-dependent (similar for what purpose)?
- What are the limitations of pure embedding-based similarity for this use case?
- How might you incorporate human judgment and feedback into the system over time?

Part 3: Practical Considerations (30%)

Discuss:

- How would you evaluate whether the system is working well?

- What would an MVP look like versus a production system?

- What are the key risks and failure modes?

- How would you handle cases where the system's notion of similarity diverges from what analysts actually find useful?

Evaluation Criteria

We are looking for:

- **Depth of thinking** – Do you identify the non-obvious challenges in this problem?

- **Technical sophistication** – Do you demonstrate understanding of relevant ML/AI techniques?

- **Practical judgment** – Do you balance theoretical elegance with real-world constraints?

- **Learning systems mindset** – Do you think about how the system improves over time?

- **Communication** – Can you explain complex technical concepts clearly?

Submission Format

- Written document (PDF or Markdown)

- Maximum 2,000 words

- Diagrams welcome but not required

- Code snippets optional (pseudocode acceptable)

Time Expectation

We expect this to take approximately 3-4 hours. We are not looking for a production-ready design – we want to understand how you think about ambiguous, open-ended AI problems.

Notes for Candidates

Some things to consider that may not be immediately obvious:

- The firm's analysts are the domain experts, not the AI system

- Historical deal outcomes are not ground truth for 'good' decisions
  (the firm passed on deals that later became very successful)

- The system needs to be trustworthy – analysts won't use a black box

- The firm evaluates deals across different sectors and geographies with different investment teams

We deliberately have not specified the 'right' answer because there isn't one. We want to see your reasoning process and technical creativity.