

INSEAD MBA Course: Building genAI Products and Business

The INSEAD MBA 2024J
Participants of [INSEAD MBA Course](#)
with the support of AI

March-April 2024

Abstract

This document highlights seven of the ideas discussed in the INSEAD MBA course, "Building genAI Products and Business" during the March-April 2024 period. It also describes some course group projects, highlighting "the 2024 state of AI driven class projects in an MBA program". These are projects that would have been unimaginable even a few years ago, and indicate both how far technology has gone and how far MBA education can go. And it is only 2024...

Introduction

Generative AI¹ is revolutionizing the landscape of technology by opening up unprecedented opportunities and enabling the creation of innovative products and services. For instance, AI-driven platforms are transforming artistic creation by allowing users to generate unique artworks from textual descriptions, while LLMs are redefining customer service with advanced, conversational chatbots that can handle inquiries and support with minimal human oversight. Additionally, AI in drug discovery accelerates the identification of viable new compounds, potentially reducing the time and cost of bringing new medications to market. However, these advancements also introduce significant risks, such as the potential for creating deepfakes that undermine trust in digital media, or biases in AI algorithms that could lead to unfair treatment in areas like lending or recruitment. As such, the rise of generative AI, and AI more broadly, is a double-edged sword, presenting both exciting possibilities and formidable challenges.

New² ventures are also emerging that utilize AI to automatically generate written content, from marketing copy to full-length articles, tailored to specific audiences and styles. In design and architecture, AI tools are enabling professionals to rapidly prototype and iterate on complex designs, significantly reducing the time from concept to creation. Additionally, in entertainment, generative AI is being used to create

¹ This paragraph was written by an LLM with the prompt "Can you write an opening paragraph about how generative AI is creating today new opportunities, enables new products and services, but also can lead to new risks, giving also some examples?"

² This paragraph was written by an LLM with the prompt "Can you write a paragraph describing a few examples of businesses and products that one can build using generative AI today?"

immersive and interactive gaming experiences, where environments and narratives adapt dynamically to players' actions. Another impactful application is in personalized medicine, where AI models generate customized treatment plans based on patients' genetic data. These examples showcase the vast potential of generative AI to not only enhance existing products and services but also to pioneer entirely new business models.

Some Ideas during the Course

These are 7 of the (many) ideas discussed during the course³:

1. Generative AI is revolutionizing language (image, video, speech, sound, etc) processing with its innovative approach of translating text into numerical vectors (called "embeddings"). This method is fundamental to the functionality of Large Language Models (LLMs), which operate by performing simple mathematical operations – like visualizing the concept of a "queen" (a vector of numbers, the "embedding" of the word "queen") by subtracting "man" (another vector of numbers) from "king" (a third vector of numbers) and adding "woman" (a fourth vector of numbers). This technique showcases the capacity of AI to comprehend and produce language with extraordinary accuracy. However, this advanced capability comes with an insatiable appetite for data, propelling the need to potentially train these models on synthetic data. As the demand for data grows, the shift towards synthetic datasets could become a necessary evolution to sustain and enhance the developmental trajectories of LLMs and AI more broadly, ensuring they continue to refine their linguistic prowess and expand their applicative horizons. Another critical consideration is gaining a robust grasp of the technical underpinnings that drive AI technologies, such as the role of human feedback to the AI output, the use of prompt engineering or RAGs, and the role of fine tuning. Understanding these underlying mechanisms allows for a more comprehensive integration of AI into business processes, unlocking new benefits. As LLMs become more widespread and potentially commoditized, distinguishing one's offerings and maintaining a competitive edge will require innovative approaches and deeper integrations in our businesses and everyday life.
2. Acquiring a deeper understanding of the foundational technical principles that drive AI is critical for managers to leveraging its extensive capabilities, for policy makers to identifying appropriate policies, and for everyone to become better (and more responsible) users of AI. One pertinent example is the integration of AI, specifically Large Language Models (LLMs), into cultural contexts. This is achieved through feedback mechanisms that allow these models to adapt and refine their outputs, demonstrating how AI can be tailored to specific cultural (corporate, social, national, etc) nuances and languages. Such integration achieved through usage of AI within a business can create lock-in effects and competitive advantage for both the providers

³ This section is written with the support of an LLM based on 7 key points made during the course. Prompts were bullet point format.

of AI based solutions and the companies using these solutions. As access to AI models becomes largely a matter of “an API call” (with some prompt engineering, fine tuning, a RAG, or other lighter tech modules), value is created not “simply” from the AI but from its integration – technical, workflow, process, culture, etc – of the AI in business. AI startups need to focus more on these integration capabilities to provide value and create competitive advantage. There exists a looming risk as these powerful LLMs become more commonplace and accessible – an AI commodification. This trend could dilute the distinct competitive advantage that such advanced technologies can provide, pushing businesses and developers to innovate continuously to stay ahead. The most significant advantages are realized in scenarios where AI can be fully integrated, seamlessly enhancing and extending human capabilities across various business processes, functions and industries.

3. Generative AI is dramatically transforming the unit economics of the software industry, primarily through its impact on power consumption and cost structures. Traditional software models typically involve costs that scale with user numbers, but generative AI may change this dynamic. The marginal cost of adding a new user approaches zero, thanks to the efficiency and scalability of AI systems. This shift is prompting a potential reevaluation of pricing models, moving from traditional pay-per-seat schemes to more dynamic pay-per-value approaches. Such a change not only aligns costs more closely with the actual value delivered but also allows companies to focus resources on solving the most profitable problems at present. This strategic shift can lead businesses to prioritize initiatives that offer immediate returns, enhancing short-term profitability and operational efficiency.
4. The technological landscape of generative AI (GenAI) is intricate and ever-evolving, characterized by a broad spectrum of applications and capabilities. At its core are foundational models designed to understand and generate text, which form the backbone of many AI-driven innovations. These models are further adapted to meet the specific needs of different industries, potentially leading to industry specific foundation models. These, in turn, are further customised for specific use cases, leading to a collection of foundation models – and eventually agents – that are specialists at the industry and use case level. Business and national AI strategies may evolve towards developing industry specific – and culture specific – foundation models, as well as marketplaces for use-case specific foundation models that can be easier adapted to specific contexts.
5. The choice between augmentation and automation represents a critical business decision, particularly in use cases where the 'human touch' may be highly valued. This distinction can significantly influence customer preferences and even pricing strategies. One may call this the "human hand-stretched pizza" effect: in the food industry, there could be a notable difference in customer perception and willingness to pay between an "AI automated pizza developer" and a "human hand-stretched pizza." Similarly in other use cases. In contexts where personalization and artisanal quality are prized, human involvement may command a premium, underscoring its potentially irreplaceable value. This preference acts as a counterbalance to the trend

towards increasing automation, highlighting the importance of human interaction and craftsmanship in certain sectors and use cases – at least for now. In making strategic decisions about implementing AI, businesses must carefully consider the trade-offs between efficiency gains through automation and the unique appeal – in addition to better performance and productivity in some cases – of human augmentation.

6. In the realm of technological solutions, it's crucial for businesses to critically evaluate whether AI is the most appropriate tool for addressing a specific problem. This assessment involves asking a fundamental question: "Is this problem really only (or at all) solvable with AI?" By posing this query, companies can avoid the pitfalls of deploying AI for its own sake, which can lead to unnecessary complexity or resource expenditure. Instead, this question prompts a thoughtful analysis of the problem at hand, encouraging solutions that are not only technologically sound but also cost-effective and strategically aligned with business goals. This approach ensures that AI is used where it genuinely enhances value, rather than being seen as a default universal solution for every challenge.
7. When businesses build products based on AI, especially using third-party foundational AI models, they must recognize that these products will inherently adopt many of the qualities, behaviors, and risks associated with those models. This integration means that the attributes of the AI – ranging from its decision-making processes to its potential biases – are transferred into the products themselves. Consequently, if the AI model is sourced from an external provider, significant aspects of the product's performance and ethical implications and risks may lie beyond the direct control of the business – or, at least, require additional effort to refine and control. This situation can introduce unpredictability and necessitates careful consideration of the chosen AI's origins, development practices, and underlying data. Companies must therefore implement robust oversight and mitigation strategies to manage these risks effectively, ensuring that their AI-driven products align with business values and regulatory requirements.

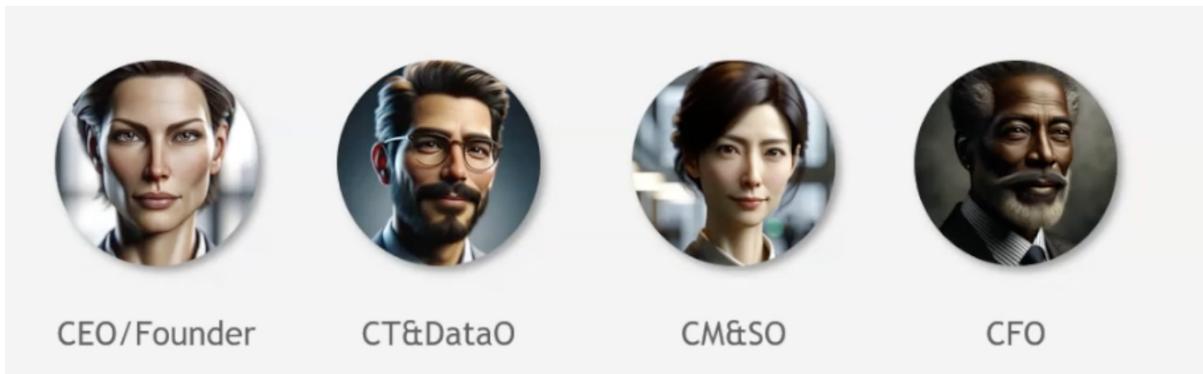
Course Projects and... the Future?

As part of the course requirements, participants had to develop a product and business idea and plan, including a product demo with code in python – with the help of both LLMs/AI and an engineer (from outside INSEAD) who was playing the role of a “Chief AI Officer” for the group. Almost half the groups decided to explore building an actual business around their project, hence not all projects are shared here. A short collection of some of the course projects, with links to demos and slides for some of them, are included in the appendix.

All groups have developed product demos, some of them being almost MVPs ready to be launched as businesses. All this was done within less than eight weeks, indicating the power of today's tools. This would have really been unthinkable even a few years

ago. It has been a long journey since even about twelve years ago when we were using [R code and markdown for data science courses](#) in the INSEAD MBA program.

Perhaps in a few years the course projects will be even closer to full blown products and businesses, including having virtual executive teams with each member of the team being an AI agent with their own personality, skills, and roles, coded as part of the course project and “hired” in the business.... One of the group slides (below) may be telling of what may be coming next.



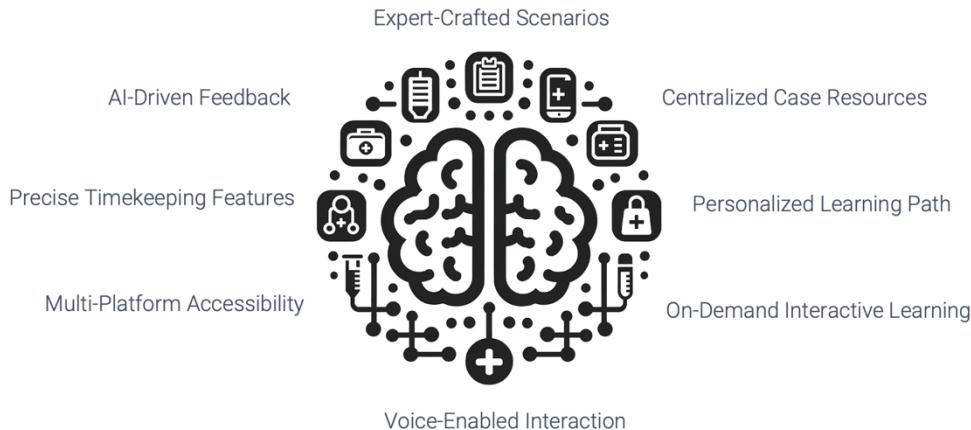
Appendix: Sample Course Group Projects

Project: CaseAI

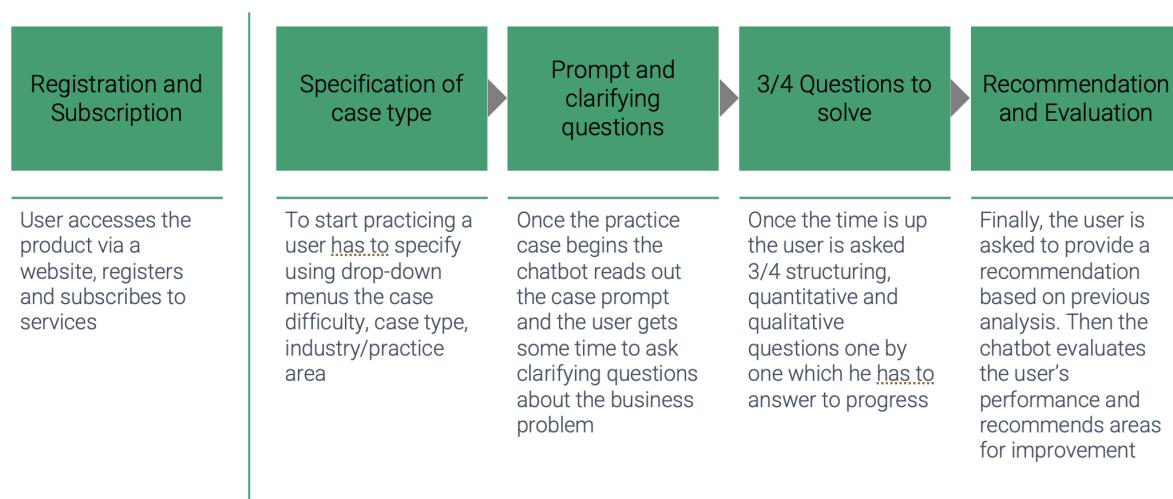
CaseAI is an innovative solution to case interview preparation. CaseAI leverages generative AI and voice recognition software to simulate management consulting interviews end-to-end, making it easier than ever for candidates to perfect their problem-solving skills.

[Slides](#); [Demo](#)

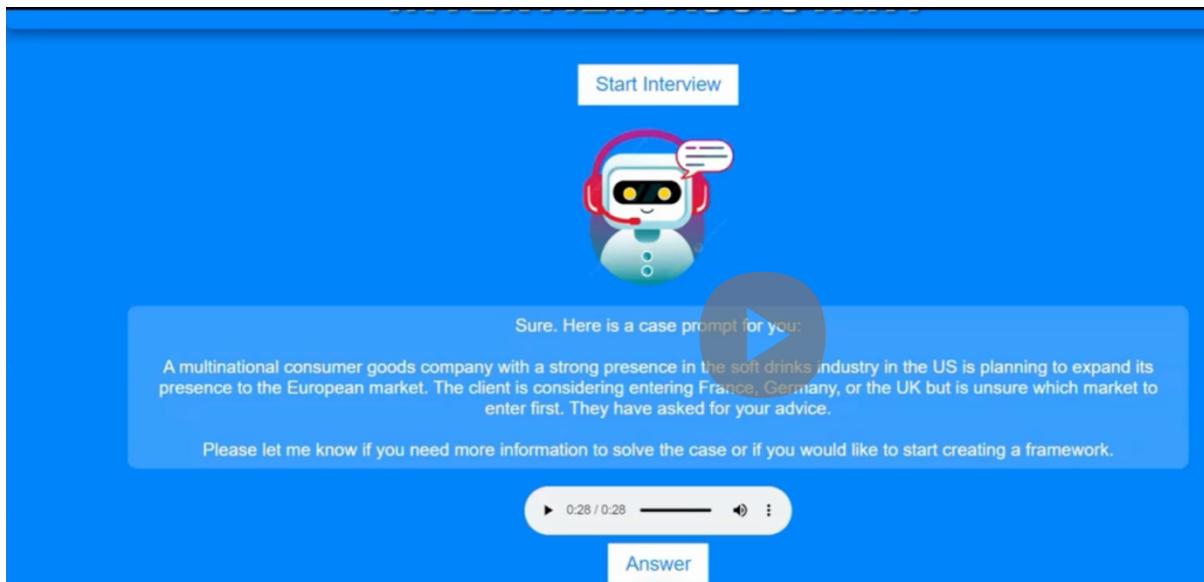
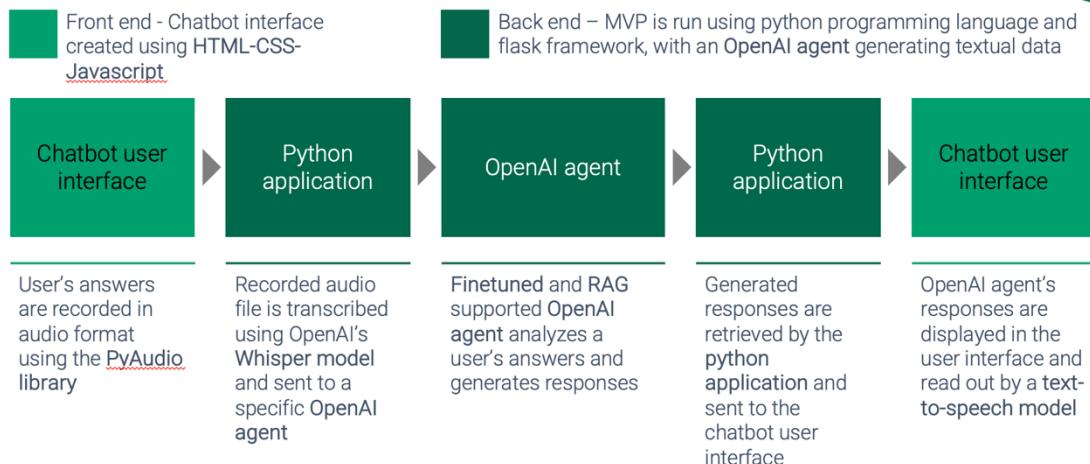
CaseAI can help students take the first step in their consulting case preparation efforts



CaseAI will cover all aspects of a typical consulting case from the prompt to the final recommendation

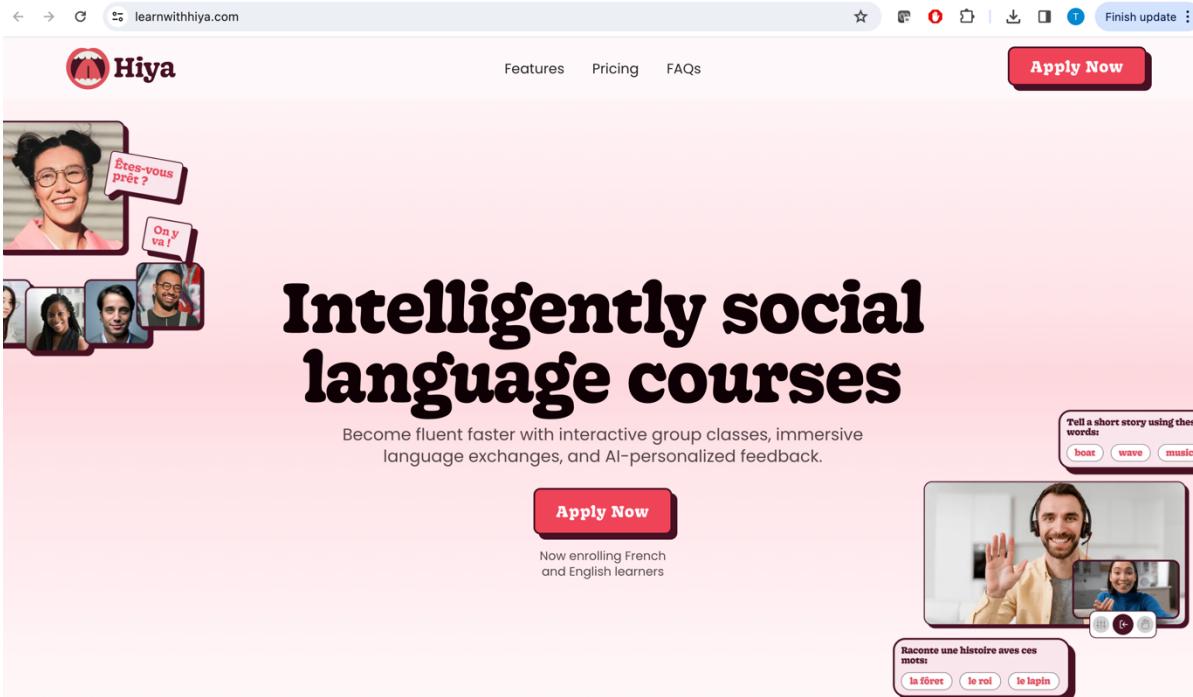


Our simple technology stack ensures reliability



Project: Hiya (<https://www.learnwithhiya.com/>)

Become fluent faster with interactive group classes, immersive language exchanges, and AI-personalized feedback.



learnwithhiya.com

Hiya

Features Pricing FAQs

Apply Now

Intelligently social language courses

Become fluent faster with interactive group classes, immersive language exchanges, and AI-personalized feedback.

Apply Now

Now enrolling French and English learners

Tell a short story using these words:
boat wave music

Raconte une histoire avec ces mots:
la forêt le roi le lapin

We're putting the people back into language learning

 **Interactive Group Lessons**

Join a cohort taught by one of our world-class language instructors.

 **Immersive Language Exchanges**

Connect with native speakers in peer cohorts for authentic language practice.

 **AI-Personalized Feedback**

Let us track your progress and deliver personalized feedback and exercises.

Product Demo ✨

Transcript

Dorianne : Nous avons lu l'histoire de Sophie et Julien, ils vont voyager la semaine prochaine. Vous pouvez me parler de l'histoire?

Riley : Oui. Julien et Sophie partent en voyage et font la réservation des tickets. Ils achètent les tickets d'avion pour aller à Berlin.

Dorianne : oui, ils font la réservation des billets d'avion. - le verbe faire avec les pluriel se dit ils font. Aussi, on utilise le mot "un billet d'avion" et le mot ticket pour un reçu, par exemple un "ticket de caisse"

Feedback

You have committed an error typical of learners at the A2 level. Here you have incorrectly conjugated the verb faire in present tense. Rather than faisent, the pronoun ils takes font. As an irregular verb, faire can trip students up especially if they haven't reviewed it in a while!

Here are some practice exercises for the verb faire in present tense:

1. Julien (faire) la liste des choses à emporter pour le voyage.
2. Sophie (faire) une réservation pour un restaurant à Berlin.
3. Ils (faire) leur enregistrement en ligne avant de partir.

[Add to Riley's Profile](#)

What we've done so far



Launched landing page



Conducting two pilot cohorts



Built Minimum Viable Product

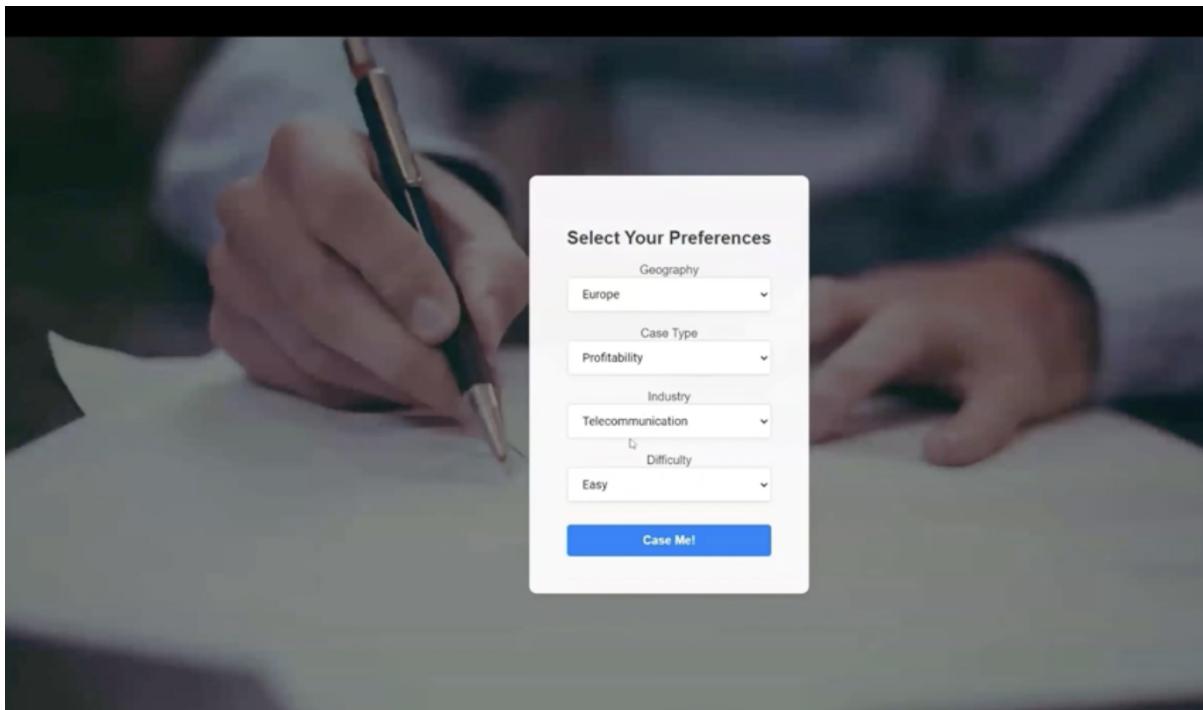
Project: CaseMe.ai

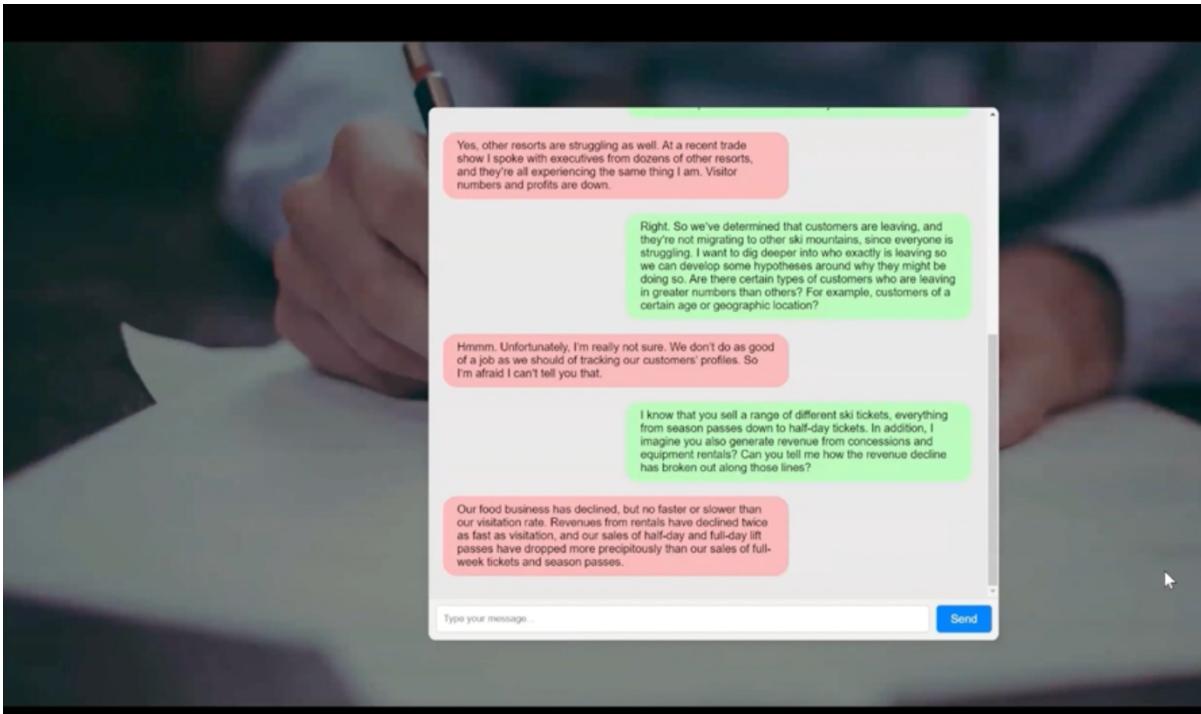
CaseMe.ai is your most reliable and complete case partner. From now on, you no longer need to send invites to all your consultant friends. For companies, CaseMe.ai will help you get to know your applicants even before you met them.

CaseMe.AI's features offer a bespoke experience for both B2C and B2B segments



Target users/ Use cases	Product features	High-level tech specs
<p>B2C</p> <p>1 Job seekers, including undergraduates, MBAs and career transitioners</p>	<ul style="list-style-type: none"> ▪ Cases on multiples industries & geographies ▪ Different types of cases ▪ Different skills tested ▪ Varying levels of difficulty ▪ Qualitative and quantitative feedback for both improvement areas and elements to maintain 	<ul style="list-style-type: none"> ▪ Voice and image recognition for a more realistic experience ▪ Web-based to begin with (MVP), with a proprietary app created in later stages ▪ +50 main languages (MVP will be in English)
<p>B2B</p> <p>2 Companies who want to optimize their recruitment processes through:</p> <ul style="list-style-type: none"> ▪ Initial screening phases, before first official round of interviews ▪ Interviewer bias elimination 		





Our solution stands on solid technical foundation and architectures



TECHNICAL ARCHITECTURES

- The frontend and backend will be constructed using the FLASK framework, providing fast and lightweight web application development
- The backend will interface with OpenAI's GPT-4 and WhisperAI, as well as other AI models and tools that may be integrated in the future.

MODEL TRAINING

- The application of RAG will enable AI to generate answers not just based on built-in knowledge but also by querying external data sources, enhancing the quality and relevance of answers.
- During training, the model will be fine-tuned using user interaction data to improve the response quality
- As time progresses, user feedback and interview outcomes will be used to further optimize the model, making it more precise in simulating the interviewers in the consultancy industry.

DATA PIPELINE

- User-input information through the interface will be transmitted via the FLASK backend and securely stored in the database.
- This information will be used to **customize personalized interview questions**, with answers being generated in real-time and displayed to the user through the frontend.

TOOLS & METHOD

- In the initial phase, OpenAI's API will be used to integrate GPT-4, providing language understanding and generation capabilities.
- Speech recognition** will be implemented through WhisperAI to support user voice inputs.
- The implementation of RAG will require an open-source vector database (like Milvus) to enable the AI model to efficiently retrieve and reference large amounts of information.

13

Project: DiscoverEase

DiscoverEase is a legal-tech tool that leverages AI to streamline the discovery process in litigation, initially targeting both contingency-fee law firms and traditional hour-billing

firms as customers. DiscoverEase uses topic modeling and a user-friendly interface, and can be fed case-specific technical data before a case for a more customized approach to document sorting and analysis. Our future hope is to create a suite of products for other legal-tech use cases.

DiscoverEase's business model centers around offering discovery services to legal associates and generates revenue by a classical SaaS model

Product

- Customized Dashboard per Case. The Product is built locally on Mistral 7 using Hugging Face embedding models.
- Intelligent Document Manager. Automatically classify, prioritize, and flag client's legal documents for each of their cases.
- Interactive ChatBot. The client can engage with a smart chatbot, designed to provide immediate, informed responses from each case dataroom.
- Data Privacy and Safety. We will ensure total confidentiality for the clients' data which will be siloed and accessible only by the client's team.

Customers

- Three customer groups are targeted to place DiscoverEase:
 - Contingency fee litigators
 - Traditional Law Firms (billable hours)
 - Corporate in-house teams

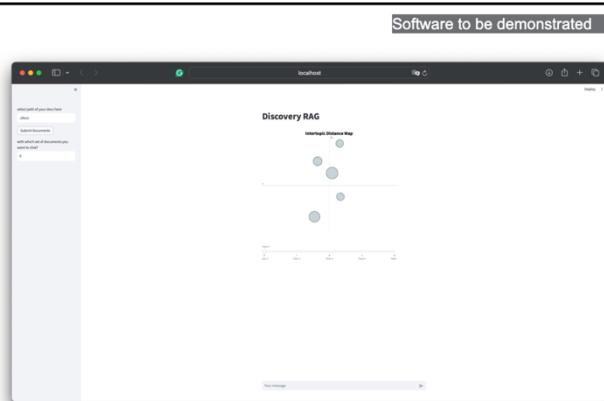
Economics

- Revenues:** Classical SaaS model with main revenue stream will be subscription fees from legal firms
- Gross Margin:** 60% at steady state
- Main Cost Items:** Salary of software developers and marketing/sales.
- Fundraising:** \$500K for 30% of the company.

See details on next slides

DiscoverEase assists in the discovery phase by 'clustering' and 'prioritizing' and 'flagging'

Demo version of DiscoverEase:



Source: DiscoverEase Demo version

Key features of DiscoverEase:

Data Inputs:

- Supports the **upload of all data relevant to a lawsuit**, incl. receipts, emails, documents
- Allows for the **addition of background information** about specific topics, such as short-term training, to enhance contextual understanding.

Advanced Outputs:

- Clustering and Tagging of Material:** Utilizes AI to intelligently **organize uploaded documents** into clusters, making navigation easier. Automatically tags materials for quick retrieval and reference.
- Prioritization of Documents:** **Analyzes and ranks** documents based on their relevance and importance to the case, ensuring that critical information is accessed first.
- Flagging of Key Findings:** **Highlights and flags** essential insights or discrepancies within the data, drawing attention to potential key findings in the lawsuit.

Project: PROcure

PROcure revolutionizes B2B software procurement by connecting SaaS buyers, primarily SMEs with limited IT procurement knowledge, with providers facing high

customer acquisition costs. Our platform leverages genAI to streamline procurement and sales management, mitigating integration failures and reducing the risk of new funding shortages for SaaS companies.

PROcure is a **platform** for **SaaS buyers** and **sellers** to simplify procurement and manage sales efficiently

Platform
connecting buyers and sellers of SaaS solutions from two different worlds

SaaS buyers
mostly concentrated in SMEs with limited IT procurement specific knowledge that lead to integration failures

SaaS sellers
experiencing high customer acquisition costs that reduce likelihood of survival in the current tight market for new funding

https://colab.research.google.com/drive/1jht06sdaMim0hZ_NMQ4855f6GK4Tn2MD#scrollTo=Uc2BSH_L9tog

SoftwareSelect.ipynb

```
import time

while run.status != "completed":
    run = openai.beta.threads.runs.retrieve(
        thread_id=thread.id,
        run_id=run.id
    )
    time.sleep(2)
    #print(run.status)

thread_messages = openai.beta.threads.messages.list(
    thread_id=thread.id
)
#print(thread_messages.data[0])
for message in thread_messages.data:
    for content in message.content:
        print(content.text.value)

*** What are your needs? Non-sql relational database manager
```

Disk 83.16 GB available [16] #update assistant to respond to questions

The screenshot shows a Google Colab notebook titled "SoftwareSelect.ipynb". The left sidebar displays a file tree with sections like "Documentation" containing PDFs for Amazon services (DynamoDB, EC2, Lambda, S3) and Neo4j, and a "sample_data" folder. The main workspace contains two code cells. The first cell's output is a block of text asking for needs regarding a non-SQL relational database manager, based on provided documents. The second cell's output lists two options: "Amazon DynamoDB" and "Neo4j", each with a brief description of its features. The bottom cell is partially visible, showing an update command for an assistant.

```
#print(thread_messages.data[0])
for message in thread_messages.data:
    for content in message.content:
        print(content.text.value)

What are your needs?Non-sql relational database manager
Based on the documents provided:

1. **Amazon DynamoDB**
   - **Price Range:** Pricing details not explicitly provided in the document, but it operates on a pay-as-you-go model.
   - **Reasoning:** Amazon DynamoDB is a highly available and durable NoSQL database service that supports key-value, wide-column, and document-oriented data models.

2. **Neo4j**
   - **Price Range:** Pricing details not explicitly provided in the document, but it offers both a Community edition and a Enterprise edition.
   - **Reasoning:** Neo4j is a graph database management system known for its efficiency in handling highly connected data structures.

[16] #update assistant to respond to questions
openai.beta.assistants.update(
    assistant_id=assistant.id,
    instructions= "You are a technical expert in a software specified by the user. The user will provide you with a task or question, and you must respond with a detailed answer or solution." )
```

The screenshot shows the Neo4j Browser interface. The left sidebar displays a file tree with several PDF files under the 'Documentation' folder. The main area shows a code editor with Python code for printing thread messages and a list of questions about Neo4j integration and performance.

```
#print(thread_messages.data[0])
for message in thread_messages.data:
    for content in message.content:
        print(content.text.value)
```

1. What are the Python integration options available?
Neo4J can be integrated with Python using the Neo4j Python driver, which is officially supported by Neo4J. This

2. Can you provide details about the performance benchmarks for typical use cases especially for large datasets
Details about performance benchmarks for Neo4J, particularly for large datasets and complex queries, are usually

3. Can you explain the theory behind graph databases used by Neo4J?
The theory behind graph databases, including Neo4J, is based on graph theory from mathematics. Graph databases

4. Can Cypher be extended with own user defined functions?
Yes, Neo4J allows Cypher, its query language, to be extended with user-defined functions (UDFs) and procedures.