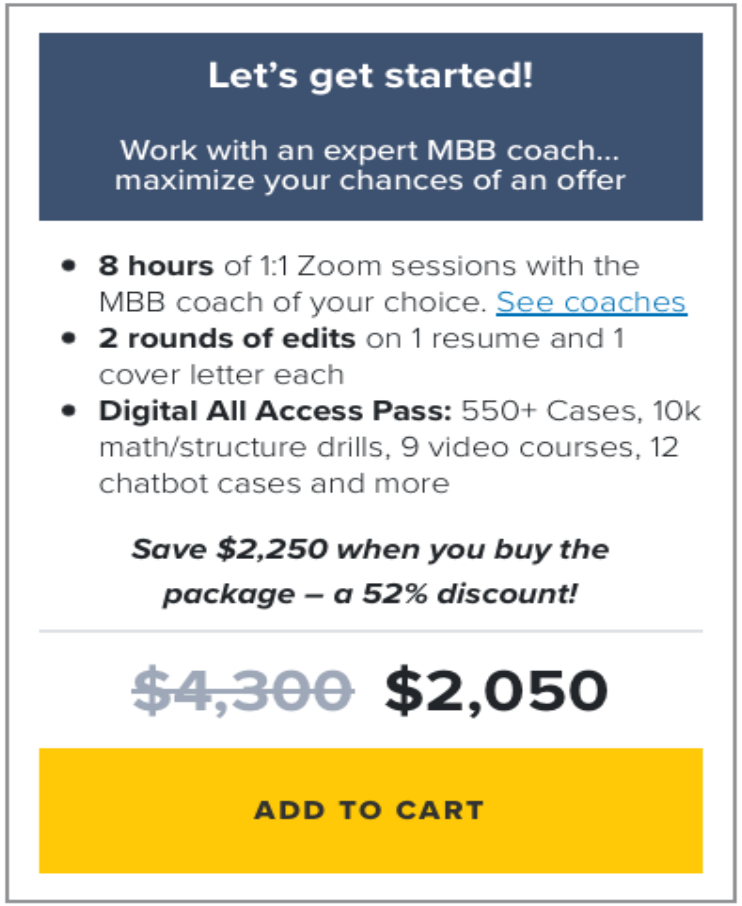
Motivation

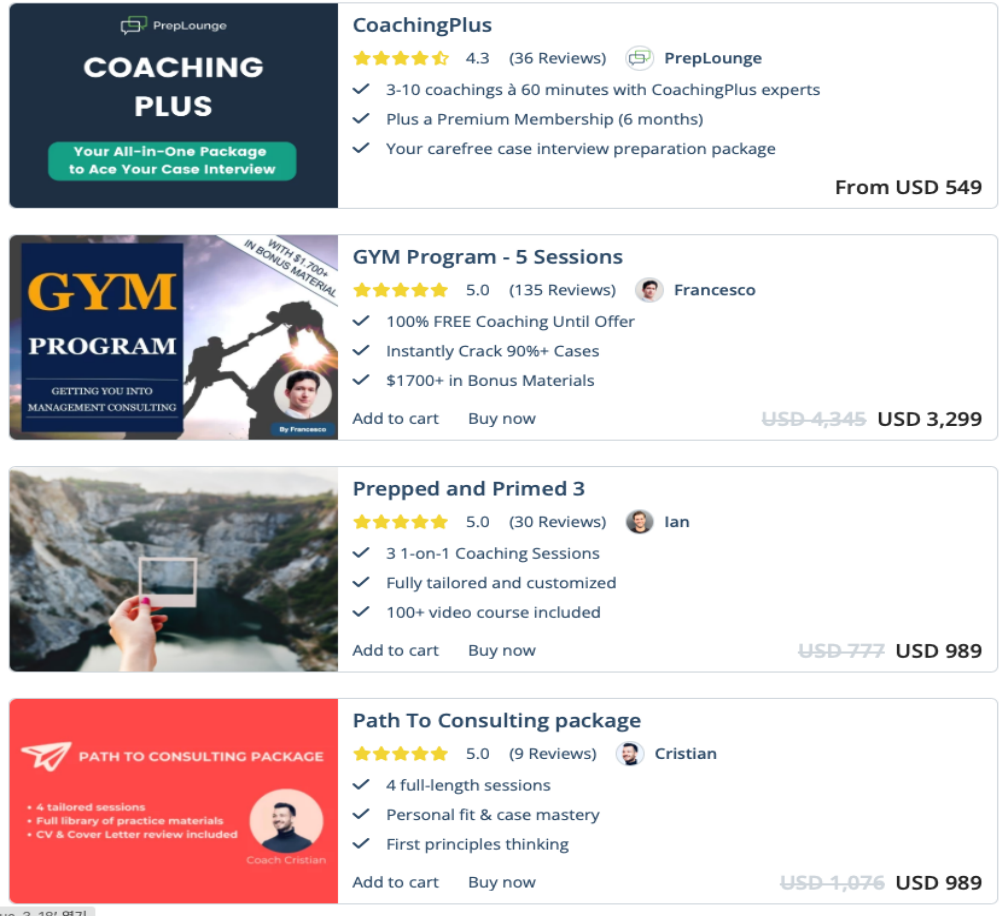
Why Consulting Case ChatBots?

Interviews are always **NERVOUS!**

Gap in the Market

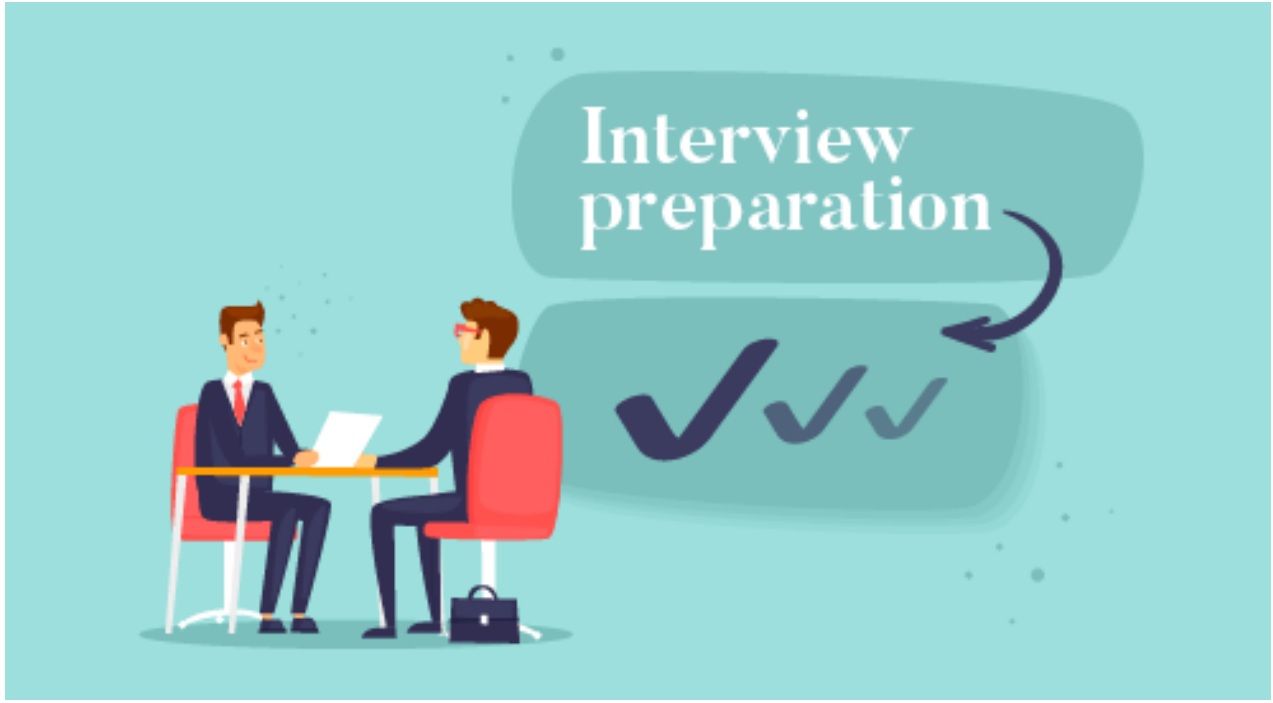
- Current price of mock interviews are **overly expensive**
- More than **$2000** with a **50% discount**!





Scope of the product

- Specifically, **Business consulting case interviews**
- Which provides **interactive** interviews of **specific case studies**
- Requires **knowledge of the field**



Goal & Target Customers

- Want to build an **interactive chatbot** that can practice mock interviews and provide evaluations and feedback
- Target Customer: **Job candidates preparing for business case interviews**

Cacey the ChatBot:
An Interactive Business Consulting Interview Assistant

Team 7. Lindenau Tim / Nielsen Sigurd Frank Thorlund / Jihoon Han / Donghwa Kim / Jaehoon Hahm /

Members

Jihoon Han

Product Owner

Tim Lindenau

Architecture

LLMs

Nielsen Sigurd Frank Thorlund

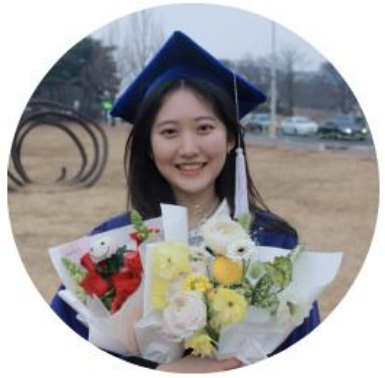
UX/UI

Evaluation

Jaehoon Hahm

MLOps

ML-Expert



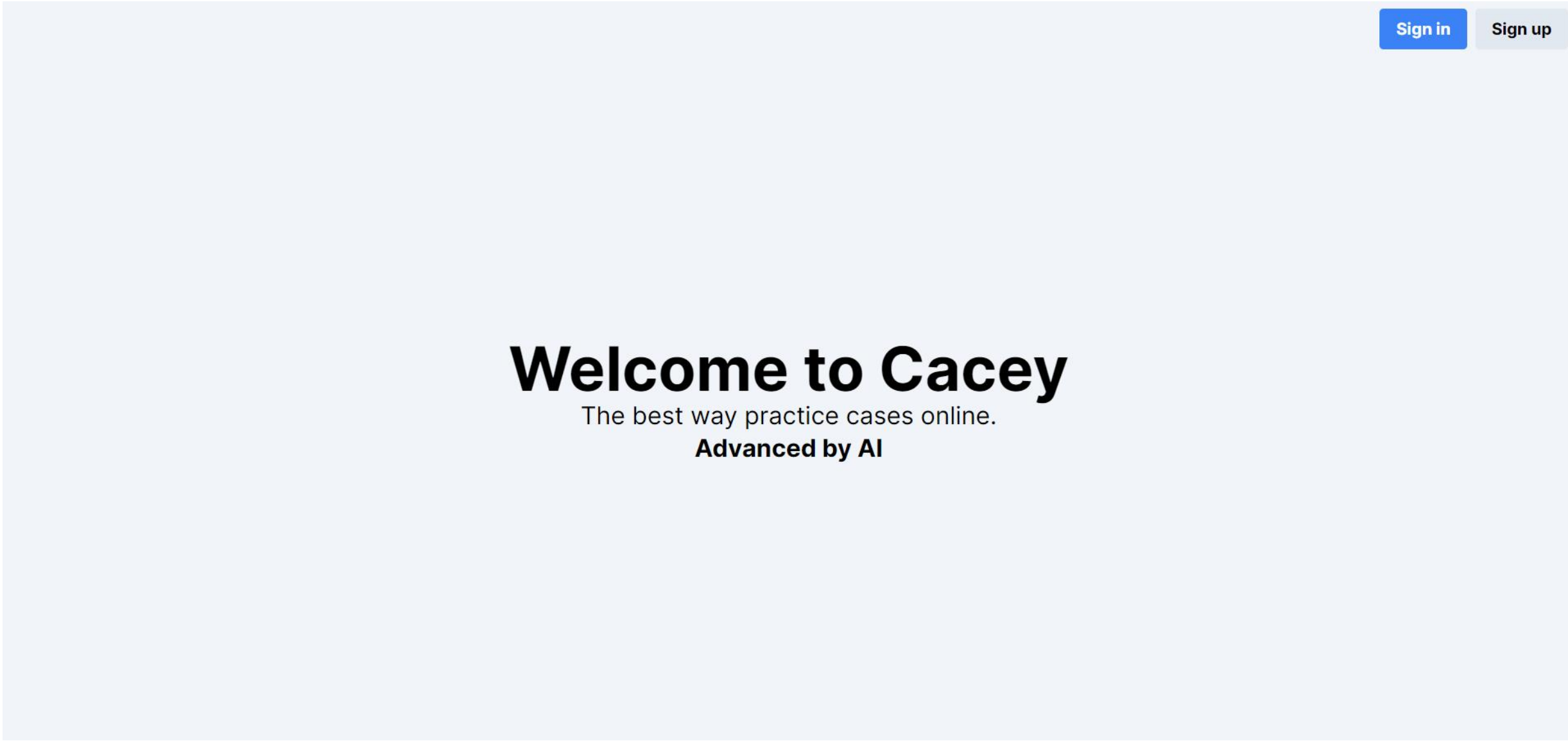
Donghwa Kim

MLOps

ML-Expert

Live Demo

Live Demo


Sign in
Sign up
Welcome to Cacey
The best way practice cases online.
Advanced by AI

Architecture

Three Factors when Deciding for an Architecture

1. Moving Fast

Similar Technologies in Frontend and Backend so **everyone can participate** everywhere

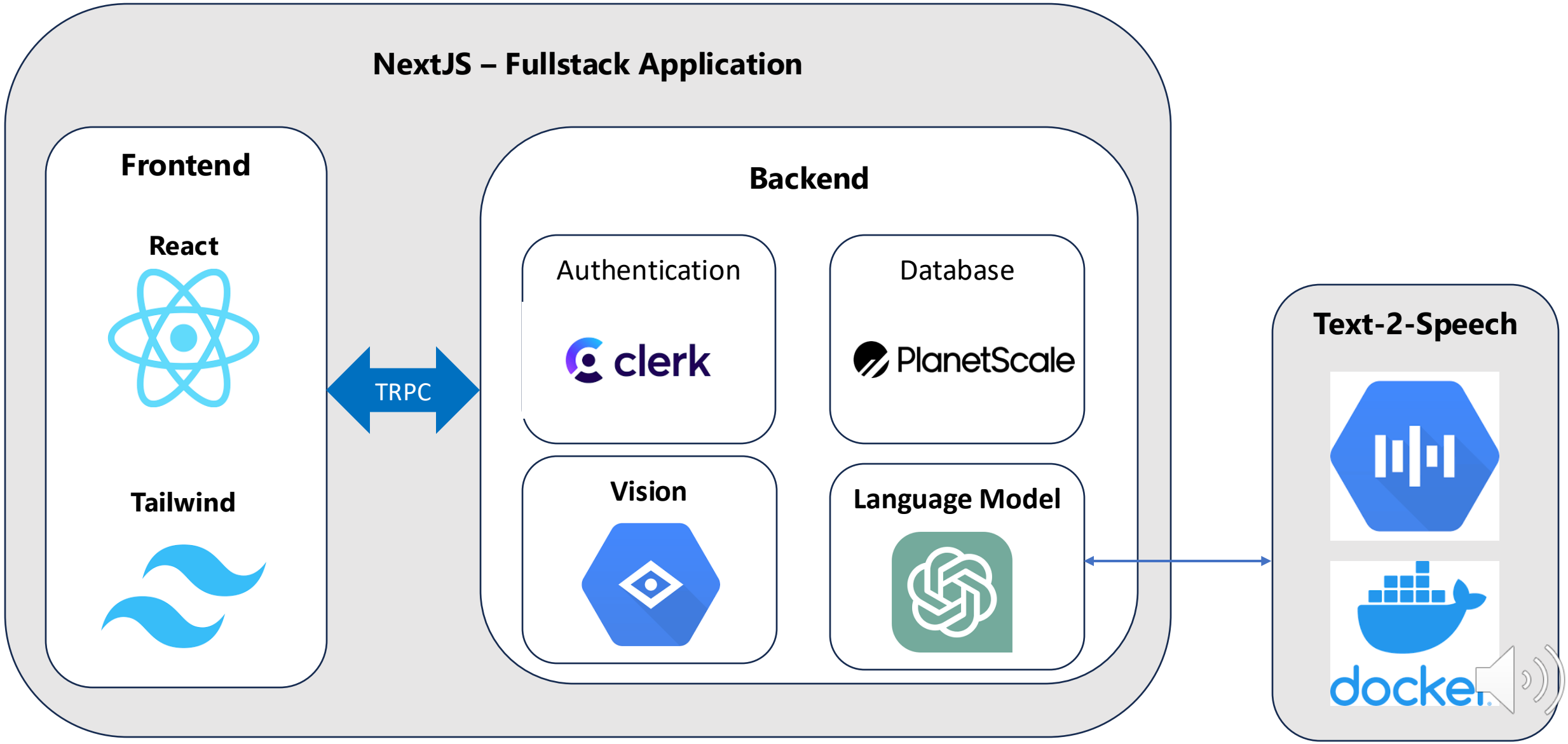
2. Developer Experience

Full **Typesafety** and automatically generated **Type** hints to prevent Bugs and support Developers

3. Costs

Low Prices and **Generous Free Tiers** to cheaply scale to the first thousand customers

Architecture Overview



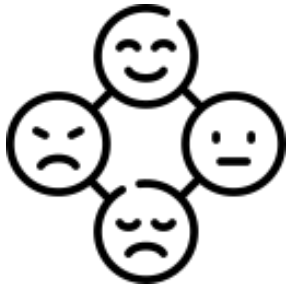
D

Main Features

4 Key Components For Successful Interview Training

1. Chatbot Interview Partner

2. Speech To Text



3. Image Guided Emotion Analysis

4. Case Evaluation

4 Key Components For Successful Interview Training

1. Chatbot Interview Partner

2. Text To Speech

3. Image Guided Emotion Analysis

4. Case Evaluation

Cases are Complex, You Cannot just use ChatGPT for the Whole Interview

Example Prompt Chatbot

"You are to take the role of an interviewer in a Case Interview for a Management Consulting Firm. …

The case is the following: A Company is trying to enter the market of ,…

One reference solution for the case could be the following:

1. *Framework*
    1. *Bucket 1: Market Attractiveness*
    2. *…*
2. *…*

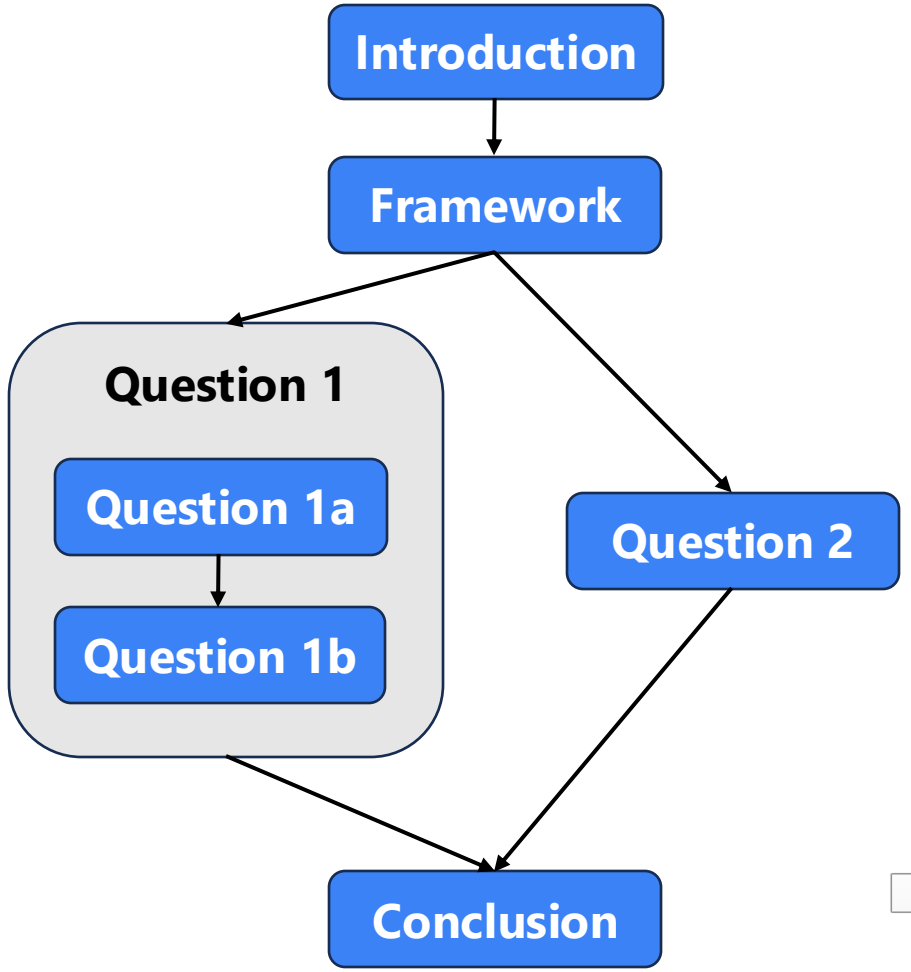
Start the case now:"

- Approach like this **does not work well**
- ChatGPT struggles with **long-term dependencies**
- ChatGPT has issues properly guiding to next structure points
- **High complexity in prompt tuning and missing control**

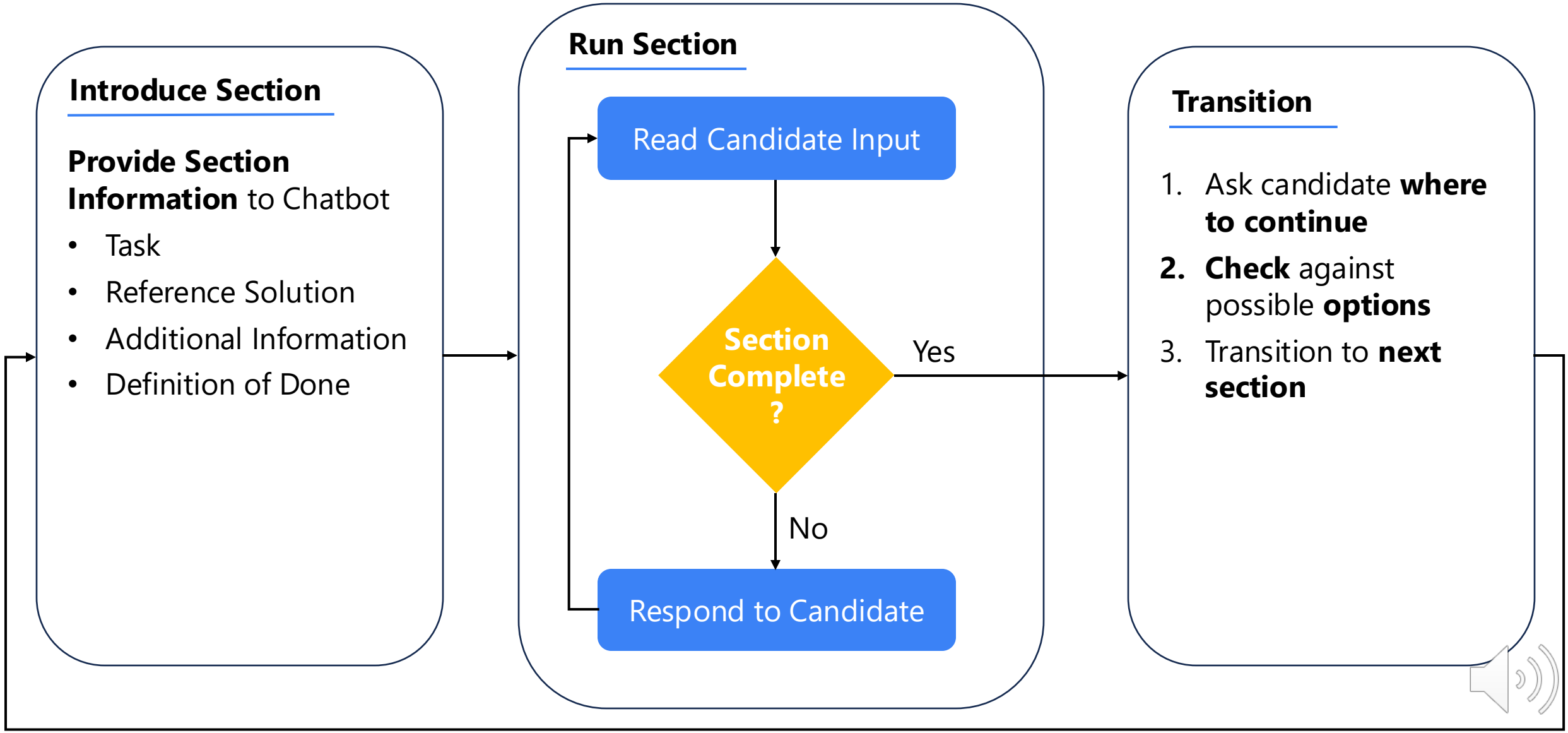
Cases are Complex, You Cannot just use ChatGPT for the Whole Interview

- From **solving more than 30 cases** we know that **each case follows a similar general structure**
- **Question:** Can we create a **state-machine** to follow this structure
- Introduce special **file format** to **parse arbitrary** cases to state-machine
- For each stage we can **individually instruct ChatGPT**
  - Provide additional information
  - Provide reference solution
  - Check whether section finished

Case Parsed into Structure



A State-Machine Helps to Reinstruct ChatGPT for Specific Sections



4 Key Components For Successful Interview Training

1. Chatbot Interview Partner

2. Speech To Text

3. Image Guided Emotion Analysis

4. Case Evaluation

Speech To Text Provides an Additional Modality To More Truthfully Imitate Interview Situation

Motivation why Speech

- Closer to real interview
- Analyze Speed, Clarity, (Sentiment)

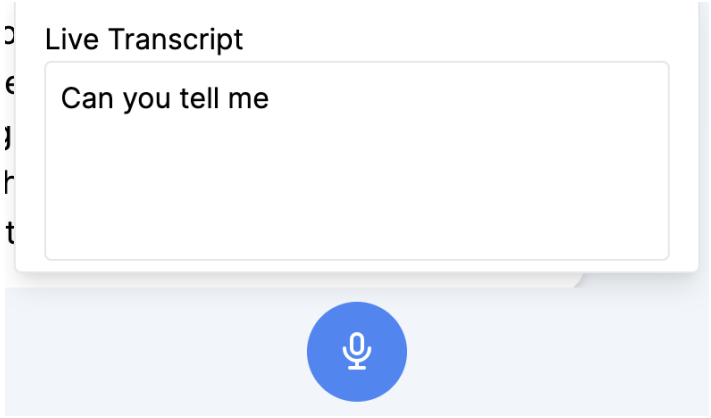
Our Solution:

- Decision for Google Speech due to **real-time streaming capabilities**
- Live Transcribe Text
- Compute **Speed** and **Clarity Scores**

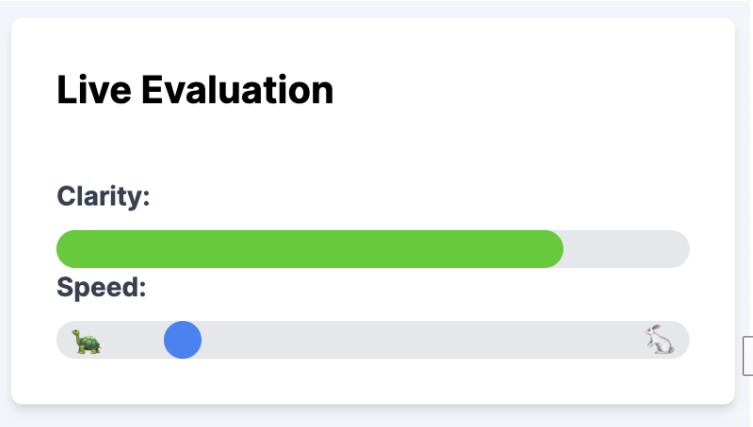
Challenge:

- Google Speech **requires File System**
- File System **not available** in Browser
- Build **additional service** to connect Browser and Google Speech

Live Transcription



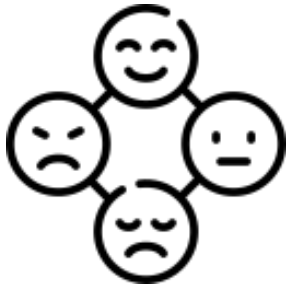
Live Evaluation



4 Key Components For Successful Interview Training

1. Chatbot Interview Partner

2. Text To Speech



3. Video Guided Emotion Analysis

4. Case Evaluation

Emotion Analysis (Video)

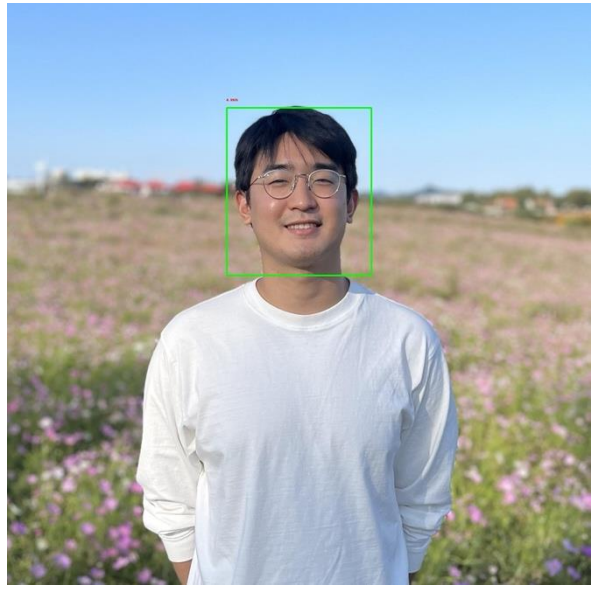
Possible Solutions for Emotion Analysis:

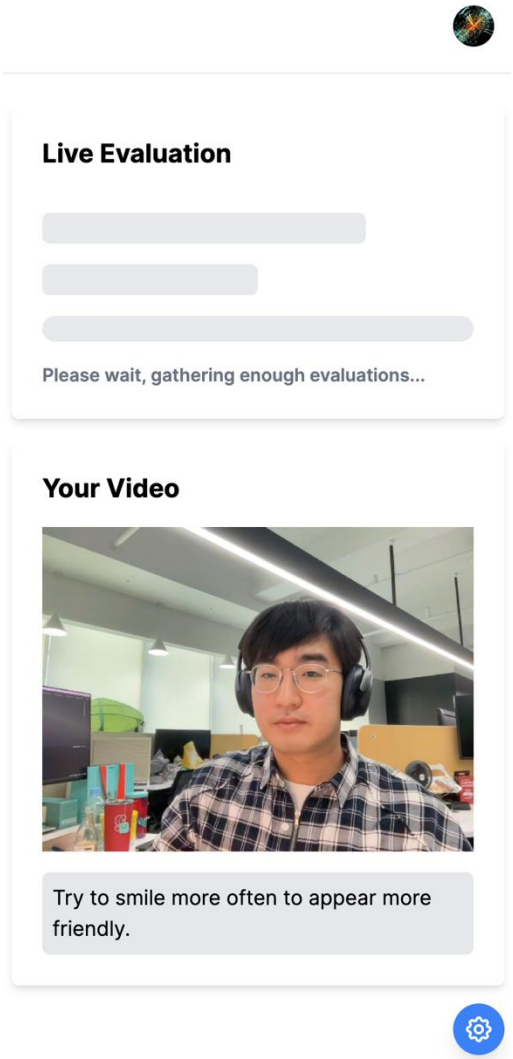
1. Pretrained machine learning model for emotion analysis
2. Google Cloud Platform: Video Intelligence API (expensive)
3. Google Cloud Platform: Vision AI, ImageAnnotator (convenient!) → **our chosen solution**

Method

1. Record frame every 10s for the input video.
2. Detect faces, analyze 4 emotions in likelihood scores.
3. Save it into emotion likelihood DB for every frame
4. Pass the average score to LLM for sentiment analysis

Likelihood	Evaluation Score
VERY_LIKELY	4
LIKELY	3
POSSIBLE	2
UNLIKELY	1
VERY_UNLIKELY	0





4 Key Components For Successful Interview Training

1. Chatbot Interview Partner

2. Text To Speech

3. Image Guided Emotion Analysis

4. Case Evaluation

Evaluating Cases Generates Value For Users

Purpose

- Evaluate candidate performance upon completing a case
- Value in feedback and pointers on how to improve

Challenges

- Evaluations are subjective judgements made on expert intuitions
- Each case may be evaluated differently
- No data available

Questions

- How can we introduce expert intuition in case evaluations?
- How do we make the framework general enough to work with all cases?

Prompt Templates Build On Expert Evaluation Criteria Provide General Framework For Feedback

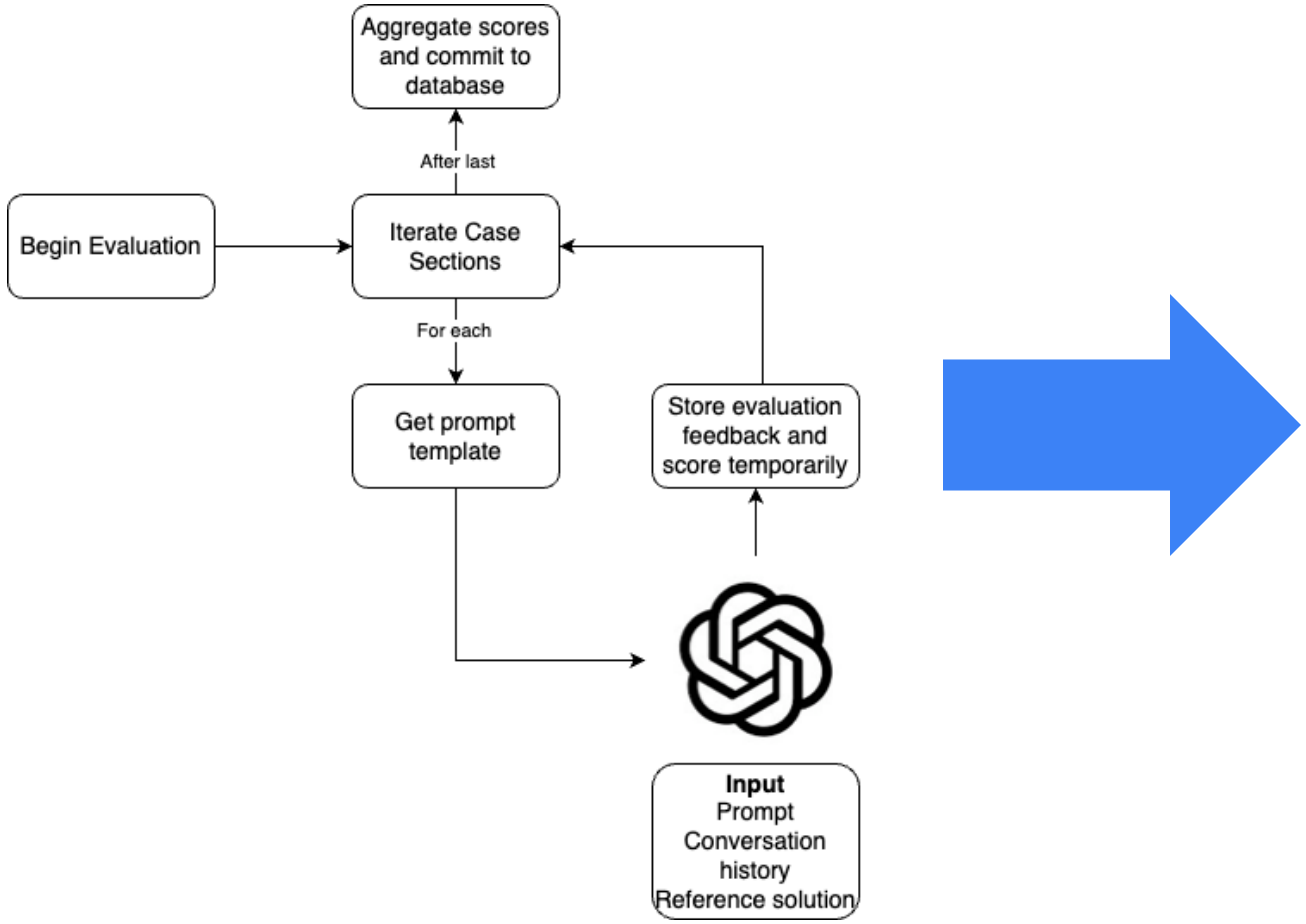
Build prompt templates for each section type for scoring

Prompt templates build on expert evaluation criteria

Compare candidate answers to reference solution

Focus on constructive feedback for the candidate. What should they do better next time

Case Evaluation



5 / 10

Your introduction is off to a good start. You asked a relevant question to clarify your understanding of the case. However, you missed an important step, which is summarizing the case back to the interviewer. This step is crucial to ensure that we are on the same page and that you have a clear understanding of the task at hand. In future interviews, make sure to include a brief summary of the case in your introduction.

Here's an example of how you could have summarized the case

"So, if I understand correctly, our client is a biotech company that has developed a groundbreaking treatment for Alzheimer's disease. This treatment is unique because it slows the progression of the disease rather than just treating its symptoms. However, there is a potential risk to the launch of this treatment due to a rumored shortage of

E

Business Model

Conducting one Interview Costs Around $1

Tokens Per Interview:

From Experience we know each Interview has:

- 2500 Output Tokens
- 30 000 Input Tokens

Cost Per Token (GPT-4.5 Turbo):

- Input: $0.01 / 1K tokens
- Output: $0.03 / 1K tokens

Cost Per Interview:

$0.01*2.5 + $0.03 * 30 = **$0.93**[1]

[1]Negligibe extra costs from hosting and additional services

Our Business Model – Subscription for Just $70 / Month Much Cheaper Than Competition

Our Service Available for Just **$70** a Month **Much Cheaper** Than The Competition

Key Assumptions:

- 20 Interviews per Month
- 2 month of preparation
- More than 1M people doing consulting interviews each year[1]
- 0,1% will use our product

Running the Numbers:

- Number of Users: 10,000
- Costs: $372,000
- Revenue: $1,4M

Profit Potential ≈ $1M

[1]Based on Assumption that McKinsey has 1M applicants per year

Execution

Two Factors For Effective Team Organization

Factor 1 - Weekly Meetings

- Discuss last week's results
- Decide next week’s priorities
- Split parts based on skills and knowledge

Factor 2 - Effective Tooling

- GitHub (Code)
- Shared Figma (Design)
- Draw.IO (Diagrams & Documentation)
- Kakao Talk

Project Timeline

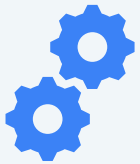
	10/16	10/26	11/02	11/09	11/16	11/20	11/23	11/30	12/6	12/8
Ideation & Scoping	✓									
Prompt Engineering	✓	✓								
Architecture Design		✓	✓							
Develop Fundamental Components			✓	✓	✓					
Develop UI (Frontend)				✓	✓	✓				
Connect All Components Together							✓			
Refine & Test										
Final Presentation										

Have a Look At Our Timeline in the Appendix

Three Key Learnings Throughout the Project

Ideation

- ✓ Good thing to **spend time on searching** topic and competitor solutions
- ✓ **Verifying** that product **idea** is sound and can provide value



Proof of Concept

- ✓ **Verify functionality in first draft** produced in Python with simple CLI application
- ✓ **Fail fast** and ensure that the solution is feasible

Architecture

- ✓ Choose architecture based on the **skills of the team members**.
- ❖ **Choice of coding language** is important to allow all members of the team to contribute to code base.
- ❖ **Learning while doing** requires more **effort**.

Future Work

Three Ideas for Future Improvements

1. Improve Chat Quality

Introduce **second LLM** to **evaluate responses** before showing to Candidate[1]

2. Add More Cases

More **cases can easily be added** to the library. Manual labor of **30 minutes required per case**

3. Improve Evaluation

Gather **more datapoints** for evaluation. Examples include **Sentiment, Tone, and Pose**

[1]Additional Information provided in the Appendix

THANK YOU

Appendix

Filter and Improve Chatbot Responses

The Challenge

- In the state-machine, chatbot required to **provide output in well-defined format**
- Example: *Did The Candidate Finish This Section → Answer: True/False*

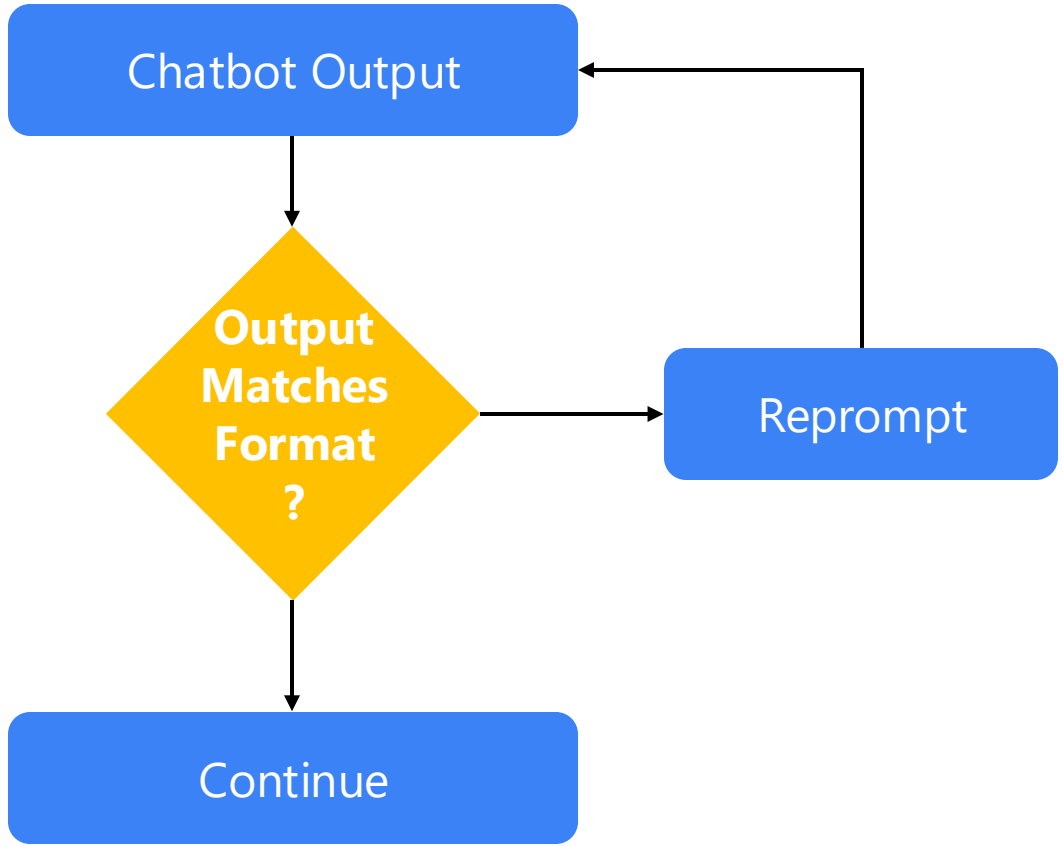
Our Solution:

- Custom Safety Filters to match patterns and reprompt the model

Future Work:

- Safety Filter only **checks semantics** but **not content**
- Output **quality** can be **improved** by having a **second model** evaluate the output

Safety Filter Architecture



Slide Added by Tim Lindenau after Deep Dive

DevOps – Easy Deployment, Automatic Scaling and Separation between Production and Development Environment

1. Vercel

- Deploy **Next.Js** application on **serverless** infrastructure
- Master branch automatically published
- Other branches published to **private preview builds**

2. Google Cloud

- **Cloud Run** to deploy **containerized speech-to-text** service
- Vision API
- Vertex AI language models

3. Planetscale

- Automatically **scaling** database
- **Production** and **development** environment **separated** by environment variables
- **Git-like branching** feature to test-features more easily

4. Clerk

- Service to guarantee stable and secure authentication
- Production and development **environment separated**
- Secures **application** and **speech-2-text**

Slide Added by Tim Lindenau after Deep Dive

Project timeline

	10/16	10/26	11/02	11/09	11/16	11/20	11/23	11/30	12/6	12/8
Ideation & Scoping	✓									
Prompt Engineering	✓	✓								
Architecture Design		✓	✓							
Develop Fundamental Components			✓	✓	✓					
Develop UI (Frontend)				✓	✓	✓				
Connect All Components Together							✓			
Refine & Test								✓	✓	
Final Presentation										✓