

# Vocal AI

## Product spec.

A product spec is essentially the definitive guide to what the product should do from a user and business perspective.

### **Executive Summary/Overview**

VocalAI is an AI-powered assessment tool that evaluates students' thinking skills through oral responses to culturally neutral, real-world scenarios. The tool will allow for real-time analyses of students' responses to the exam questions. VocalAI will not only analyse the student's final answer, but it will also capture students' spoken reasoning, analyse their argumentation quality, and provide unbiased scoring recommendations to human examiners. The tool will also help examiners generate questions based on culturally neutral, real-world scenarios, reducing bias and enabling inclusivity.

The assessment and questions are modelled after the A-Level curriculum. By combining automated speech analysis, argument structure detection, and body language and emotions analysis from video/voice, VocalAI reduces cultural and linguistic bias, increases assessment reliability, and supports more scalable and equitable evaluation.

### **Problem Statement**

Traditional assessments rely heavily on written responses, which can overlook students' reasoning, creativity, and communication skills—core elements of critical thinking. This assessment process may also introduce bias related to language proficiency, writing style, and cultural familiarity. Human examiners may unintentionally favour fluent speakers or culturally familiar reasoning patterns.

In the A-level curriculum, educators face the challenge of evaluating students' conceptual understanding and analytical depth through limited assessment formats.

Emerging speech recognition and AI-driven evaluation technologies open new possibilities for oral assessments that capture real-time thinking, argumentation, and reflection. The challenge lies in designing a fair, accurate, and pedagogically sound oral assessment system that enhances rather than replaces human judgment

### **Goals & KPI**

#### *Primary Goals*

- Generate culturally neutral questions modelled on the thinking skills curriculum.
- Evaluate students' oral reasoning using defined Critical Thinking criteria.
- Evaluate students' thought processes and problem-solving skills using defined criteria.

- Provide consistent, unbiased scoring recommendations.
- Highlight students' reasoning structure for human examiners.
- Reduce dependency on English fluency.

#### *KPI*

- 90% accuracy rate in alignment with expert scores
- 70% reduction in time spent generating culturally neutral questions.
- 90% accuracy in identifying argument components (premises, conclusions).
- Improved student experience (measured via satisfaction surveys).
- 20% reduction in time spent for marking students

#### **Target Users**

- Students: Diverse language backgrounds, aged 17–19, preparing for their A-level exams.
- Examiners and Educators: Responsible for teaching and grading thinking skills exams.
- Institutions and Exam Boards: Seeking scalable, fair assessment solutions.
- ESL students benefiting from non-linguistic evaluation.

#### **User Stories/Use Cases**

- As a student, I want to answer questions orally so I can demonstrate my reasoning without being limited by writing ability, and final answers.
- As an examiner, I want the AI to highlight argument structure so I can quickly verify whether the student applied CT principles and problem-solving skills.
- As an institution/exam board, I want to create culturally neutral questions so assessments remain fair and globally applicable.
- As an ESL student I want to be able to illustrate my answers in my native language.

#### **Features & Requirements**

##### **A. Question Generation**

1. Generate culturally neutral, real-world scenario questions. Questions will be designed in line with the Cambridge International A-level Thinking Skills curriculum.
  - The first section will focus on students' problem-solving skills, inspired by paper 1 and paper 3. (questions focuses on logic and might need some primary mathematical skills). To answer the question, students need to think out loud and explain their thinking process to find the right answer.
  - The second section will focus on Critical thinking, inspired by papers 2 and 4 (questions present document/article with real-world scenarios and students need to analyse these documents and answer the questions that follow.
2. Provide expected reasoning paths and marking guidance internally.

*Future features:*

- Difficulty tiers.
- Ability to auto-calibrate questions based on student performance.

## **B. Oral Response Capture & Analysis**

- Real-time speech-to-text transcription (multilingual) to allow for response analysis and translation if needed.
- Video analysis for body language and emotions.
- Analysis of any written notes/ text highlighted or underlined, with time stamps to analyse their process of thinking.
- For section 1(PS), Problem analysis and solution extraction:
  - The issues/challenge
  - Solution
  - Method/steps taken to get the answer. (in this, the VocalAI will produce a road map explaining the student's answer)
- For section 1(PS) Detect, map and assess students' thought processes:
  - Problem framing: Clarity in identifying information sources and the logical Relationship between them.
  - Strategy selection: The logic behind the steps taken to solve the problem
  - Option evaluation: The accuracy of the presented solutions (is the answer right or wrong)
  - Trade-off reasoning: clarity in comparing options.
  - Decision-making under uncertainty.
- For section 2 (CT), Detect and assess:
  - Relevance of information.
  - Evaluation of evidence.
  - Strength of reasoning.
  - Coherence and clarity of constructed arguments.
  - Formation of well-reasoned decisions.
  - Use of an analogy and hypothetical reasoning

## **C. AI-Generated Assessment & Recommendations**

VocalAI will give a descriptive assessment of the student's ability to achieve the following Assessment objectives<sup>1</sup>.

AO1: Use and evaluate information. Weighing 40%

AO2: Create and communicate reasoning. weighing 60%

It will also will recommend an overall score to each section according to the following description

---

<sup>1</sup> Based on Thinking skills syllabus for 2028-2030

Grade/ student score	Use & Evaluate Information (AO1)	Create & Communicate Reasoning (AO2)
A* 71- 100 <sup>2</sup>	Exceptional ability to filter, interpret, and critically assess complex data. Accurately distinguishes relevance and recognises subtle implications and hidden assumptions.	Highly sophisticated, coherent reasoning. Innovative, fully justified arguments/solutions. Clear, logical, refined communication anticipating counterpoints.
A 62-70	Very strong ability to select and interpret relevant information. Recognises most assumptions and evaluates information soundly.	Well-structured reasoning with strong justification. Clear communication with logical flow and some self-checking.
B 52-61	Good use and interpretation of relevant information. Identifies common assumptions but may miss subtle ones. Generally accurate evaluation.	Logical reasoning though sometimes standard or underdeveloped. Mostly clear communication with occasional lapses.
C 43-51	Competent with straightforward information. Recognises obvious assumptions. Evaluation may be inconsistent or superficial.	Reasoning is coherent but simplistic or underdeveloped. Arguments/solutions may lack depth or full justification.
D 36-42	Struggles with complex information. Often misses assumptions or relationships. Evaluation shallow or flawed.	Reasoning fragmented or unclear. Arguments/solutions weak, unstructured, or unjustified. Communication lacks coherence.
E 26-35	Very limited ability to use or evaluate information. Fails to identify relevance or relationships. Evaluation very weak or incorrect.	Minimal or illogical reasoning. Arguments/solutions unclear or unsupported. Communication is disorganised with no self-reflection.

#### *Additional AI insight outputs*

- Confidence analysis: hesitation patterns, tone shifts, body language scoring.
- English proficiency estimation (to reduce bias).

#### **D. Examiner/ students Interface**

- Visual highlighting of student problem analysis, roadmap and argument structure.
- Side-by-side comparison of:
  - Student transcript
  - AI reasoning analysis
  - Audio/video playback
- Adjustable scoring rubric aligned to official exam mark schemes(only for examiner).
- Override and justification logging for human examiners (only for examiner).

<sup>2</sup> Numeric grades are based on Cambridge grade thresholds for Thinking Skills (9694) – June 2025, <https://www.cambridgeinternational.org/Images/740354-thinking-skills-9694-june-2025-grade-threshold-table.pdf>

## E. Reporting & Feedback

- Student feedback on thinking skills, strengths and weaknesses, and areas to improve.
- Examiner analytics and scoring consistency dashboards.
- Exportable reports for institutions.

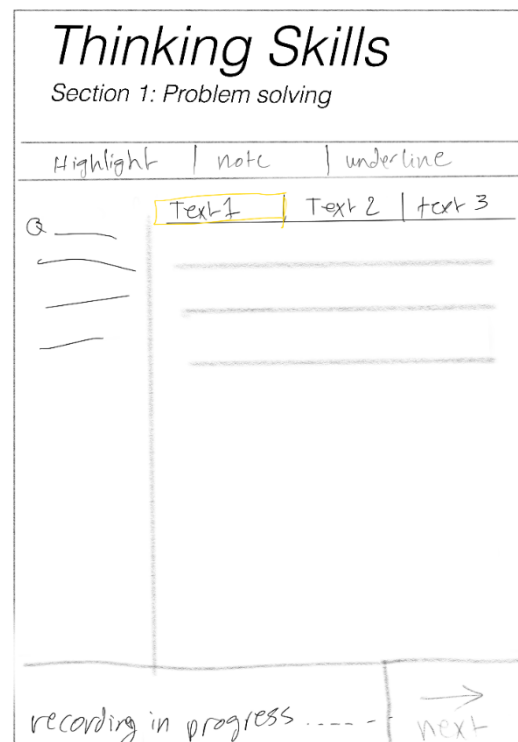
### User Experience/Design Guidelines

Students → present scenario → think aloud → AI analyses reasoning → AI highlights structure → examiner reviews → final human-approved grade.

To start the exam, students will need to log in with their ID and password provided by the examiners.

After that, students will be able to read the instructions and sign a consent form indicating that the exam is being recorded, and AI will be used in the assessment.

Before the exam begins, students will be prompted to enter their demographic information “How do you describe your cultural background, i.e. in terms of your ethnicity, upbringing, religion, worldview?” and choose their preferred exam language from a list: “How do you describe your cultural background, i.e. in terms culturally appropriate and contextually accurate content.



They will start with the first section, Problem solving; video and voice record will start as soon as they start the exam. Each thinking process will be recorded and analysed. Students will be able to highlight and underline the text, write notes to identify an element.

Questions will be displayed next to the text, to allow for active analysis and answers.

After submitting the exam, the Student will get an estimated grade by VocalAI. VocalAI will send the examiner feedback and a score recommendation. The examiner will be able to review this score, validate it and send the mark back to the student within a few days.

During the exam, if a student shows signs of stress or lack of confidence, the AI will generate motivational quotes to help them regain their confidence. These quotes will appear randomly and will be tailored to the student’s culture and language. Based on the emotional

analysis, the student may also be advised to take a few minutes between questions to breathe. This ensures that students can maintain their performance.

Examples of quotes include: “Relax! You’ve got this.” “Keep going!” “Focus, believe, achieve.” “One step at a time.” “Stay calm, stay confident.” “Small progress is still progress.” If the pop-up text does not help calm the student, the AI will proceed to use an empathetic voice.

Motivational quotes are important to help students gain confidence and raise their self-esteem. It is proved that self-esteem could predict critical thinking dispositions<sup>3</sup>.

### **Ethical concerns.**

- Students taking the exam will sign a consent form.
- Oral assessment should not be the sole tool for assessing students' thinking skills; a hybrid mode is recommended to allow students with poor public speaking skills/shyness to get good grades.
- Human examiners will always approve final grades to avoid any misunderstanding by the AI. However, to avoid any bias, the student will get an estimation by the AI right after they submit their exam, and Human mark should not be more than 5% different.
- Students must select their exam language from a predefined list of supported languages. The AI's performance may vary across languages, but additional languages will be added as the tool is further developed.
- data will be anonymised after grading.
- All recordings will be deleted from the server after two years of the exam.

### **How cultural background and upbringing can affect person's Thinking skills ?<sup>4</sup>**

1. **Language and test performance.** Measured differences in critical-thinking scores between Asian and Western students were substantially explained by English proficiency: lower performance on English-language critical-thinking tests can reflect language limits, not weaker reasoning per se. ScienceDirect
2. **Different preferred reasoning styles.** People from East Asian backgrounds more often use dialectical or contextual reasoning (emphasising contradiction, multiple perspectives, and synthesis), whereas many Western educational contexts emphasise analytic reasoning (formal logic, argumentation). These styles affect how people approach problems, what counts as a convincing answer, and how they perform on standard tests. ScienceDirect+1
3. **Nonverbal / social-cognitive cues shape thinking in interaction.** Cultural norms about eye contact, gestures, proxemics and other body language change how people signal attention, confidence, and intent — which in turn shapes social inference, persuasion, and

---

<sup>3</sup> <https://pmc.ncbi.nlm.nih.gov/articles/PMC10312417/>

<sup>4</sup> <https://pdfs.semanticscholar.org/624a/9e2728b886d8ae6d718ae126a63c2ff9c1.pdf/>  
<https://www.sciencedirect.com/science/article/pii/S1041608010000816>

classroom/workplace reasoning (who speaks up, who is taken seriously, how arguments are interpreted).

4. **Educational practices and values shape cognitive habits.** Different schooling methods (rote memorization vs. debating, teacher-centered vs. inquiry-based) create long-term habits: attention to context vs. attention to abstract rules, tolerance for ambiguity, and comfort with disputation. These habits affect both what skills are strong and how people display them in assessments.
5. **Measurement & interpretation issues.** Many standard critical-thinking instruments assume a particular (often Western, analytic) notion of “good thinking.” That can bias comparisons unless language, test format, and cultural reasoning norms are accounted for.