The Importance of Q&A in Academic Presentations

Academic Presentations

- Sharing research findings
- Engaging in Q&A discussions

The Challenge of Preparing for and Reflecting on Q&A

- Requires deep understanding of the paper
- Limited opportunities for structured practice and reflection

→ AI-supported practice and feedback for conference Q&A

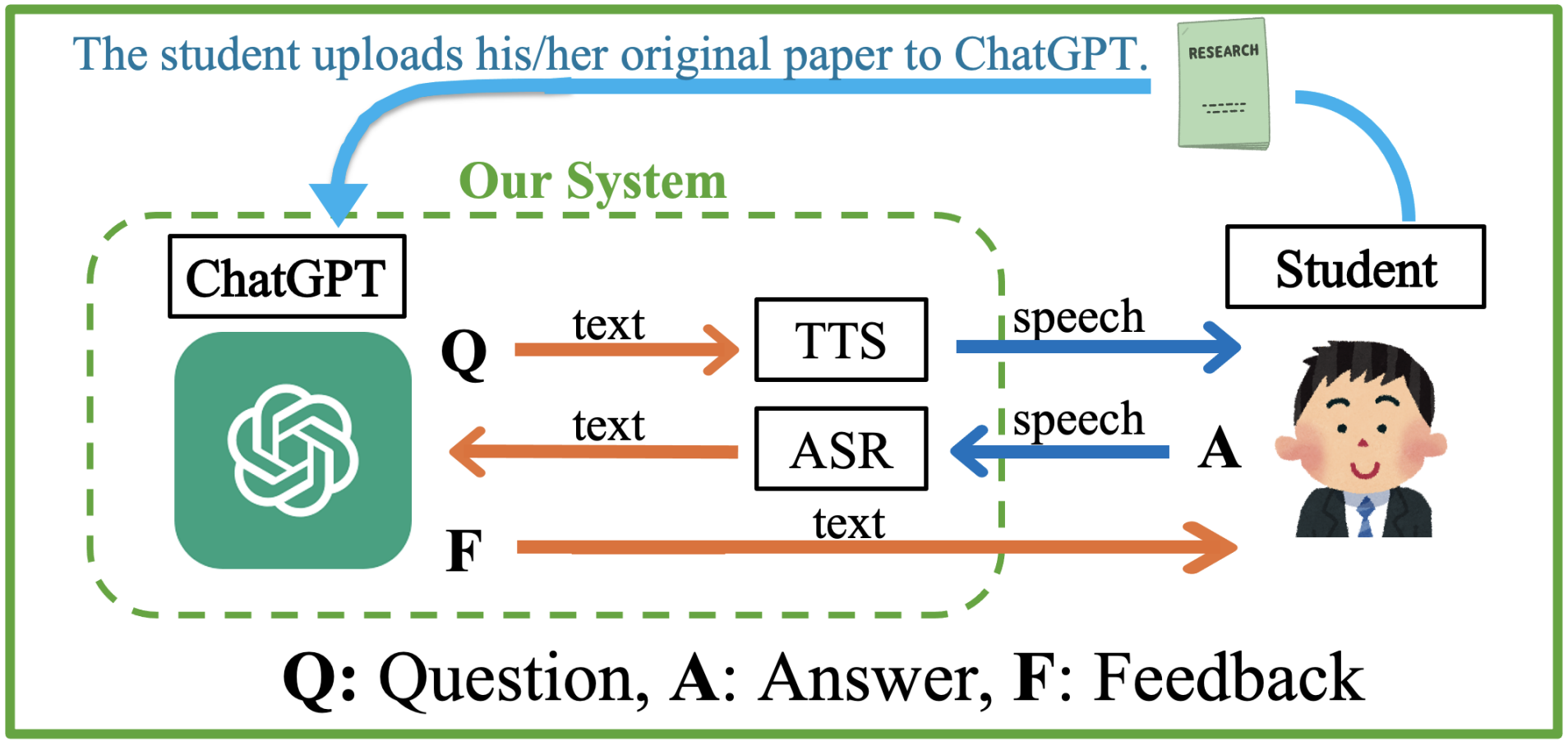
Case 3:

A ChatGPT-Based Oral Q&A Practice System for L2 Speakers at International Conferences

Mayuko Aiba, Daisuke Saito, and Nobuaki Minematsu. GPT-based simulation of oral Q&A to support students attending first conference. JALTCALL Trends, 1(1), Article 2163, 2025.

System Architecture

Large Language Models (LLMs) with ASR + TTS can help students.



Design and Evaluation of the Q&A Generation System

System Development

- We developed a customized Q&A chatbot using the GPT customization feature in ChatGPT Plus.
- The system was built on the **GPT-4 Turbo** and designed specifically for academic conference papers in engineering.

Experimental Setting

Participants	8 Engineering Master's Students
Paper Type	International Conference Papers
Presentation Experience	No more than twice
Question Generation	6 different configurations
Evaluation Method	Author-based relevance assessment

GPT customization feature

GPTs feature allows users to create a dedicated chatbot with predefined behavior and objectives.

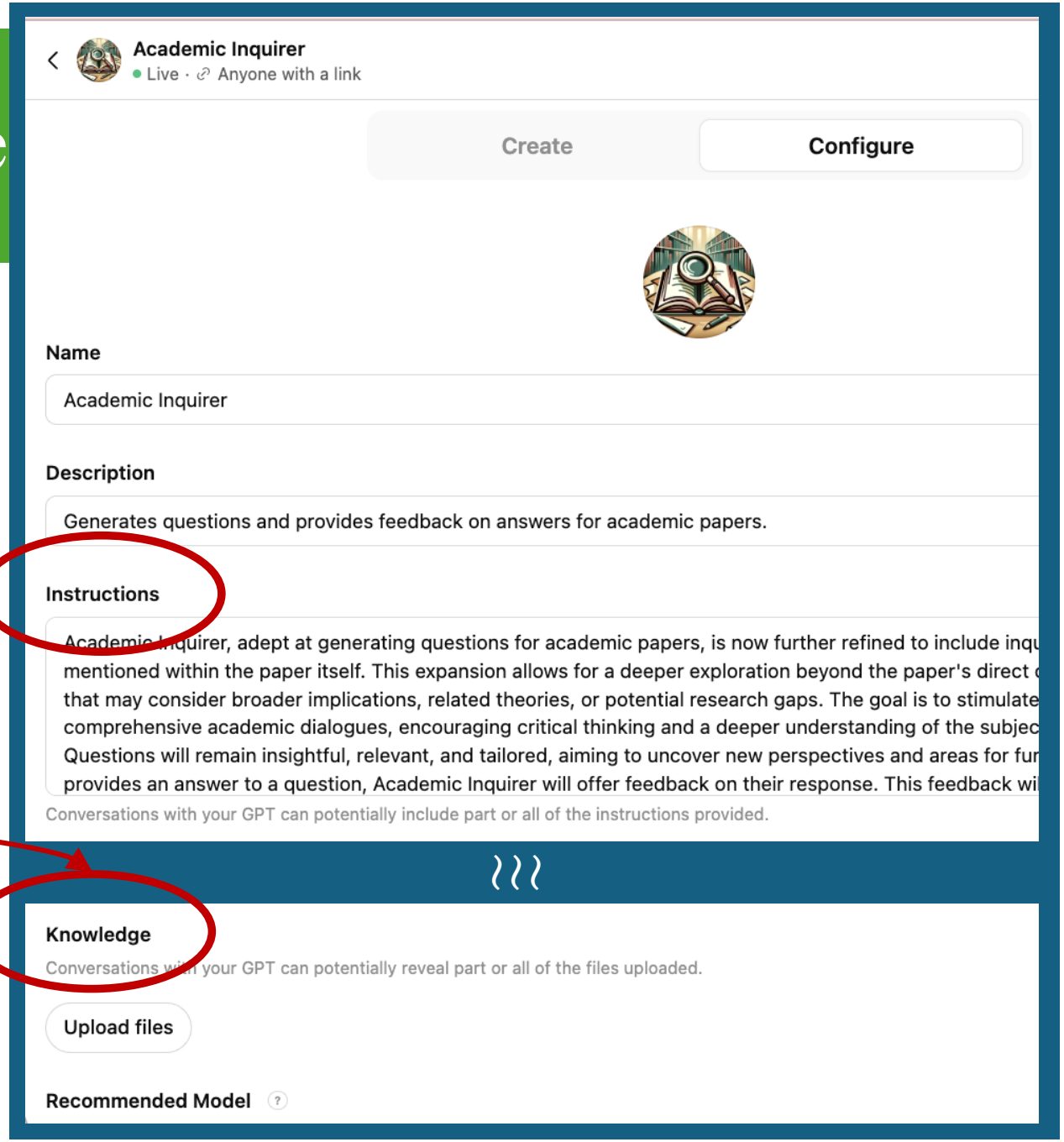
Key Components

1. Instructions

   Define the chatbot's role and response style

2. Knowledge

   Provide background materials or reference information to guide output generation



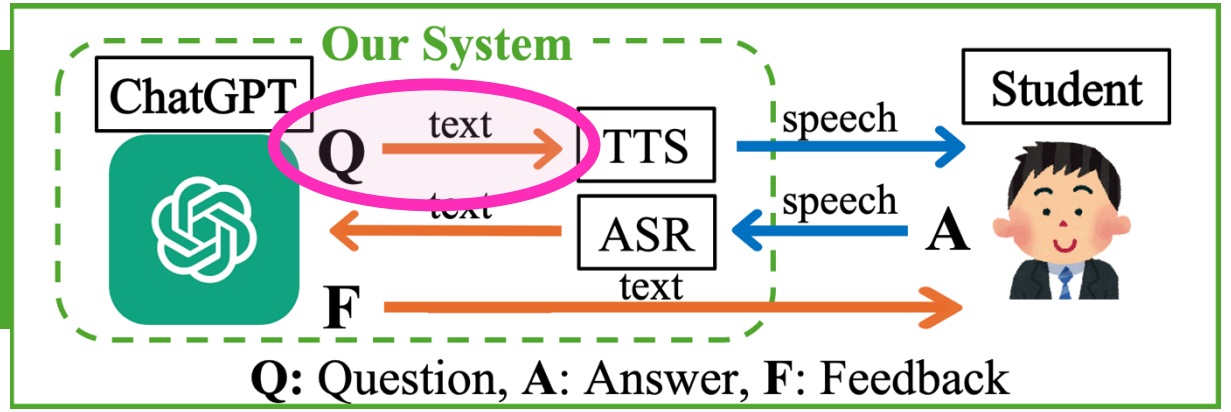
Custom GPT for Q&A Generation

Summary of *Instructions*:

- Purpose is to generate questions for academic papers
- Generating questions specifically for engineering papers
- No restrictions on the difficulty, field, or number of questions generated at once

→ Ensures consistent and structured question generation

Six Configurations



Six configurations, labeled as A-F, are examined.

Scope of question generation	references uploaded	references not uploaded
Each section	A	D
Each keyword	B	E
Entire scope	C	F

Evaluations on Generated Questions by Students

Evaluation by the students who authored the papers (8 participants).

- They rated each question on a scale of 1 to 5 across five criteria: Relevance, Clarity, Specificity, Inspiration, and Expectation.
- Additionally, they assessed whether the answer to each question is contained within the paper.

Each participant evaluated approximately

60 questions × (5 criteria + 1 additional check) = 360 items

in total.

Preferences Over Six Configurations

Students rated their preference over configurations.
Each participant allocated a total of 100 points to A to F.

Scope of question generation	references uploaded	references not uploaded
Each section	A	D
Each keyword	B	E
Entire scope	C	F

When a question was presented, its configuration, A to F, was also shown, but they did not know the meaning of the configuration.

Evaluations on Generated Questions

Scope of question generation	references uploaded	references not uploaded
Each section	A	D
Each keyword	B	E
Entire scope	C	F

The students most preferred **F**, which neither specifies keywords nor sections and does not explicitly incorporate reference papers.

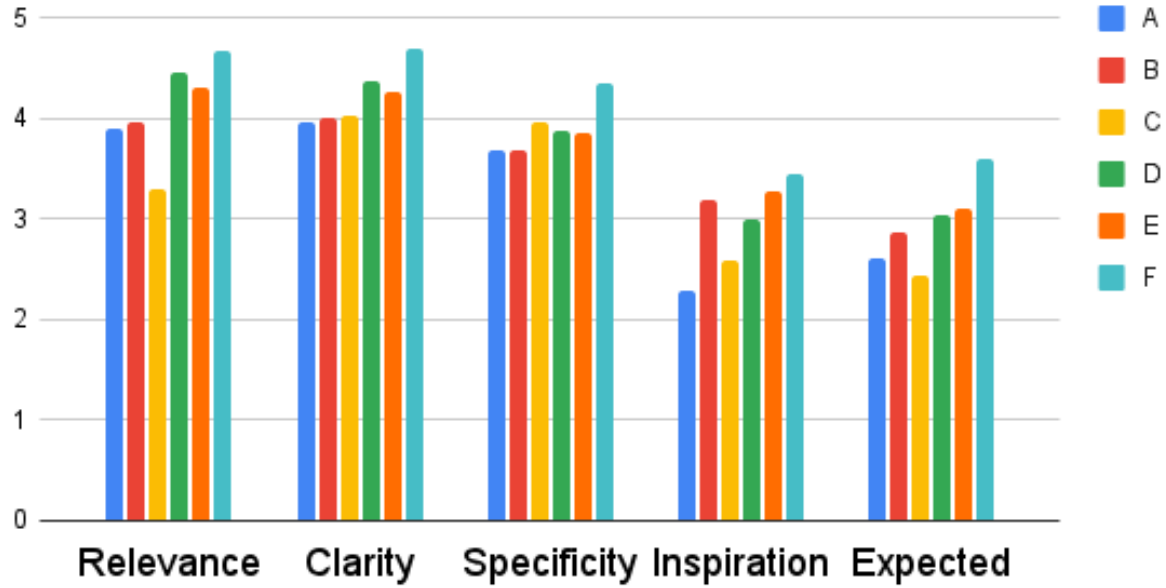
Overall, modes that references were not uploaded received higher evaluations.

The averaged preference over the six configurations

	A	B	C	D	E	F
Students	10.0	15.0	9.1	18.4	21.0	**26.5**

references uploaded references not uploaded

$$\Sigma = 100$$



Students' Evaluation of Each Criterion for Each Mode

Insightful Questions

Questions whose answers are NOT found in the paper,
and *Relevances* are 3 or more can be considered insightful.

	A	B	C	D	E	F
Relevance≥3	35.2	35.4	31.7	28.0	39.6	**40.0**

The proportion of insightful questions was highest in Mode F.

All students expressed a desire to use it, particularly in Mode F.

A Possible Explanation

Why were configurations without references preferred?

Conference Q&A

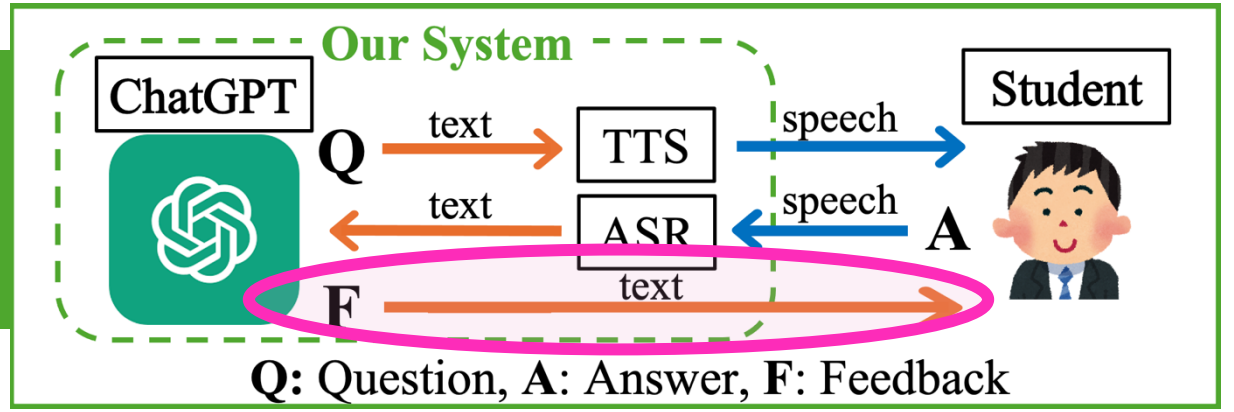
- Questions are based on short presentations (10–15 minutes)
- The audience does not read the full paper

Journal Peer Review [Liang+, 2023]

- The entire paper is carefully examined

The nature of questions differs between international conferences and journal peer review.
In conference settings, less detailed questions were rated more highly.

Automated Feedback After Q&A Practice



Language-Level Feedback

- Recommended corrections on English expressions
- Revised sentence suggestions

Speech Analysis

- ASR transcription
- Confidence scores
- Word count and lexical diversity
- Fluency and comprehensibility scores

Limitation: Focuses mainly on linguistic performance

→ Does not directly evaluate response content or academic validity

Case 4:

A Study on Feedback Generation Methods for Post-Presentation Q&A Based on Structuring Dialogue Understanding Using LLM

Mayuko Aiba, Daisuke Saito, and Nobuaki Minematsu. 2026. Incorporating Respect into LLM-Based Academic Feedback: A BI-R Framework for Instructing Students after Q&A Sessions. In Proceedings of the 16th International Workshop on Spoken Dialogue System Technology, pages 288–301, Trento, Italy. Association for Computational Linguistics.

From Practice to Content-Focused Feedback in Q&A

- Why Q&A matters:
  Discussion brings new perspectives and clarifies understanding.
  → To benefit fully from Q&A, presenters must accurately interpret questions and respond appropriately.

- Observed challenge:
  Students may misunderstand the question or its purpose

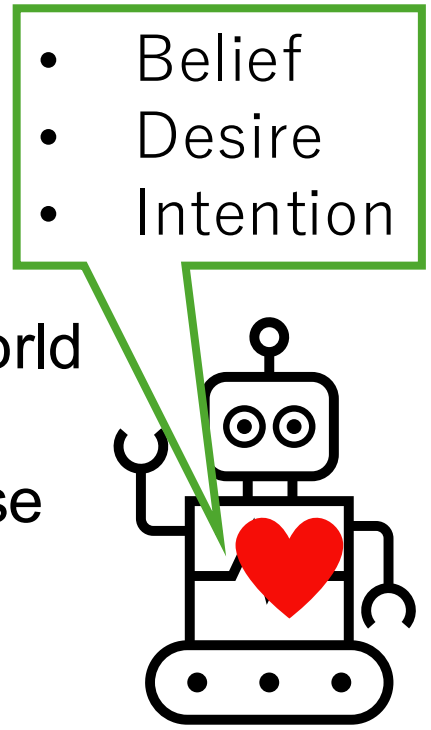
- Our objective:
  Automate a supervisor-like feedback using an LLM
  → **Structure the LLM's reasoning via prompting**

BDI Model

BDI Model(Bratman, 1987)：Belief–Desire–Intention

Provides a framework for modeling decision-making in rational agents

Widely adopted in AI and multi-agent systems



Core Components:

- **Belief:** The agent's understanding or assumptions about the world
- **Desire:** The goals the agent aims to achieve
- **Intention:** The committed plan of action chosen to achieve those goals

Modeling Conference Q&A as BDI-Structured Dialogue

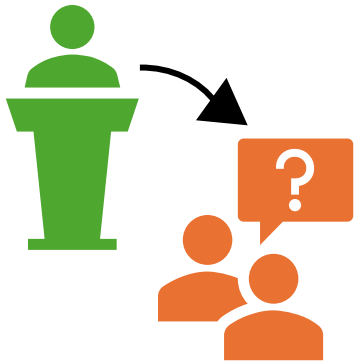
We interpret the process through the BDI framework.

Three phases of interaction:

1. Presenter gives a talk
2. Questioner asks a question
3. Presenter answers to the question.

Presenter talk

Questioner asks



Presenter answers

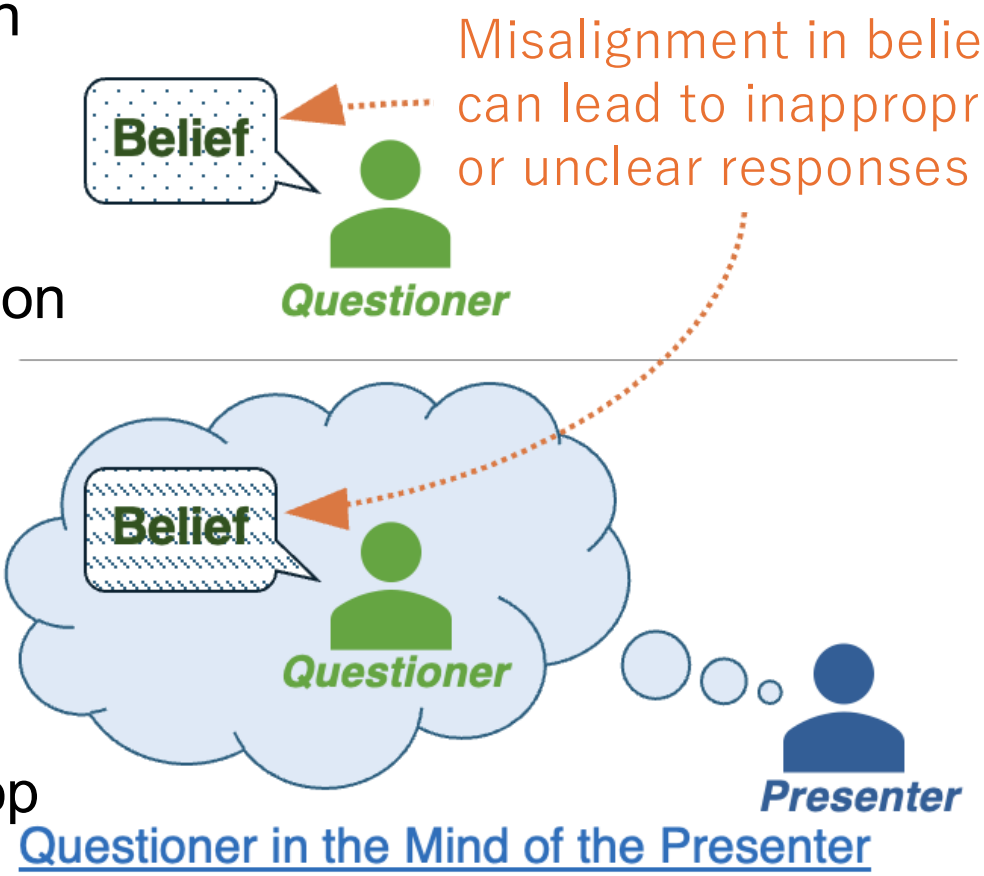
Interpreting Q&A Through the BDI Model

Questioner

- Belief: An interpretation of the presentation based on personal knowledge and values
- Desire: To confirm understanding or contribute meaningfully to the discussion
- Intention: To ask questions that deepen the discussion or introduce new perspectives

Presenter

- Belief: An interpretation of the questioner's thinking within the context of one's own research
- Desire: To respond clearly and accurately to the question
- Intention: To answer the question and further develop the discussion based on the research content



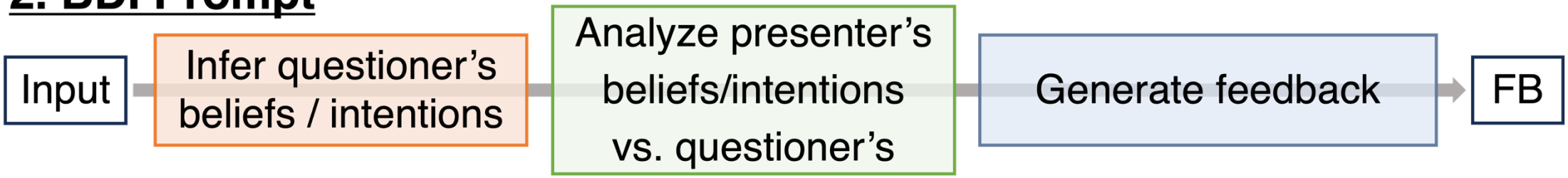
Questioner in the Mind of the Presenter

The Flow of Proposed Prompts

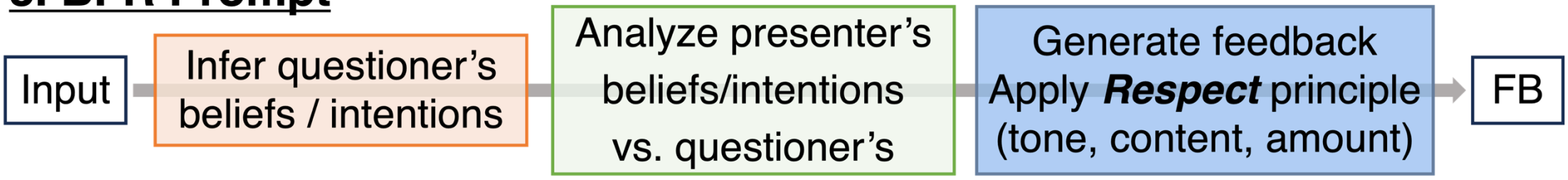
Input: paper PDF, presentation slides, transcript of Q&A session

1. Baseline Prompt

2. BDI Prompt



3. BI-R Prompt



While BDI models logical inference, it does not address pedagogical considerations.
→ We introduce Respect as a guiding principle for student feedback.

Experimental Overview

We conducted two user studies to evaluate the effectiveness of feedback generated by different methods.

- **Master's student study** (N=6)**:**
  Pairwise comparison between the proposed methods and a baseline (Baseline vs. BDI/BI-R)
- **PhD student study** (N=3)**:**
  Simultaneous comparison of three types of feedback (Baseline vs. BDI vs. BI-R)

Feedback was generated by **ChatGPT (GPT-4o)** for each Q&A interaction.

Although this approach is expected to be applicable to L2 contexts, the present study was conducted in Japanese with native Japanese speakers.

Evaluation Criteria

Students who made presentations were evaluated subjectively based on the following two criteria:

- **Usefulness**:
  Perceived effectiveness of the feedback for improving research and presentation quality (5-point scale)
- **Preference**:
  Selection of the most natural and acceptable feedback among alternatives

Master's Student Study (Pairwise Comparison)

Usefulness and Preference of Feedback by Prompt Type (master's students)

Comparison	Usefulness (Base)	Usefulness (BDI/BI-R)	Statistical Test	Preference Rate
Baseline vs. BDI (N = 5 participants)	3.22	3.24	t(26)=-0.17, p> .05	0.568
Baseline vs. BI-R (N = 6 participants)	3.39	**3.94**	t(39)=2.42, **p< .05**	**0.748**

Comments from the students

- Baseline vs. BDI: "Too detailed" / "Overly critical" (3 comments)
  → Overly detailed feedback may increase cognitive load and hinder learning (Shute, 2008).
- Baseline vs. BI-R: "Easy to accept" (1 comment),
  "Considerate of the presenter's perspective" (1 comment)

Subjective Evaluation of Feedback Quality

Preferences varied across participants.

PhD-level Q&A sessions contained more clarification questions than master's students.
Consequently, the concise baseline feedback may have been relatively more preferred.

Usefulness and Preference of Feedback by Prompt Type (PhD students)

Participant	Baseline	BDI	BI-R
P7(n=12)	4.55, 0.58	4.25, 0.33	4.25, 0.08
P8(n=7)	4.14, 0.14	4.29, 0.43	4.43, 0.43
P9(n=27)	4.48, 0.41	3.81, 0.07	4.52, 0.52
Overall	4.39, **0.38**	4.12, 0.28	**4.40**, 0.34

Reanalysis Excluding Clarification Questions

When excluding clarification questions:

Participant	Baseline	BDI	BI-R
P7(n=9)	4.56, 0.44	4.44, 0.44	4.44, 0.11
P8(n=6)	4.17, 0.17	4.17, 0.33	4.33, 0.50
P9(n=17)	4.41, 0.35	3.94, 0.06	4.53, 0.59
Overall	4.38, 0.32	4.18, 0.28	**4.44, 0.40**

Comparison	Statistical Significance
Baseline vs. BDI	Not significant (*)
Baseline vs. BI-R	Not significant (*)
BDI vs. BI-R	Significant (**)

- PhD students tend to infer the intent and core issues of questions independently, potentially reducing the perceived value of BDI- and BI-R-based feedback.
- Feedback incorporating Respect was more considerate and better aligned with the presenter's perspective.

Summary and Future Work

Summary

- While the BDI prompt alone did not consistently outperform the baseline, the BI-R prompt demonstrated clear improvements in usefulness and preference depending on question type and student expertise.
- Incorporating Respect shapes feedback reception beyond politeness, directly influencing feedback preference.

Future Directions

- Validating in larger and more diverse settings, including non-native English presentations and other research domains.
- Analyzing the causes of ambiguous or inappropriate responses using the proposed framework.
- Incorporating multimodal cues for more adaptive feedback.