



Teachers College
COLUMBIA UNIVERSITY

LanPIP

**A Specialized LLM-empowered Pipeline for
Customizable Language Learning**

Xinmeng Hou (Faceia)

Agenda

01 Motivation

02 Overview

03 Details

04 User Studies

05 Example Video

01

¡Hola! soy Faceia

- B.A. from UCLA
- M.A. from TC, Columbia
 - **ESL teaching at Community Language Program**
- Visiting Scholar @ SJTU



01

Motivations

Dealing with **diverse**
learning needs

01

02

Help students catch up the
lesson they were absent

Reliable outside classroom
learning **resources**

03

04

Assess student language
performance

01

Why not ChatGPT ?

- 01** Students know nothing about LLM nor promoting LLM
- 02** LLM chatbot is not specifically designed for education.
- 03** Can't ensure students learning quality

01

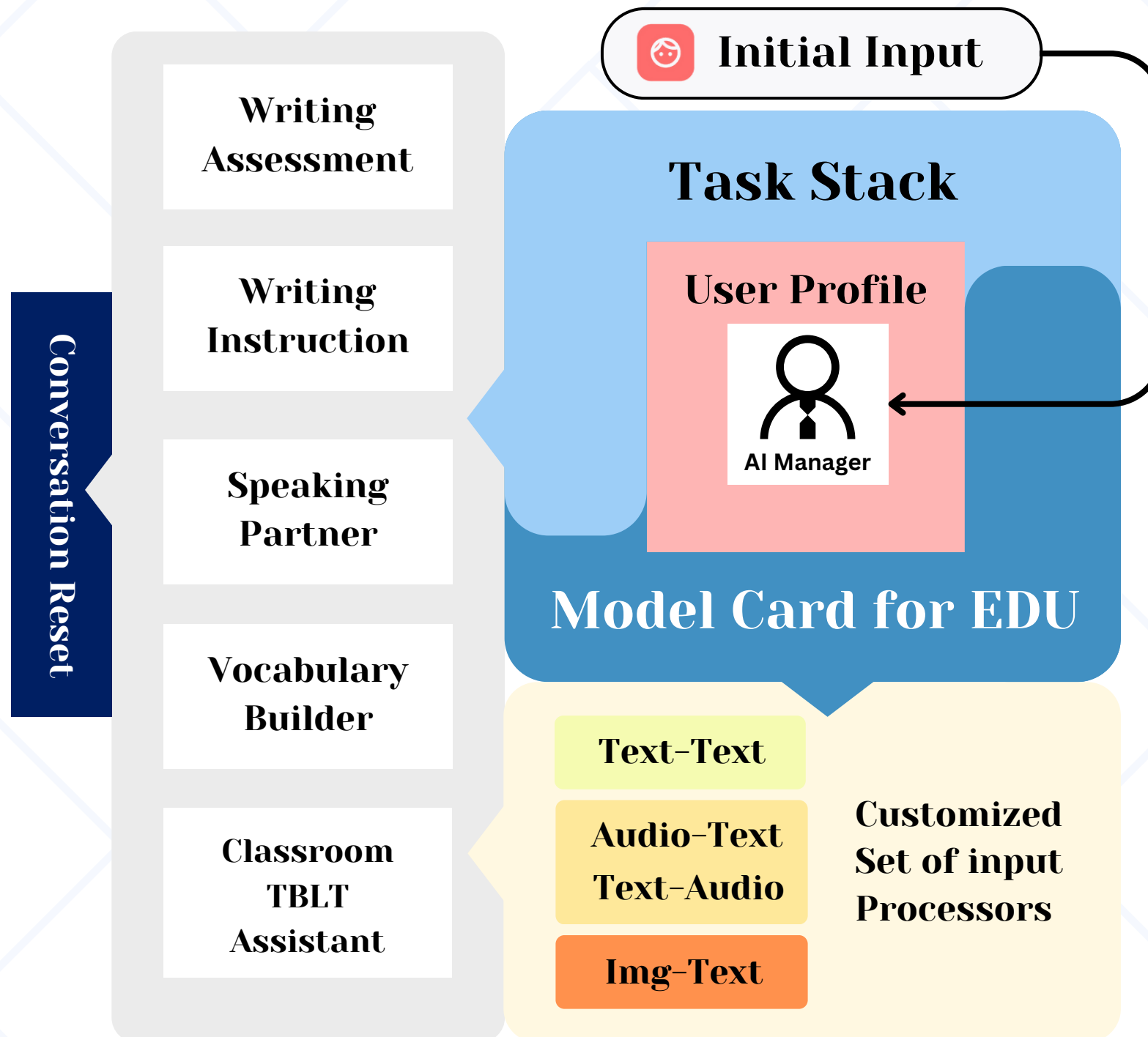
Why not ChatGPT ?

- 01** Students know nothing about LLM nor promoting LLM
- 02** LLM chatbot is not specifically designed for education.
- 03** Can't ensure students learning quality



**Multi-LLM
Agent**

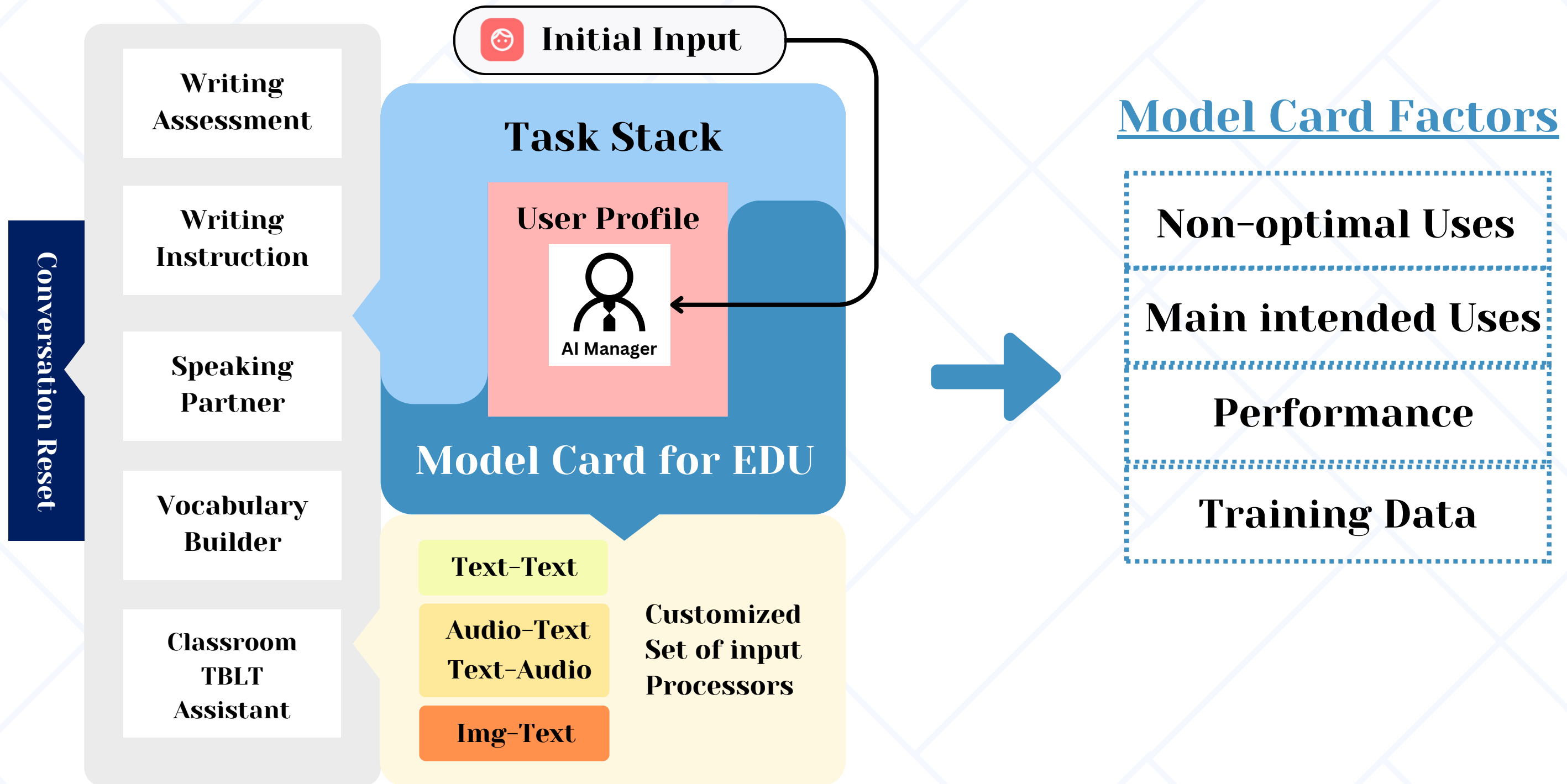
Pipeline Overview



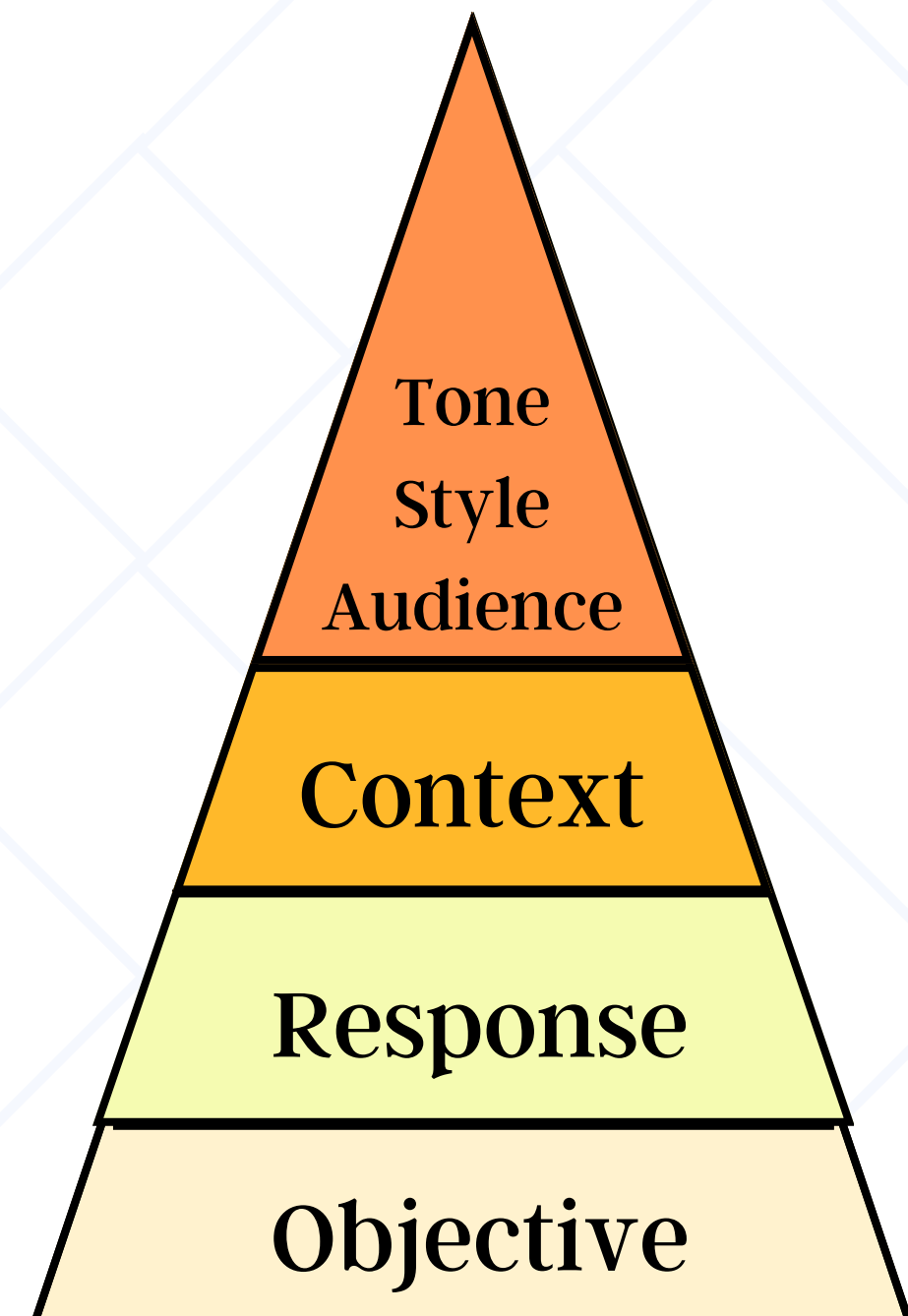
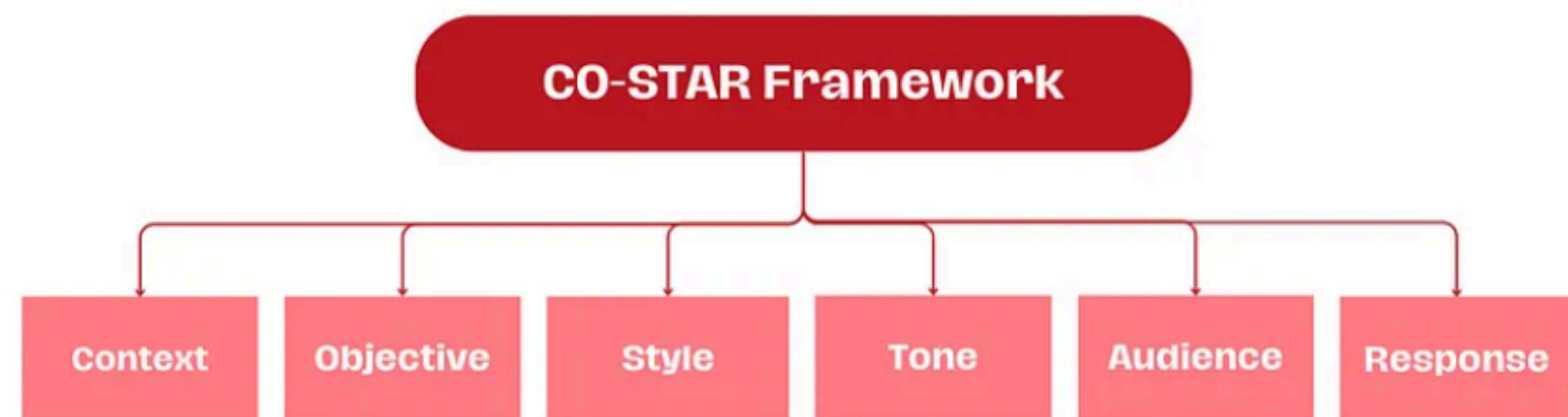
- APP implemented with **Streamlit**
- **Models:**
 - **T2T**
 - GPT-4, GPT-3.5-turbo, Claude 3 Opus, and Llama2-70b-chat
 - **A2T**
 - Whisper-1
 - **T2A**
 - tts-1
 - **I2T**
 - GPT-4 and Claude 3 Opus

02

Pipeline Overview



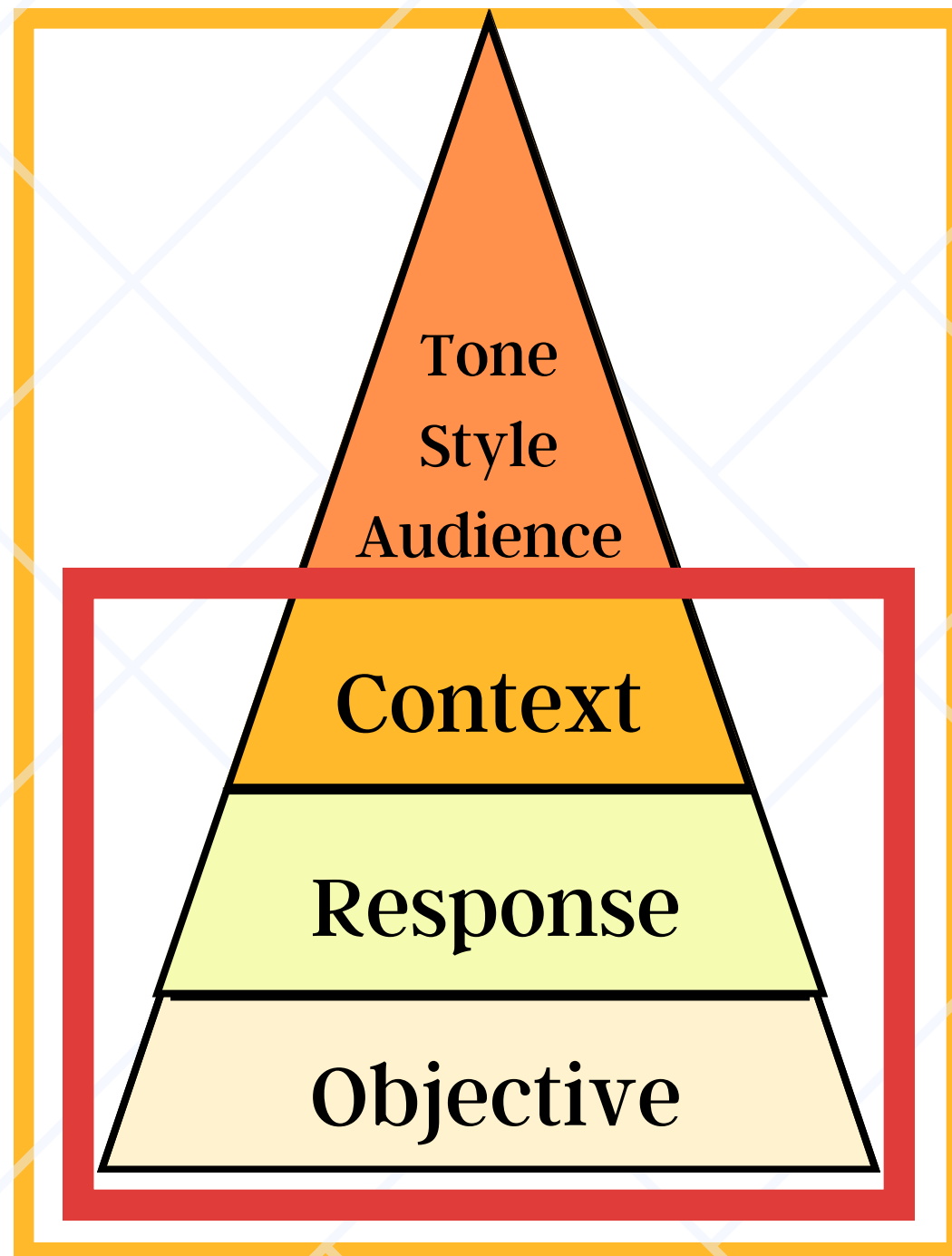
Prompt Design



Teo, S. (2024, April 21). How I won Singapore's GPT-4 Prompt Engineering Competition. Medium. <https://towardsdatascience.com/how-i-won-singapores-gpt-4-prompt-engineering-competition-34c195a93d41>

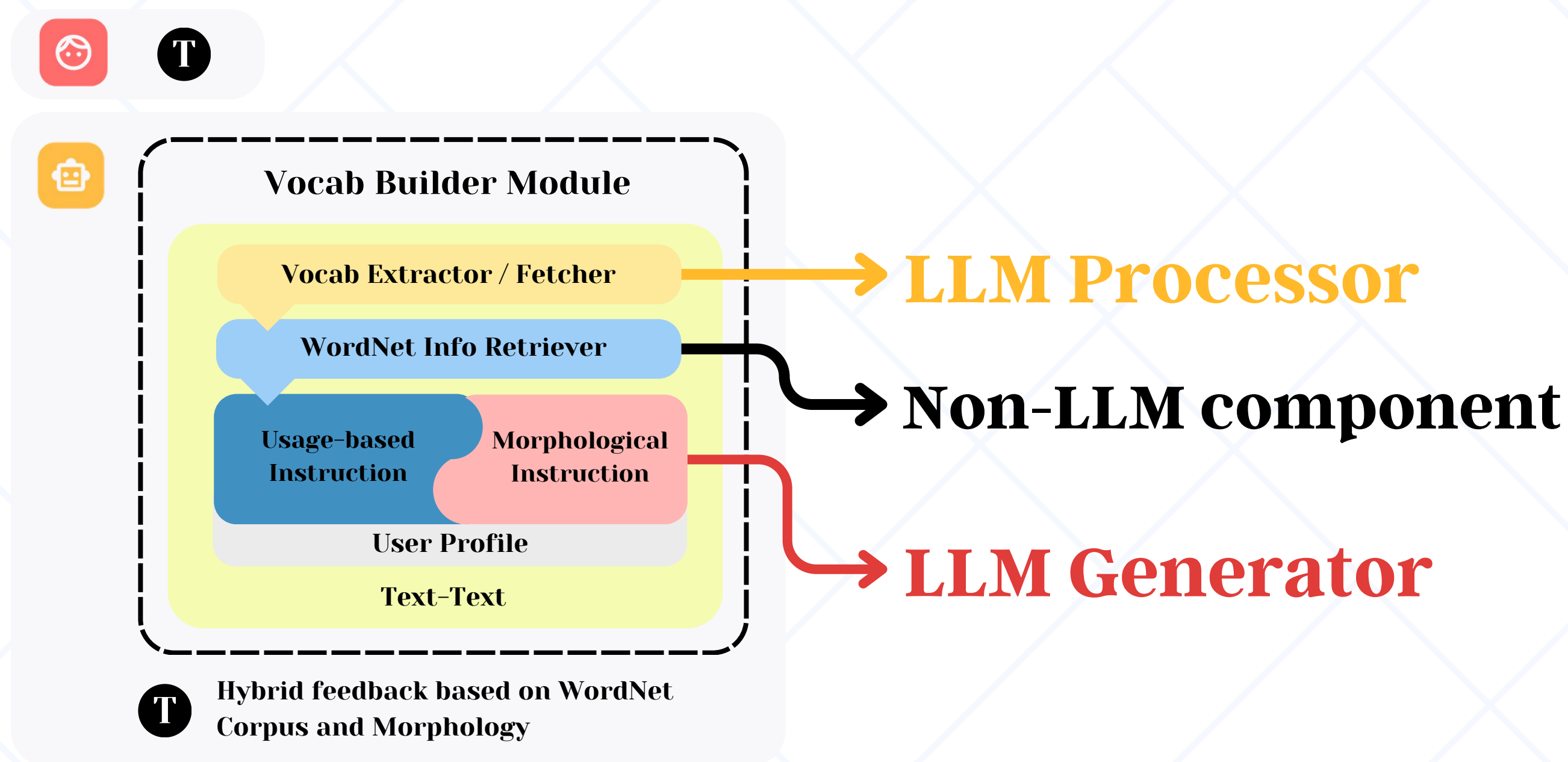
Prompt Design

Generator

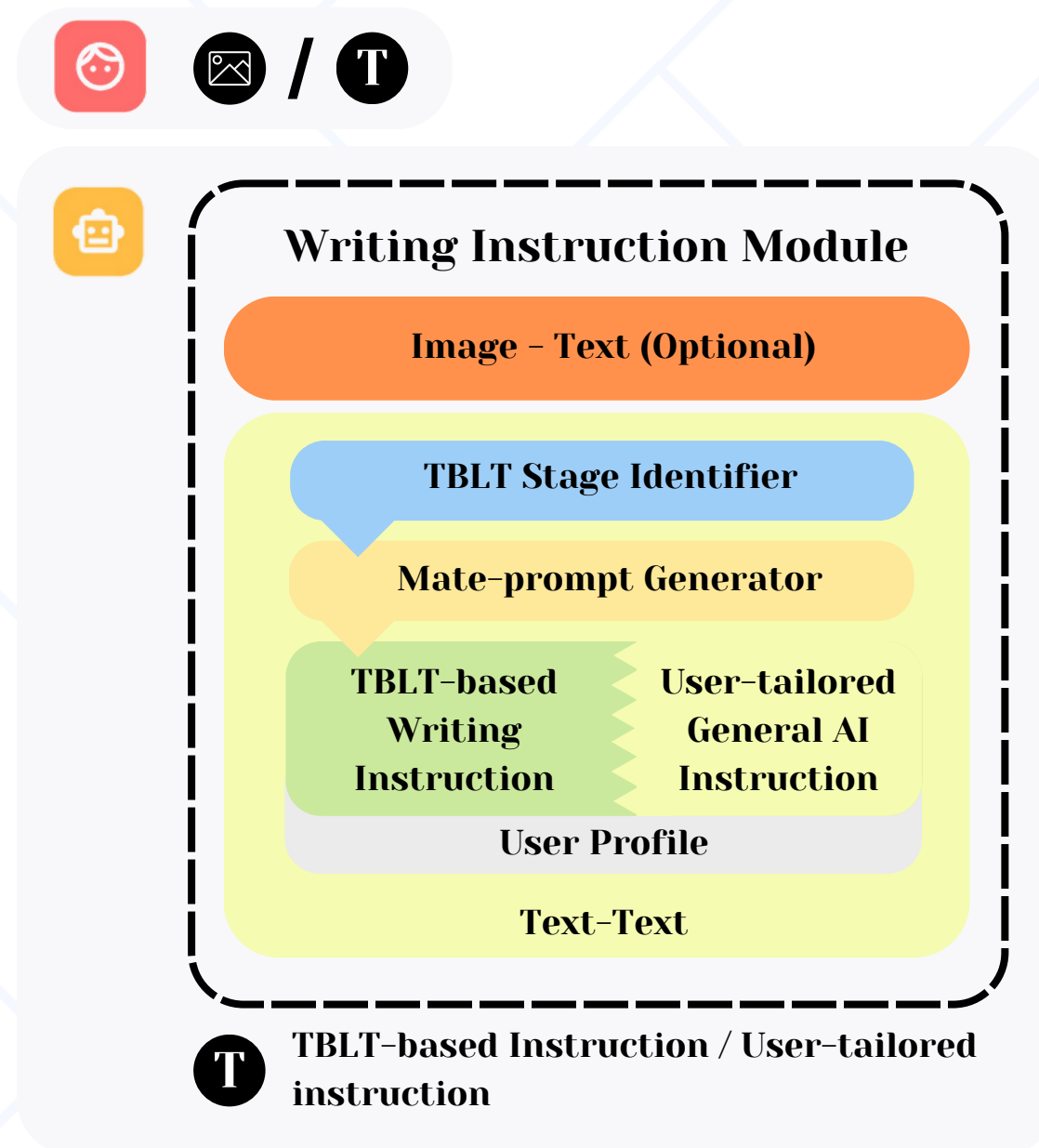


Processor

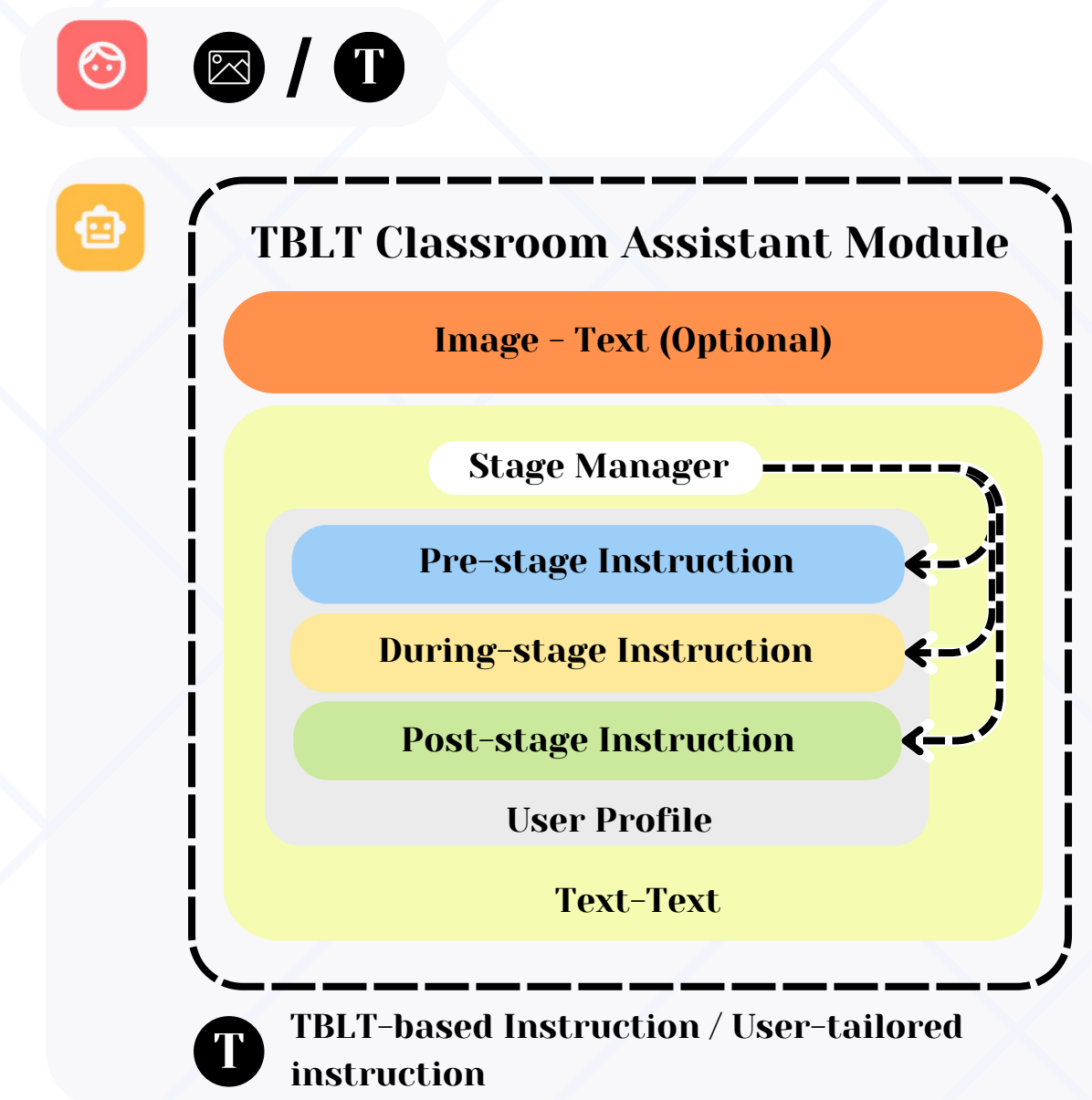
Resource Fetching



Two Types of TBLT Design

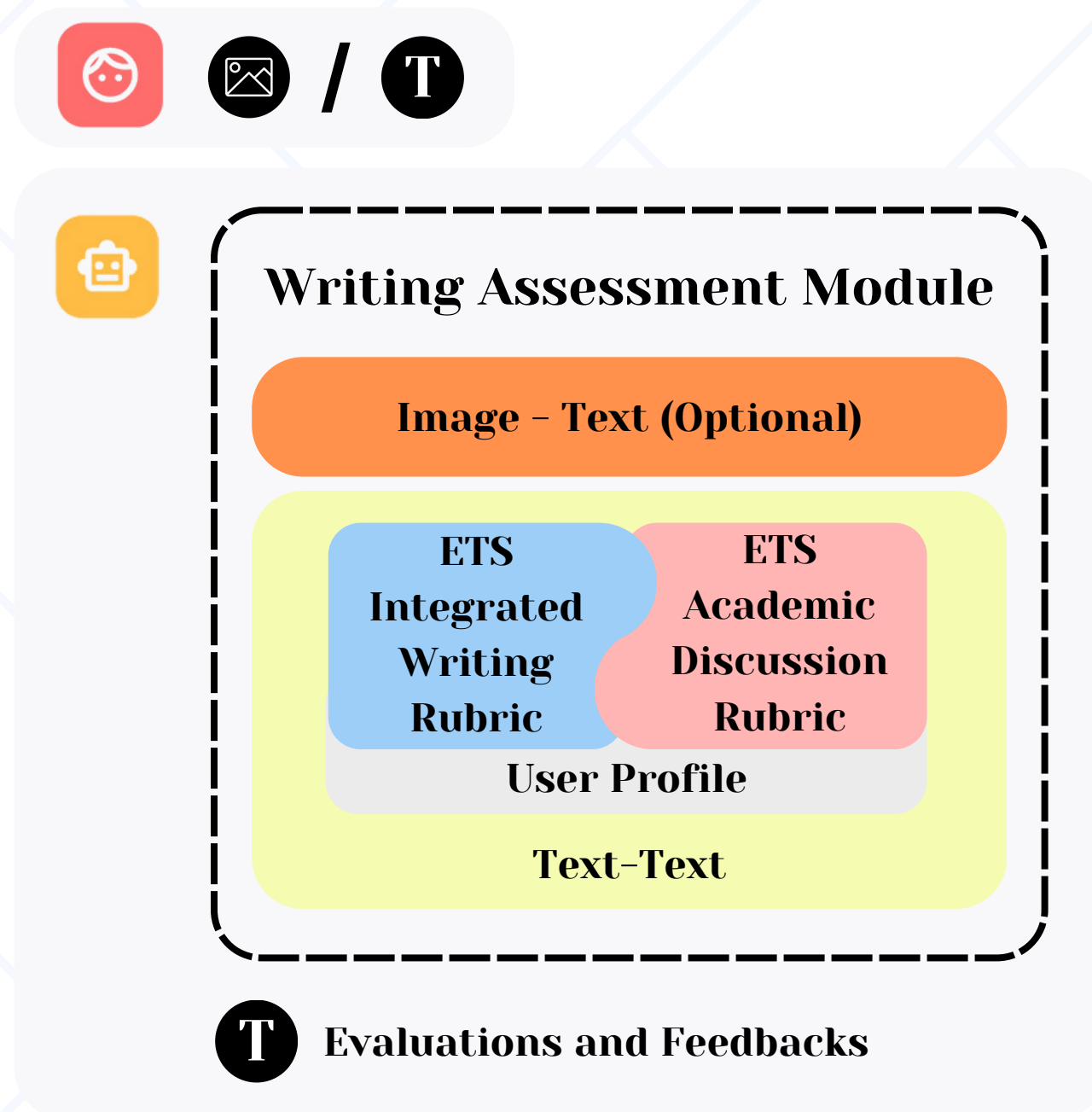


Student-centered Adaptive Instruction

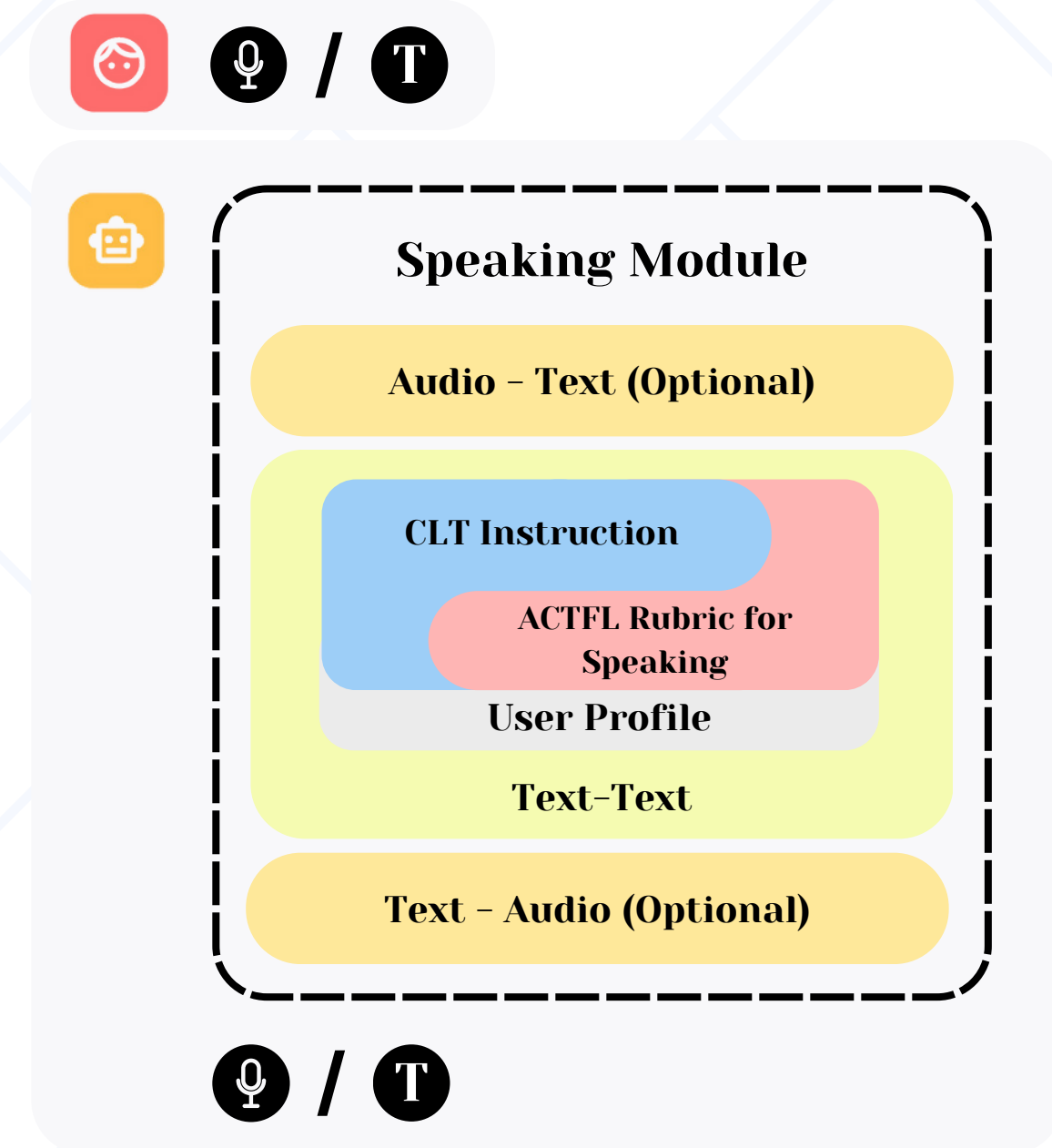


Teacher-directed Instruction with Defined Lesson Plan

Assessment & Feedbacks



Assessment Only

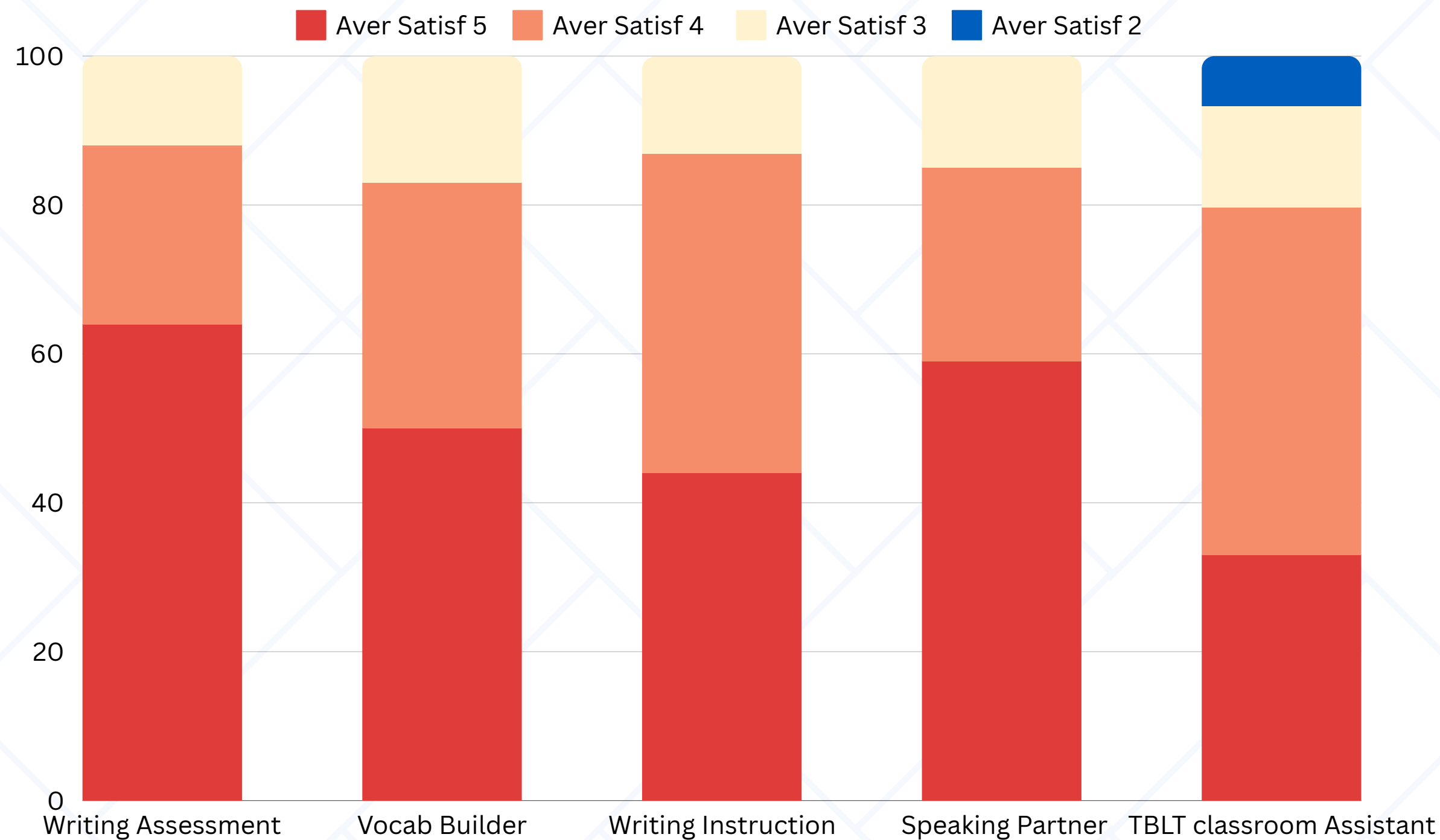


Assessment for Feedbacks

User Studies

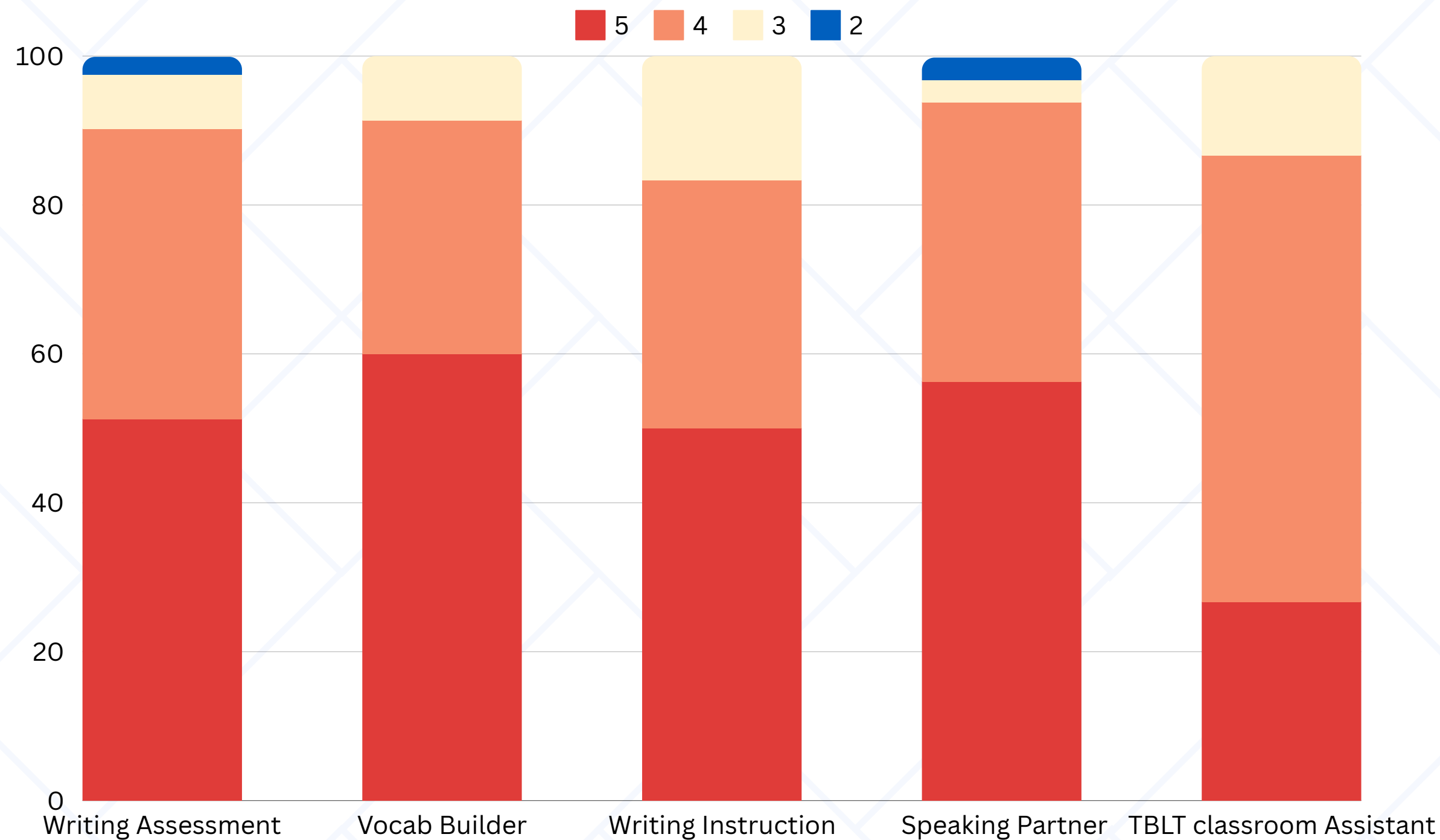
- Number of responses per functionality:
 - Writing assessment: 33
 - Vocabulary builder: 36
 - Writing instruction: 35
 - Speaking partner: 41
 - TBLT classroom assistant: 15
- Satisfaction level calculation:
 - Multiple questions asked for each aspect
 - Overall satisfaction level = average of all scores for that aspect:
 - Level 5: 4.5 to 5.0
 - Level 4: 3.5 to 4.49
 - Level 3: 2.5 to 3.49
 - Level 2: 1.5 to 2.49
 - Level 1: below 1.5

Evaluation on Input Understanding (6 Qs)

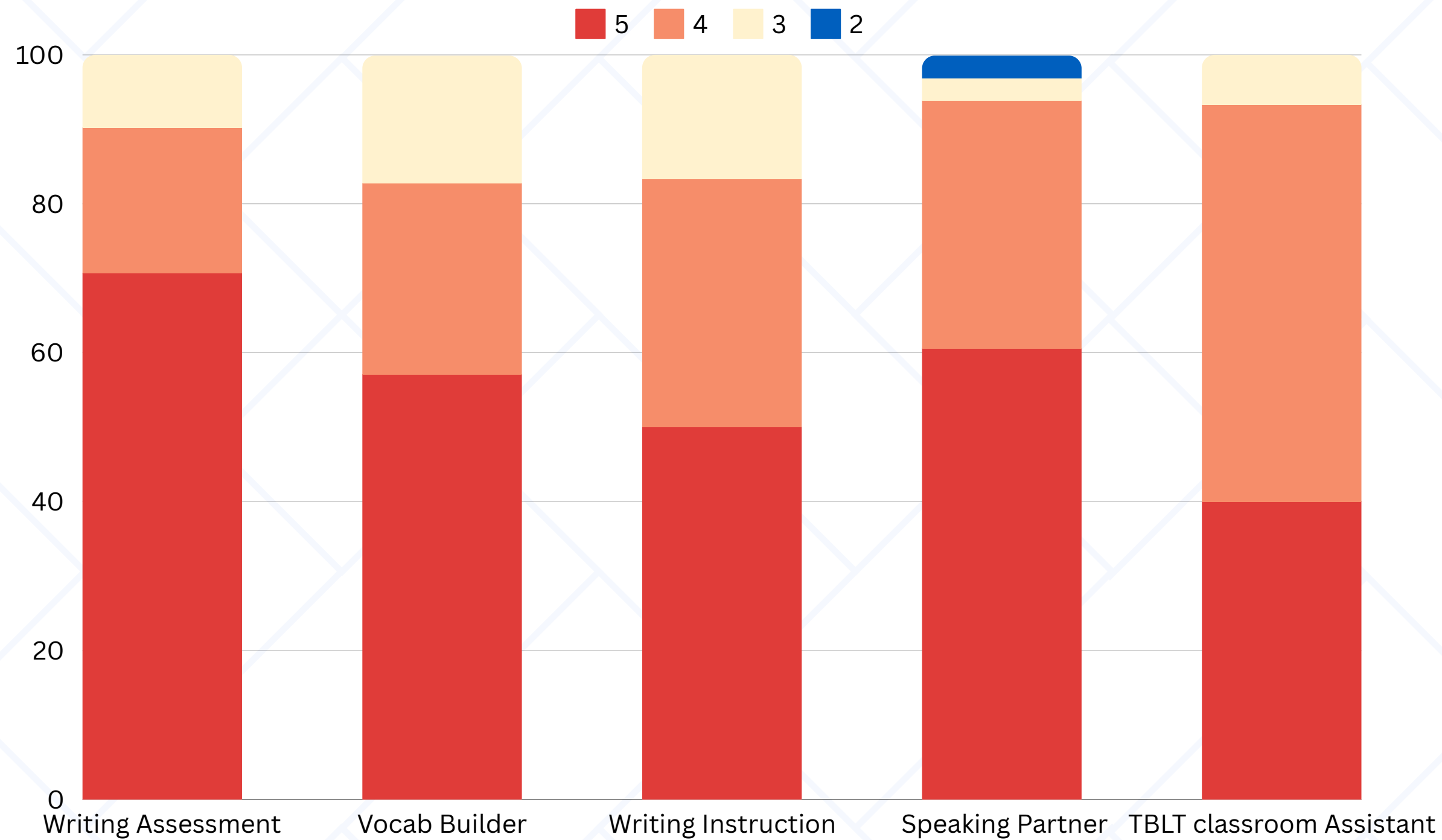


04

Response Evaluation (8 Qs)



Pedagogical Approach Evaluation (7 Qs)



05

Writing Assessment Demo

Deploy

LanPIP - Login Page

Please contact Faceia at fh2450@tc.columbia.edu if you encounter any trouble logging in.

Username

Password

Login

Thank you for using! Please help us fill out the survey of the function you experienced!

[Speaking AI Evaluation](#)

[Vocabulary AI Evaluation](#)

[Writing Assistant AI Evaluation](#)

[Writing Assessment AI Evaluation](#)

[TBLT system on itinerary](#)

05

Vocab Builder Demo

×

Deploy ⋮

This application is currently implemented with:

- Writing Assistance 📝
- Vocabulary Building 📖
- Speaking Practice 🗣️
- Writing Assessment/feedback 📄
- An example of TBLT classroom assistant on task "Write about Your Travel Plan" 🗺️

Upload only

Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG

Browse files

LanPIP

If you are stuck or want to switch to another function or want to start a new conversation, please try to refresh the page and log in again.

🤖

How could I help you today? 😊

Hi, I'm Yuki, an 18-year-old Japanese student in my first year at a US college. I'm writing a paper on COVID-19's impact on global education for my English class. As a non-native speaker, I'm having trouble with the specific vocabulary related to the pandemic. I'd like help learning and understanding more COVID-19 related terms to use in my paper

>

Thank You!

