

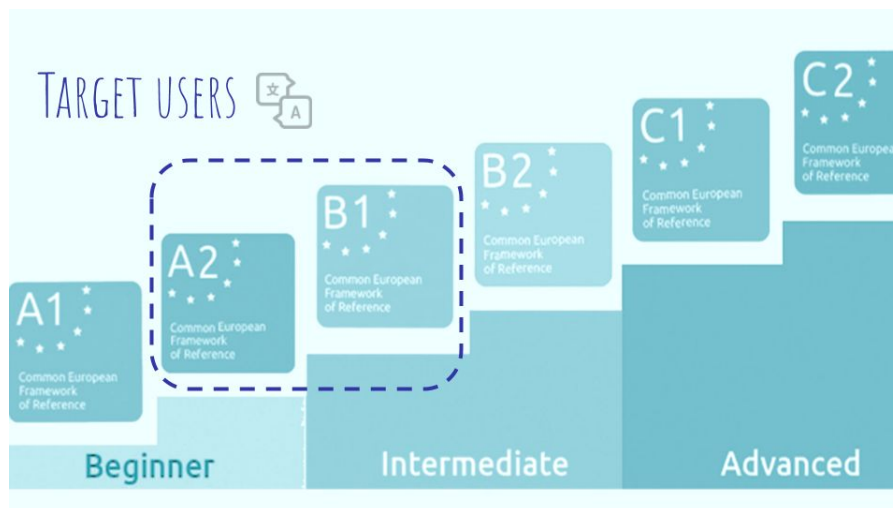
# User Research Summary

## Question(s) you were answering

- When engaging with AI, would users feel more engaged or anxious?
- Would they feel less embarrassed about making mistakes?
- Would they try harder or less hard to communicate well (operationalized by the usage of language support tools)?

## Who did you test with?

- Chinese English Language Learners living in the US



- One Female Chinese ESL student at a community college; move here from China with her husband, no job; spouse VISA
- female visiting scholar (AI condition)
- and one visiting scholar currently taking English Lessons on campus (human condition)

## How did you recruit and what compensation was given?

- Reached out to friends who fall under our target user description and use them as gatekeepers for finding users from the local Chinese community
- Sent messages on the Chinese student groups (Chinese students who came to the state this year, asking for participants who have lived in the US under 2 years.
- We will send out one \$20 gift card to each participant

Test plan (environment test took place in, tasks tested or experiment setup)?

- Run the experiment either in Nathan's office in Gates or LDT studio when space is free from people outside of this study to ensure psychological safety

Test Plan/Experiment Design:

- For the experiment, get ready:
  - [Task Introduction](#);
- Brief debrief about what they should be expecting
  - A simple conversational task
  - Telling them who they are conversing with (conditions)
  - Freedom to use whatever tools they want to use to prepare and use during the task
- Obtain the signed consent form for recording
- Zoom Link and **record** <https://stanford.zoom.us/j/3492439756>
- Experimenter set up:
  - One experimenter (facilitator) in room with the participant to explain in participant's native language (Chinese)
  - One experimenter (controller) hiding in another room, controlling the mechanic's actions and speech through the control panel
  - Zoom video conferencing to send user video and audio (including both user's and mechanic's speech) to the controller in the other room
- Shown them the instruction in Chinese
- When they're ready, confirm if they understand the task. Key is to do whatever they would normally do.
- Finish the task
- debrief

[Detailed] Experiment Setup:

- Model real-life situations of conversing with a mechanic to describe car problems, schedule an appointment to fix it, and arrange transportation options such as car rental with the mechanic.
- Experimental procedure (should be done ideally by at least 2 experimenters)
  - 1. Debrief the participants about the goal of the experiment and tasks to expect, right to withdraw, and ask for informed consent.
  - [Ask for consent to record the session; Screen record the engagement if testing virtually]
    - For data collection:
    - If the virtual meeting, use zoom or hangout to record the entire session.

- Ask the participant to share their screen and do screen recording.
  - Participants can hear the speech from the “AI” (pre-recorded)
  - Observer observes and notes:
    - if they understand the synthesized speech
    - The need to react quickly
    - If they use tech to help
- 2. Tell the participants either they are conversing with a real person or an AI
- 3. Give the participant the slides with the mechanic background and written instruction in Chinese; after the participants click “continue”, they see the mechanic and a window with text of reminder information
  - Read and confirm if they understand the instruction in Chinese (if necessary)
- 4. Conduct the task
- 5. WoZ the AI and observe without interruption
- 6. Debrief + [Discussion Guide](#)

## User Testing Notes

May 21, 2019 - female visiting scholar; AI

Notes:

- [After putting on the VR set]
  - Why the instruction is not in English? I need to translate it...
- She did not prepare for the script or look up words during the process
  - “I don’t think it was necessary. In real life, I would just ask the person to explain it, point to it, or ask someone else.”
- She did not catch it when the mechanic asked if the car is “2005” [ideally, we want the participants to spot the intentional mistake and correct the mechanic]
  - Maybe we need to add a line that says “I don’t see any 2005 Camry”
- She was very goal-oriented, insisted to accomplish the goal
  - She mistakenly thought she had to get her own car back by Monday, and insisted on that.
  - “Can you find another way to solve the problem? ... can you call the engine shop to delivery the engine during the weekend?”
- During the final debrief, the participant said she kind of knew that it was not an AI
  - **AI is too good - “It can understand my English” “its reaction is too good”**
  - **The delay is too long.**
  - **Wish there could be a pop-up when she asks for words she didn’t understand.**
  - She has never been in a situation like this before
  - She thinks she learned from this experience, maybe words
    - it’s hard to tell for us if this is true or not. The question may be leading for her to say yes she learned. But it’s hard to ask this question without leading. Though through her personality, she seems to be honest about it.

- She asked why we have a guy's voice and not a female voice.
- Maybe next time:
  - Ask if they see this as a valuable experience
  - Maybe we need to add a line that says "I don't see any 2005 Camry"

May 24, 2019 - Female Chinese ESL student at a community college; move here from China with her husband, no job; spouse VISA

Notes:

- Fit target group description really well
  - This participant fits very well with our target population of people who are living in the US, lacks the opportunity to speak and practice English but want to get better
  - Attending ESL classes now. Have not got to the level to graduate from ESL but close.
- She was very excited and impressed by our prototype, particularly because she thought the AI can understand her English.
- Spent a long time preparing for notes
  - Those notes are the things she wanted to say and not the things she expected the other would ask
- Used google translate to prepare for the notes

Verbal Feedback

- She was very impressed and excited about this technology and felt sad that it wasn't an AI.
  - Felt she could use practice tools like this to practice English
  - Wish for more scenario
- "the VR looks very real"
- "some times before I go to a store, I think about what I want to say before going in"
- **"if that is a real ai, I would really like it" "the current ones can't recognize me very well"**
- "calling a store is very frustrating. sometimes they just hang up on me if they don't understand me"
- **"good for beginners learning English, but maybe not enough details for intermediate English level"**

Survey responses:

How did you feel about speaking English with the mechanic in this scenario?

---

对方状态很接近真实状态，用于英语学习十分有帮助，可以激发使用者的潜能，在以后的生活中减少对于修车情景的恐惧或惊慌；不过人物在回答用户问题时 首先，反应时间相对较长会有明显不是真实人类的感觉，其次回答用户问题用语相对单一对比于正常人类，如果增加些主动和用户互动，比如谈论更多的信息，可能会使机师有更自然的表达更有益于接近人类的感觉。

---

The responses from the other person were very close to reality. It's great for learning English. It can motivate learners to achieve their potentials and reduce the fear and panic around getting their cars fixed later in real life. But when responding to user's questions, the reaction time was relatively long. It gave me the impression that it's definitely not human. Second, the responses were relatively monotonous, compared to human's. If there are more active interactions, such as chatting about more other information, it may make the mechanic's responses more close to human.

How did you feel about making mistakes while speaking English in this scena...

---

担心他可能会听不懂

感觉很好，首先已知是软件，所以在练习时可以放松精神，多次尝试解释或说明自己的意图，其次，画面和人物整体感觉非常接近真实场景，在以后遇到真实场景时减少惊慌感并且在多次练习与场景人物的对话时，可以更多了解机师可能的表达方式，在真实场景中更好理解他人并表达自己

I feel great. First of all, because I know it's a software, so I was more relaxed during the practice, and attempted to explain myself and my intentions. Second, the scene and the character altogether feel close to real life. It lessens the fear and panic for future encounter of this situation in real life. With repeated practices with the character, I can learn more about different possible responses from the mechanic, and learn to better understand others and express myself in real life.

How did you feel about speaking English in a VR environment in this scenari...

---

非常好，很有兴趣做更多的场景

Very good. I'm interested in practicing for more scenarios.

May 31, 2019 - visiting scholar; currently taking English Lessons on campus; Human condition

- Didn't spend too much time to prepare
- Briefly looked up words for translation on her phone
- After the end of the experiment survey, she said she got confused when the other person said "Sorry, I didn't understand that. Can you say it again?"
  - I said everything in the beginning, and she/he didn't understand.

- I thought it was a game and the AI (she thought it was an AI; not sure if she thought so or got that impression from the survey she just took) has to follow steps
- Made a few grammatical and word choice mistakes
  - Had problems when trying to say “warranty”; used “guarantee” instead
  - The other person did not understand, so she repeated.
- Overall, she said it was really interesting. Even mentioned it will be great for her kids to learn.
  - She also gave some suggestions on how to make the task less open-ended
  - Such as giving a few options for the users to choose from and say

#### Survey responses:

How did you feel about speaking English with the mechanic in this scenario?

不知道应该回答到什么程度合适

I don't know to what extent I should answer his questions (from the mechanic) and give my answer.

How did you feel about making mistakes while speaking English in this scena...

担心他可能会听不懂

I worried that he may not understand me.

How did you feel about speaking English in a VR environment in this scenari...

非常有趣，不用觉得说错了而难为情

Very interesting. I don't have to feel embarrassed about my mistakes.

June 3, 2019 - visiting scholar; currently taking English Lessons on campus; Human condition

- Didn't take any notes or prepare for the script once she knew the information will be on the screen too.
- Acknowledged that she didn't have enough vocabulary. She wasn't sure if she was able to complete the task first but, but started it without preparing for it.
- Expressed that she learned new words such as "warranty" and new phrases she can use in this situation.
- Thought the the task was interesting.
- Said she would like to have her kids use it. Maybe we can have different levels of difficulty.

June 3, 2019 - visiting scholar; currently taking English Lessons on campus; Human condition

- The participant didn't want to do the task cuz she thought it might be too hard for her.
- Mentioned that if she had 50% vocabulary she might feel comfortable to do it, but she had less than 20%.
- She said she doesn't drive and have no knowledge in car, so don't know how to do this. But if the task is about something else that she knows, such as food and health she will be able to talk about it.
- Next step:
  - leave the participant do the study without the presence of the experiment
  - We may need to have several situations, or different topics
- Used the word “为难” (nervously embarrassed; difficult) to describe her feeling about the task
- Psychological safety?

June 3, 2019 - visiting scholar; currently taking English Lessons on campus; AI condition

- The participant mentioned that she probably need to look up a lot of words, but looked things up only once.
- When she didn't understand something, she asked "pardon" first. When the AI repeated the response without rephrasing, she asked about the word she didn't understand.
- Wonder if we could have different topics to choose from
- Reported feeling nervous, but went along without using too much technology support.
- Next step:
  - I (Joyce) don't think she interpret the last survey questions the way we intended to. Her responses were pretty brief and maybe because she needed to be somewhere else? But we may consider revise the Chinese survey questions.

## Feedback from Heuristic

- High Priority:
  - Splash screen
  - Thinking/Text bubbles
  - Exit, maybe "save", function
  - Consistent bullet point

- Acknowledgment - do nodding if feasible
- Replace the CRV
- Make the pointing animation
- Medium;
  - Controller and check on the checklist
  - First no car - one pops up - The wrong car - disappear - right car pops up
  - Make a dictionary button and have users speak into it to look up words
- Low:
  - Facial expression to correspond to talking or not talking
- Script changes:
  - The bot should use more language like “Got it” or “Hmm” or “Ahh” or other filler language tokens.
  - To make the experience more realistic, the mechanic should be very knowledgeable about all the available rental cars
  - Handing over the keys or change the wording
  - Make the end goal clear [Severity 4]
  - Users may feel confused about the long silence and not be sure how to resume the conversation or whether to provide a different response. A solution would be to add voice lines for prompts (e.g. “Sorry, I didn’t catch that,” “I can’t find that model, did you mean something else?”)
  - The user doesn’t receive enough verbal feedback from the bot. This problem is twofold, since the user may not be sure whether their response was interpreted correctly, but also, as a language-learning application, the user is missing out on an opportunity to practice reinforcement of language by hearing their response repeated. This could be resolved by repeating back what the user said in the speaker’s voice.