Welcome to the MultiVRse

CS377E | Spring 2019
Nathan D., Glenn D., Joyce H., Angelica W.

MultiVRSE

A VR world that immerses a user in **realistic**, **culturally aware**, scenarios with **AI-powered** language support tools that evoke nuanced and goal-oriented language learning experiences.

Experiment design

HUMAN VS. AI
ON THE OTHER SIDE

(In both conditions, users are allowed to use any technologies, such as their phone or dictionary, they would normally use to help them communicate)

Research questions => Tasks

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 usage of language support tools)?
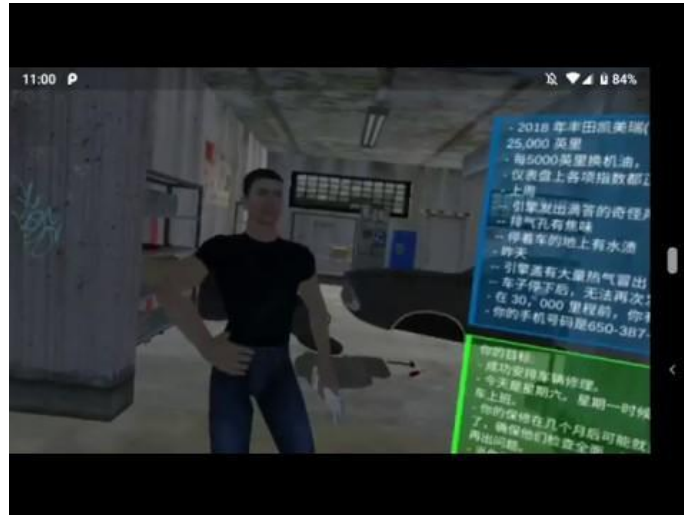
Major New design changes

Added pre- and post-test Qualtrics surveys to gather data about perception of the scenario

(manipulation check, foreign language anxiety scale based on literature)

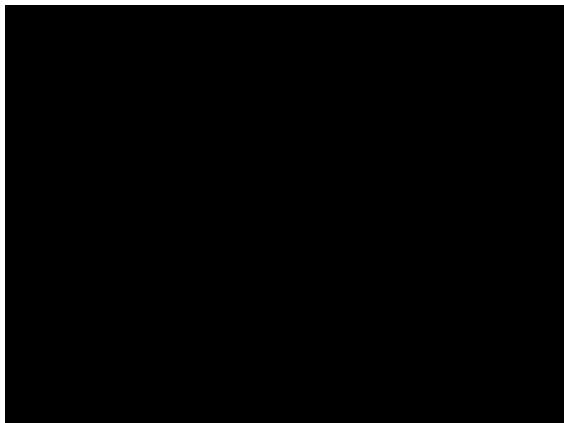# Major New design changes

Mechanic character more human-like

(instead of zombie)

Major New design changes

Add visual feedback for when the mechanic is "thinking"
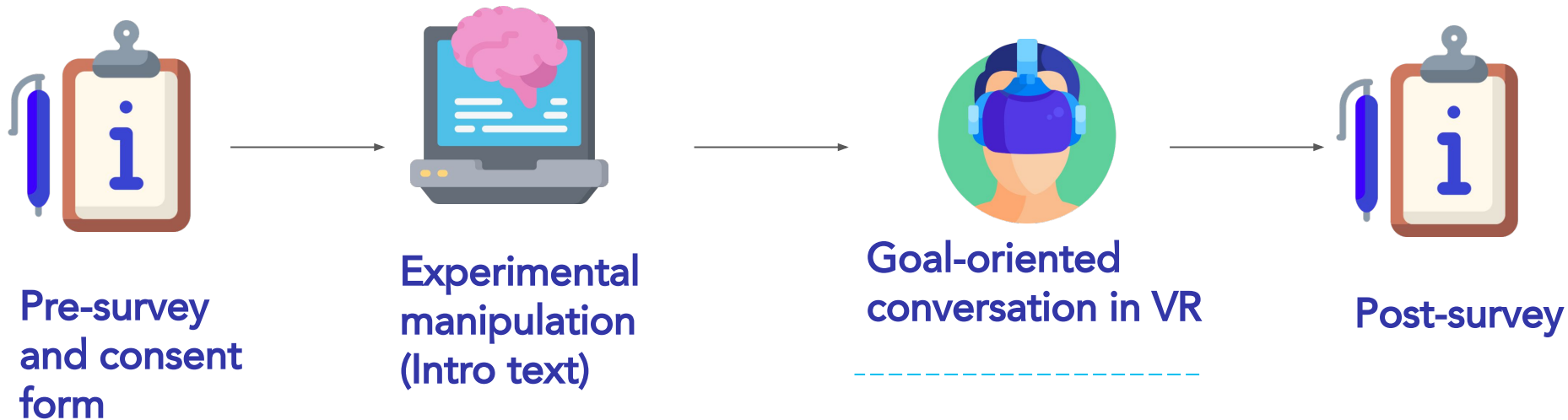
[Before]

[After]

Additional changes

- Mechanic character moves when speaking

- User's car is no longer overlapping user's position

- **Added more pre-recorded responses for mechanic ("I see.", "Got it.", etc.)**

- Reworked line about handing over keys

- Added rental cars to scene

- Two conditions for intro text (Your partner is controlled by another human / your partner is AI)

User Task Flow

Pre-survey and consent form

Experimental manipulation (Intro text)

Goal-oriented conversation in VR

Post-survey

[Behind the scenes]

Adaptive WoZ script, text-to-speech

欢迎参与我们的实验！

MultiVRse 是斯坦福大学计算机377课程的项目成果之一。项目测试参与者的数据将会被用来评估和改善 MultiVRse的界面和交互。数据将会由采访，观察和问卷形式采集。

参与为自愿性质。参与者可在任何时候退出，并无需担心后果。参与完成后，你将会得到20美金的亚马逊礼品卷。如果有任何关于这个实验的疑虑，你可以跟实验人员或者 James Landay 教授（计算机377的教授）沟通。联系方式在本文末端。

我们会把数据和姓名分别记录来保证参与的匿名性。数据只用参与者号码标记确认。除了学生研究人员和他们的导师外，这个课程以外的人无法接触到任何能追溯参与者身份的信息。

在这里，我确认我被给予了问问题的机会——任何关于这个研究和与我的参与相关的问题。我同意对我与MultiVRse实验互动和观点数据的采集。并且我同意实验人员采集照片和录像信息并用于报告和发表，只要它们不含任何能追溯我身份的信息。我理解，我能在任意时候退出这项实验。

研究人员：
Nathan Dass (ndass@stanford.edu)
Glenn Davis (gmdavis@stanford.edu)
Joyce He (jiayuhe@stanford.edu)
Angelica Willis (arwillis@stanford.edu)

计算机377 导师：
James A. Landay
CS Department
Stanford University
650-498-8215

Consent form

如果你还没有填写同意书，开始前请点击"这里"阅读。

在这个任务里，你对话的对象是人工智能驱动的。你的 VR 眼睛上有麦克风。说话时请尽量清晰明亮。

请阅读下面的情景。你可以用任何你喜欢和习惯的方式来准备这个情景，准备时间没有限制。一旦你开始这个情景，请只跟情景内的人物对话。实验组成员将不会参与你的情景对话中。实验过程中，你随时可以使用任何你需要的语言工具。

当你准备好的时候，请告诉任意实验组成员。

情景：

你在美国居住。你每天上下班单程20英里都是开着你的2018年丰田凯美瑞(Toyota Camry)（里程数为25,000 英里）。车子每开5000英里，你都会按时更换机油。至今为止，你的车子没有都出现任何问题。

上周，你发现你的车子有点怪怪的。你几次听到引擎有持续好几分钟滴答滴答的怪声音。上周三，你下高速的时候闻到通风口有焦味，但当你到达办公室时，气味消失了。并且，你在自家车库里发现你常停车的地上有水渍。车子的仪表盘上各项指标都没有显示有任何问题，一切看起来都很正常。

昨天，你从家开去上班的途中，引擎盖中喷出大量热气。你在路边停下车。在等了10分钟热气消散后，你还是无法启动你的车子。拖车公司把你的车拖到修理厂，现在你正在和修理人员讨论如何修理你的车子。在你车开到30,000里程钱，你有全额保修，所以不需要担心价格。

今天是星期六。星期一时候你得需要开车上班。你的保修在几个月后可能就里程满失效了。你的目标是修理好你的车并星期一能顺利上班，并保证修车场检查所有问题，以防保修过期后再出问题。

你的名字：[你的名字]
你的电话号码：650-387-9745

你的目标 (任务过程中屏幕上也会显示)：

- 成功安排车辆修理。
- 今天是星期六。星期一时候你得需要开车上班。
- 你的保修在几个月后可能就里程满失效了，确保他们检查全面，以防保修过期后再出问题。
- 当你觉得任务完成结束后，告诉工作人员你任务完成了。

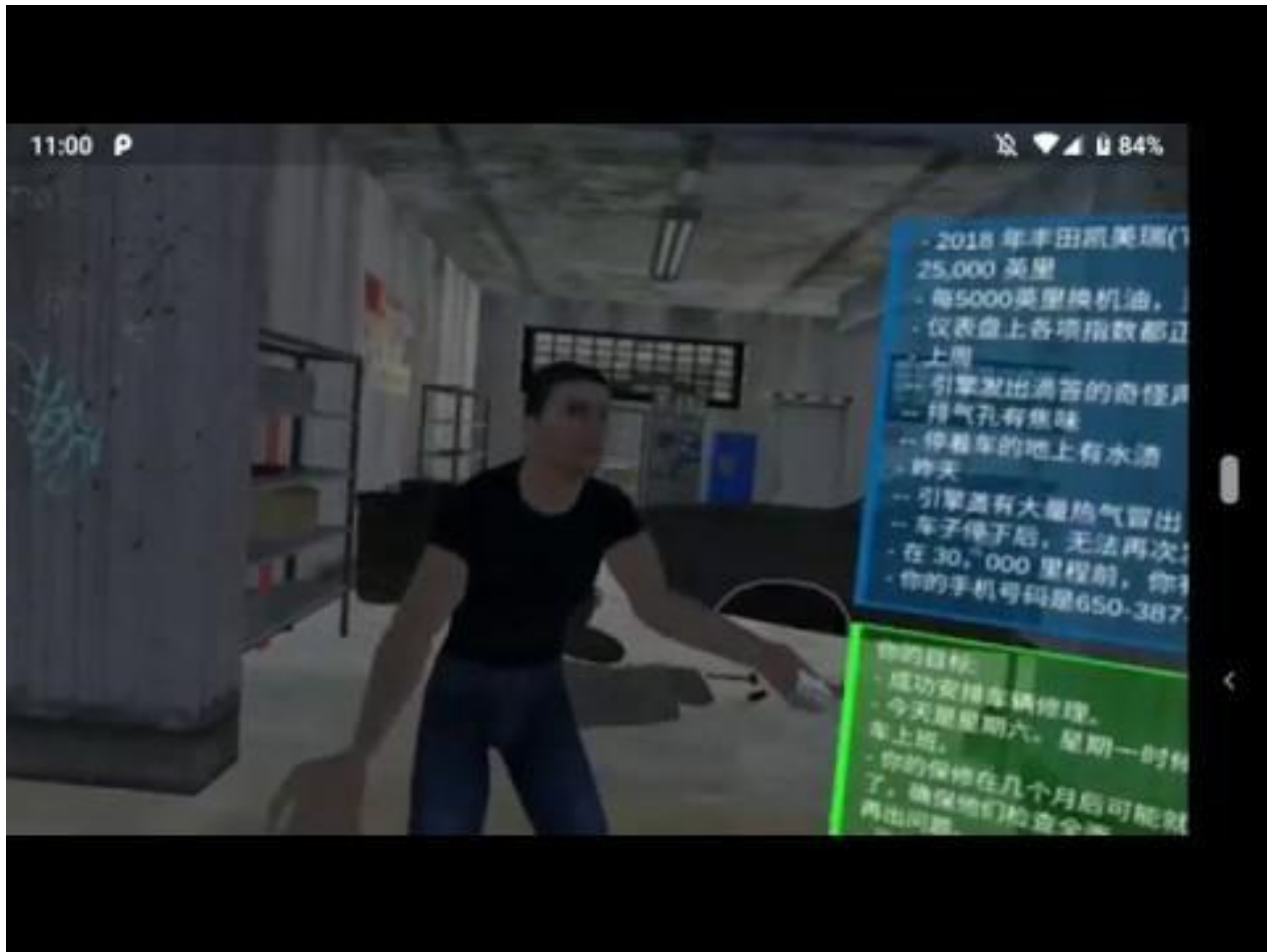## 任务提示 (任务过程中屏幕上也会显示)：

- 2018 年丰田凯美瑞(Toyota Camry)，25,000 英里
- 每5000英里换机油，至今没有出现问题
- 仪表盘上各项指数都正常
- 上周
  - 引擎发出滴答的奇怪声响
  - 排气孔有焦味
  - 停着车的地上有水渍
- 昨天
  - 引擎盖有大量热气冒出
  - 车子停下后，无法再次发动
- 在 30,000 里程前，你有全面保修
- 你的手机号码是650-387-9745

在完成你的 VR 场景模拟后，请点击"这里"填写问卷。

Updated Introductions

including experimental manipulation:
"Your partner is a human" or
"Your partner is an AI"

Video of current build

Post-test survey

- Experimental manipulation check
- Foreign language anxiety scale

Please choose the answer that best reflects your feelings about each statement.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I never felt quite sure of myself when I was speaking English with the mechanic.	○	○	○	○	○
I didn't worry about making mistakes when speaking English with the mechanic.	○	○	○	○	○
It frightened me when I didn't understand what the mechanic was saying in English.	○	○	○	○	○
I panicked when I had to speak English without preparation in this scenario.	○	○	○	○	○

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I worried about the consequences of failing to achieve the goals of this scenario.	○	○	○	○	○
In this scenario, I got so nervous I forgot things I know.	○	○	○	○	○

Speaking with the mechanic felt like speaking with a...

Definitely human	Probably human	I don't know	Probably AI	Definitely AI
○	○	○	○	○

Experimenter task flow

One experimenter (facilitator) in room with participant to explain in 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

MultiVRse control panel

Line to speak | Submit

LISTENING

Sorry, I didn't understand that. Can you say it again?

Hi there, welcome to Joe's Toyota dealership. Can I have your name please?

I see your car was towed in yesterday. The 2005 Camry, right?

We towed you, right?

Are you sure? I can't find that car.

Ok. Can you describe the problem?

When did you start noticing these problems?

Was the temperature of the car higher than normal?

Did the check engine light come on?

Was anything illuminated on the dash?

Did you hear any noises from the engine?

Where is the smell coming from?

Where were the wet spots?

Yeah, you can see the engine damage here.

Do you want a full replacement of your engine? It will cost $5,000.

Yes, you have full warranty with us, so everything is covered.

Intro 1 (AI)

Intro 2 (Human)

RESET

THINKING

Yes. | Got it. | I see. | No.

OK, let me look it up.

Your car was brought into the garage yesterday. The 2005 Camry, right?

The '05 Camry, right?

Oh my mistake. The 2018 Camry, right?

Did you notice any problems before yesterday?

Did you notice anything happening with the temperature gauge?

Was anything illuminated on the dash?

Were there any lights visible on the car's dashboard?

How often do you get the oil changed?

Did you smell anything strange?

Did you notice any leaks or wet spots on the ground?

It sounds like your engine overheated. Let me take a look at it.

Looks like you had a coolant leak. The engine is completely destroyed.

Oh I'm sorry. Let me check …

What's your phone number? I'll call you when it's ready to pick up, it may take a while to fix.

WOZ Dashboard Updates

Data Collection

Engagement with the task
    How users prepare for the task
    Time spent on the task

Usage of language support tools
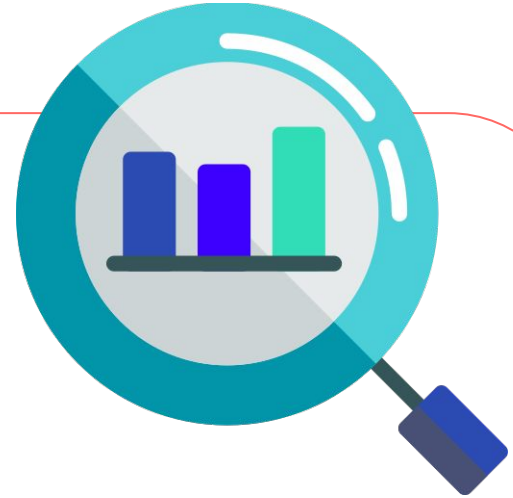   Tools they use
    Level of engagement/reliance

Engagement with the "AI"
    # times participant ask for repetition and clarification

Attitudes (Post-survey)
    Foreign Language Anxiety Scale
    Engagement with AI

User 1

Chinese Visiting Scholar recruited from Chinese Students WeChat groups

Female Chinese ESL student at a community college; moved here from China with her husband.

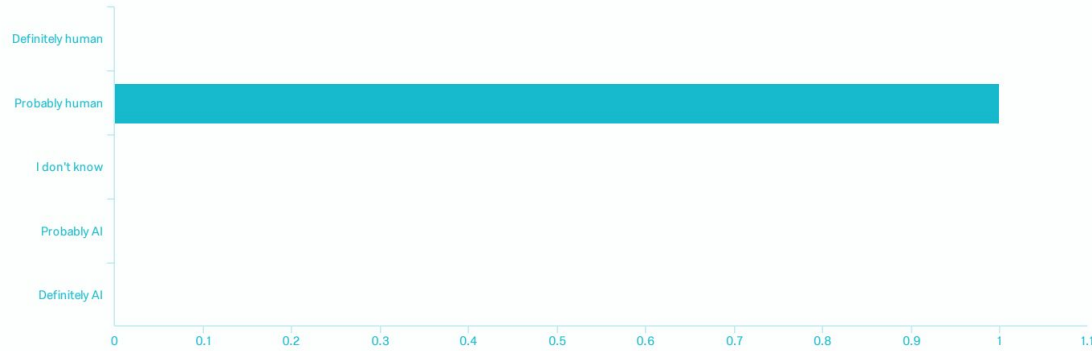Recruited from a gatekeeper from local Chinese community

User 2

Results

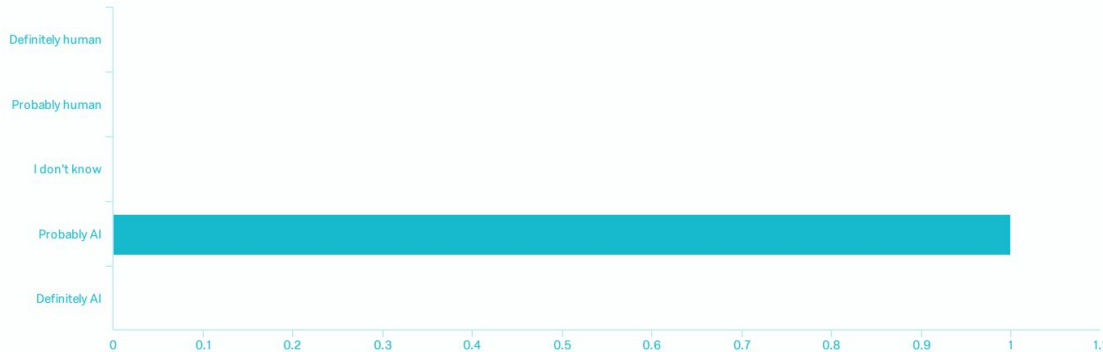## Q3 - Speaking with the mechanic felt like speaking with a...

Q5 - I believe that my partner (the mechanic) was controlled by...

Experimental manipulation check

(Condition = AI, user believed the mechanic was actually controlled by AI)

Foreign language anxiety scale

#	Field	Strongly agree		Agree		Neither agree nor disagree		Disagree		Strongly disagree		Total
1	I never felt quite sure of myself when I was speaking English with the mechanic.	0.00%	0	0.00%	0	100.00%	1	0.00%	0	0.00%	0	1
2	I didn't worry about making mistakes when speaking English with the mechanic.	100.00%	1	0.00%	0	0.00%	0	0.00%	0	0.00%	0	1
3	It frightened me when I didn't understand what the mechanic was saying in English.	0.00%	0	0.00%	0	0.00%	0	100.00%	1	0.00%	0	1
4	I panicked when I had to speak English without preparation in this scenario.	0.00%	0	100.00%	1	0.00%	0	0.00%	0	0.00%	0	1
5	I worried about the consequences of failing to achieve the goals of this scenario.	0.00%	0	0.00%	0	0.00%	0	100.00%	1	0.00%	0	1
6	In this scenario, I got so nervous I forgot things I know.	0.00%	0	0.00%	0	0.00%	0	100.00%	1	0.00%	0	1
7	I felt confident when I was speaking English with the mechanic.	0.00%	0	0.00%	0	100.00%	1	0.00%	0	0.00%	0	1
8	I felt very self-conscious about speaking English in front of the mechanic.	0.00%	0	0.00%	0	0.00%	0	100.00%	1	0.00%	0	1
9	I got nervous and confused when I was speaking English with the mechanic.	0.00%	0	0.00%	0	0.00%	0	100.00%	1	0.00%	0	1
10	I got nervous when I didn't understand every word the mechanic said.	0.00%	0	100.00%	1	0.00%	0	0.00%	0	0.00%	0	1
11	I got nervous when the mechanic asked questions that I hadn't prepared in advance.	0.00%	0	0.00%	0	100.00%	1	0.00%	0	0.00%	0	1

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132.

Usability study

Q10 - How did you feel about speaking English with the mechanic in this scenario?

A: 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 a 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.

Usability study

Q11 - How did you feel about making mistakes while speaking English in this scenario?

A: I feel great. First of all, because I know it's 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 a 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.

Usability study

A: very good. I'm interested in practicing for more scenarios.

Q12 - How did you feel about speaking English in a VR environment in this scenario?

Insights and Results

1. Our "AI" appears to be substantially better than current solutions on the market, especially at recognizing and understanding participants' speech. We may want to add some artificial "misunderstandings" to maintain the illusion of a functional AI.

2. One user spent about 10 minutes preparing before the VR scenario, writing down things she wanted to say; the other user did not prepare at all. This is likely an individual difference—most previous participants spent time preparing.
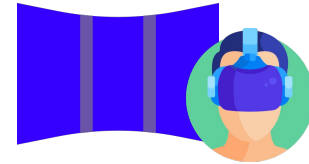
Next Experiment [RE]design Focus
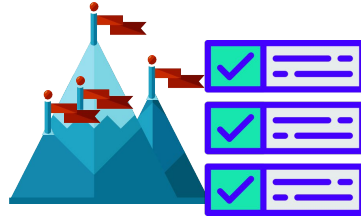
Ability for user to mark goals as completed

Language support tools within the scenario

Splash screen before starting the scenario

Remaining questions, concerns, worries

- Recruiting participants (small sample size makes it difficult to generalize)
- Hawthorne effect (presence of experimenters may affect participants' actions, feelings, etc.)
- User's language proficiency may affect results

안녕 CIAO HOLA

HALLO 你好 HELLO

สวัสดี BONJOUR

こんにちは

THANKS!

[IN ADVANCE, FOR YOUR FEEDBACK]