InstantTranslate

Final Report of Week 1 (02266)

Afaf Amajjan (AA)
Technical University of Denmark
s230920@student.dtu.dk

Leonardo Rodovero (LR)
Technical University of Denmark
s240095@student.dtu.dk

Ömer Coskun (ÖC)
Technical University of Denmark
s230926@student.dtu.dk

Rahima Akhter (RA)
Technical University of Denmark
s225152@student.dtu.dk

Zahedullah Wafa (ZW)
Technical University of Denmark
s224316@student.dtu.dk

Zaid Sheikh (ZS)
Technical University of Denmark
s224309@student.dtu.dk

1 INTRODUCTION (LR, AA)

Effective communication across language barriers is a persistent challenge in an increasingly globalized world [4]. Travelers and exchange students frequently encounter difficulties in understanding written and spoken foreign languages, especially in environments where traditional translation tools, such as smartphones or handheld devices, are inadequate and require constant attention. The **InstantTranslate** prototype, leveraging **augmented reality** (AR) glasses, aims to address these issues by providing **real-time** translation of written digits, spoken language, and dialogue between 2 individuals while focusing on the environment around the user. This seamless interaction aims to enhance accessibility and participation in diverse settings. This innovative approach not only empowers users to navigate linguistic barriers effortlessly but also fosters deeper connections in multicultural environments. With its ability to handle diverse scenarios, InstantTranslate can redefine communication, bridging language gaps in real time with precision.

2 EXISTING WORK (LR, ZW)

The competitive landscape for real-time language translation is marked by innovative solutions addressing specific user needs, yet gaps remain in providing an all-encompassing experience. **Solos' AirGo3 Smart Glasses** stand out with their AI-powered SolosTranslate, which supports multilingual group communication by converting speech to text in real-time [6]. For example, the AirGo3 glasses use advanced AI algorithms to transcribe spoken words and display them as text on the glasses' display, making them effective in group settings but limited to text-based interactions.

Similarly, the **Leion Hey AR Glasses** focus on real-time transcription and translation, catering to dynamic conversational settings [2].

Hearview Subtitle Glasses target accessibility, offering speech-to-text conversion tailored for the deaf and hard-of-hearing community [1]. While these products excel in their respective domains, they often lack integration of both text and speech translation into a single, cohesive AR platform. This is where **InstantTranslate** offers a unique value proposition, combining real-time translation of written text and spoken language with AR glasses, providing an intuitive, hands-free experience designed to bridge language barriers in diverse environments seamlessly.

3 ITERATION #1 (ÖC)

In the *InstantTranslate* project, our hypothesis about the need for an efficient and multilingual translation solution with the use of AR glasses was tested through an initial validation process. We presented our idea in the form of a landing page (Fig.1) to our fellow peers and teaching assistants. The presentation aimed to introduce them to the market segment and the features of the AR glasses, specifically text and audio translations in different languages. This validation was essential to determine whether there was real interest and market potential for our idea, as well as to collect early feedback that could guide the further development of our prototype.

The validation method included a presentation of the overall prototype and the landing page that illustrated the AR glasses core functionalities and value proposition for the prospective users. After the presentation, feedback was collected through direct comments from peers and teaching assistants (see Appendix E), as well as through a voting process, where our idea was chosen as the best among competing projects, as seen in Fig. 1. This approach gave us valuable insight into users initial impressions and specific areas for improvement that we could work on.

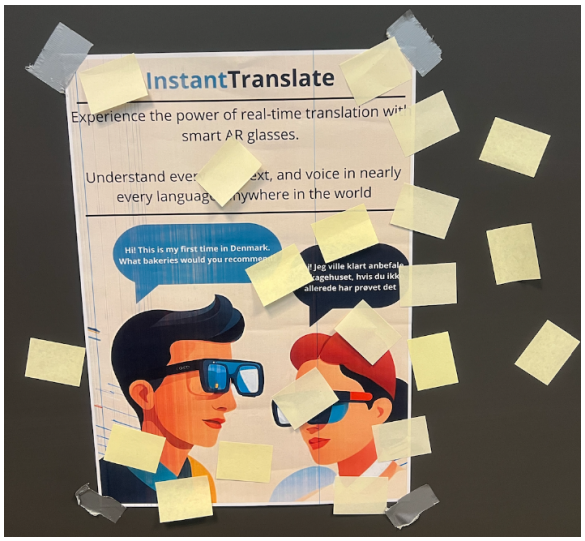

Figure 1: InstantTranslate poster

The validation results confirmed our hypothesis about the relevance of AR glasses and received positive feedback that highlighted its potential in the market. The baseline feedback pointed to several areas for improvement, including optimizing the communication on the landing page to more clearly demonstrate the interaction between the two people by making it look like they are talking to each other. In addition, we added visual cues as flags and languages with different alphabet to increase the recognition of different languages. Lastly, we added a call to action by including a QR code so that the prospective user know where they find more information about AR glasses and buy our product. These changes led to the planning of the next iteration, where we focused on optimizing the user interface of the prototype itself.

4 ITERATION #2 (AA, ZS)

What was Validated: The hypothesis was that AR glasses offering real-time text and voice translation could effectively address communication challenges in multilingual environments. Targeted at travelers and exchange students, which most of our fellow students have experiencedthe project aimed to enhance user experiences in reading signs and conversing in unfamiliar languages.

How we did: The validation process started with prototyping, including wireframes and storyboards, to visually represent the solutions functionality, such as translating text from signs and enabling real-time voice translation. Feedback was gathered by 5 fellow students through a peer review on FeedbackFruits, enabling systematic evaluation of usability and alignment with objectives. This iterative approach highlighted areas for improvement.

Outcome of Validation: The validation process yielded several key insights. Overall, the project was recognized for its innovative approach and its potential to address a pressing need within its target market. The user story map was praised for its structure and alignment with the projects goals, though opportunities for improvement were noted, particularly in enhancing task clarity and incorporating error-handling mechanisms. The wireframes, while commended for their clean design, required improvements in readability, the depiction of system states, and better clarity in specific scenarios such as noisy environments. The feedback highlighted the need to refine the wireframes by introducing clearer task flows and detailed system interaction points. Suggestions included the addition of customization options for user preferences and mechanisms to address environmental challenges such as background noise. The feedback (see Appendix E also showed that they generally had a hard time identifying who was speakers, and reading the subtitle text. To address these issues, we decided to improve the contrast between the background and the displayed text. As the feedback also suggested, we added a way to customize users settings, included the ability the adjust text size, giving the users more control over the AR-glasses.

Figure 2: Menu on Hand

As our prototype didn't require a lot of user interaction, we decided to add a new feature by adding a menu on the user's left hand. This feature allows user who cannot speak, to use their hands to activate/deactivate different features. This feature is quite similar to our eye tracing, and together these features will help users get a more immersive experience. We also change the placement of mute/unmute button, to be next you index finger making it easier for users to reach it.

5 ITERATION #3 (ÖC, ZS)

In our **think-aloud test** of InstantTranslate, we validated the hypothesis that users can intuitively switch languages and engage in a dialogue or translate text in the chosen language using AR glasses. We wanted to ensure that the prototype was user-friendly and could give an impression of how real-time dialogue would appear in a different language. This was important to demonstrate the systems ability to facilitate communication across language barriers, which is essential in both social and professional contexts.

The method for this iteration involved a think-aloud test with two students from the Lyngby campus area at the Technical University of Denmark (DTU), who are not taking this course. Students were asked to perform three specific tasks by using the AR glasses, and these tasks needed to provide an understanding of the central functions that prospective users would use regularly, to make sure the relevant parts of the user interface can be optimized [3]. **The tasks were the following:**

(1) Switch languages and engage in a translated dialogue.
(2) Focus on a restaurant menu to have it translated.
(3) Extend your hand to access settings and interact using the other hand.

Test Procedure:

(1) Participants verbalized their thoughts and actions during the test.
(2) One group member acted as the facilitator.
(3) The remaining group members observed and noted feedback without influencing participant interaction with the prototype [5].

From the think-aloud test, we learned that the participants were generally able to switch languages and engage in dialogues without major problems, which confirmed part of our hypothesis about

the usability of the system (see Appendix E). However, clearer instructions were needed in the user interface from the beginning instead of letting the system translate a dialogue right away. In the second task, we discovered that while the text recognition worked well, participants wanted some indication of how much of the text had been detected. Finally, the interaction with the settings showed that relevant user preferences for settings are included, but participants expected to find the battery percentage here, since this was not apparent in previous tasks. To immerse and guide the users in this new AR-world, we have decided to include configuration and welcome screen. This way the users can get a better understanding of the feature and how they can access them. These findings helped us guide the final iteration, where we had the opportunity to improve the user interface to make navigation more intuitive with clearer instructions and visuals, as well as optimize the accuracy of the system under translations.

6 DISCUSSION (RA, ZW)

The development and validation of the InstantTranslate AR glasses highlighted both the potential and challenges of real-time translation technology in diverse scenarios. Feedback from multiple iterations, including peer evaluations and think-aloud testing, revealed the strengths of the system and areas for improvement. One of the major achievements of our project was the validation of the core hypothesis: Users found the AR glasses effective in facilitating communication across language barriers. The system demonstrated intuitive usability for tasks such as switching languages, translating text, and engaging in dialogues. Positive feedback emphasized the innovative combination of text and voice translation in real-time, which significantly enhances accessibility. However, some challenges were identified during testing. Users expected clearer instructions before starting interactions, better visual feedback for text detection, and inclusion of battery status in the settings interface. Addressing these concerns required iterative improvements to enhance the user experience and make the interface more intuitive. Additionally, the systems capability to handle noisy environments remains an area for future exploration. Below is an example focusing on broader ethical considerations without specifically mentioning data handling.

Ethical Considerations:
From the customers perspective, ethical concerns primarily center on clear communication and cultural sensitivity when using real-time translation tools. Over-reliance on automated translations can lead to misunderstandings if nuances or context are lost, especially in critical situations. There is also the question of fairnessensuring all participants in a conversation are comfortable with the technology and understand its implications. By being aware of these factors, customers can use real-time translation responsibly and respectfully, supporting genuine dialogue rather than replacing it.

Overall, the iterative development process, supported by feedback from TA's and other students, enabled us to refine the prototype and establish its feasibility as a tool for breaking down language barriers. In future iterations, we would focus on refining the system for noisy and complex environments and incorporating user interface enhancements for broader usability.

7 CONCLUSION (RA, ZW)

The InstantTranslate AR glasses show promise in addressing the challenge of real-time language translation in a user-friendly and hands-free format. By integrating text and speech translation capabilities into a single platform, our prototype demonstrated significant potential to connect communications for travelers and students. Through validation and feedback, our project showcased how AR technology can be leveraged to provide seamless interaction in multilingual environments. Although the current prototype effectively handles one-on-one dialogues and text translation, future improvements can be made by enhancing system performance in noisy and crowded environments and further refining the interface based on user feedback. In conclusion, InstantTranslate represents a significant step forward in real-time translation technology with the potential to empower users worldwide and redefine communication in multicultural settings.

8 CONTRIBUTIONS

All group members contributed equally to the project, from developing the artefacts and prototype iterations to presentations and report completion. This collaboration ensured a well-balanced effort to include every group member's perspective in the project.

REFERENCES

[1] Hearview AI. [n. d.]. Hearview Subtitle Glasses: Real-Time Speech-to-Text Conversion. https://www.hearview.ai/products/hearview-glasses Accessed: 2025-01-09.

[2] Ltd Beijing LLVision Technology Co. [n. d.]. Hear with Your Eyes: Five AR Live Captioning Glasses. https://www.hearingtracker.com/resources/hear-with-your-eyes-five-ar-live-captioning-glasses Accessed: 2025-01-09.

[3] Rex Hartson and Pardha S. Pyla. 2018. *The UX Book: Agile UX Design for a Quality User Experience* (2nd ed.). Morgan Kaufmann, Cambridge, MA. https://www.amazon.com/UX-Book-Design-Quality-Experience/dp/0128053429

[4] Dennis Medved, Antonio Franco, Xiang Gao, and Fangfang Yang. 2013. *Challenges in teaching international students: group separation, language barriers and culture differences.* Genombrottet, Lunds tekniska högskola.

[5] Rolf Molich. 2003. User Testing, Discount User Testing. https://www.dialogdesign.dk/about-rolf-molich/ Unpublished manuscript.

[6] Solos. 2024. Solos AirGo3 Smart Glasses Expand Live Translate and Group Communication Capabilities Launching at 2024 CES. https://bit.ly/solos-airgo3 Accessed: 2025-01-09.

A LANDING PAGE (FINAL)

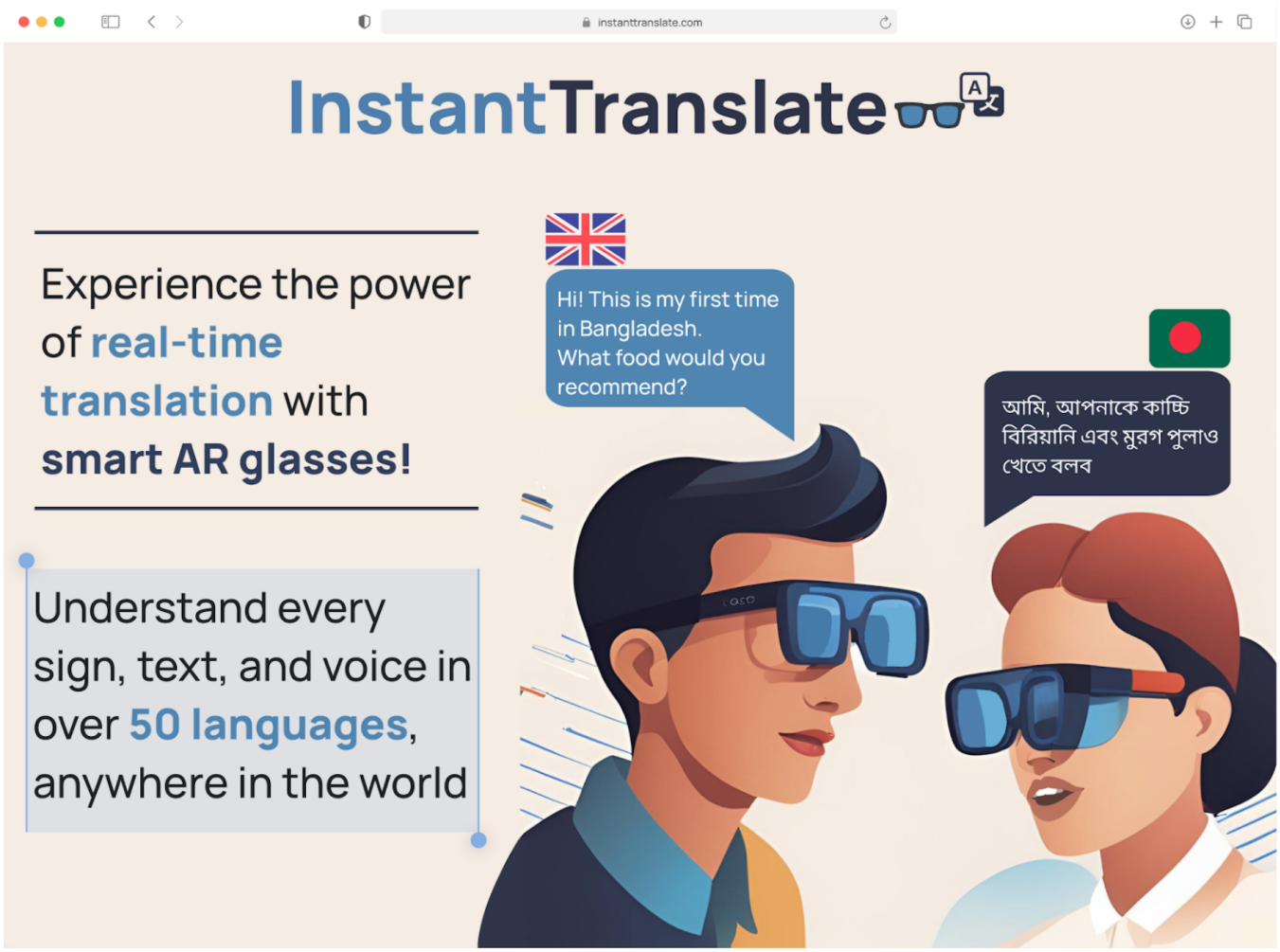

Figure 3: Landing Page

B LEAN BUSINESS MODEL CANVAS (FINAL)

PRODUCT | MARKET

1. Problem

The user can't understand foreign language in written text when for instance traveling.

The user can't engage in a dialogue with a foreigner speaking a different language.

4. Solutions

In real time, when commanded the AR-glasses will translate the text in front of the user.

When activated, the AR-glasses will listen to the dialogue and give a translation in the language chosen by the user.

The AR-glasses can be used in different environments and focus on the person who you are looking at.

5. Key Metric

List the numbers that tell you how your business is doing

3. Unique Value Prop.

Easy user friendly interface which translates in real-time voice and written translation.

Easy for the user to translate both texts and dialogues, with the smart AR-glasses.

Empowers users to communicate effortlessly in any environment

9. Unfair advantage

What do you have that gives you an unfair advantage?

Something that cannot be copied or bought.

6. Channels

What free and paid channels to your customer are there?

2. Customer segments

Persona: Lucy Mendez, 21, from Colombia, is studying in Denmark. With limited English and no Danish, she struggles to follow lectures, do daily tasks, and connect with local students. With Instant translation with the new AR-glasses, She can real-time translation solution to fully integrate academically and socially.

7. Cost structure

List out your fixed and variable costs.

8. Revenue streams

Identify your revenue model, and calculate back-of-the-envelope numbers for lifetime value, gross margin, breakeven point, etc.

Figure 4: Lean Canvas

C USER STORY MAP (FINAL)

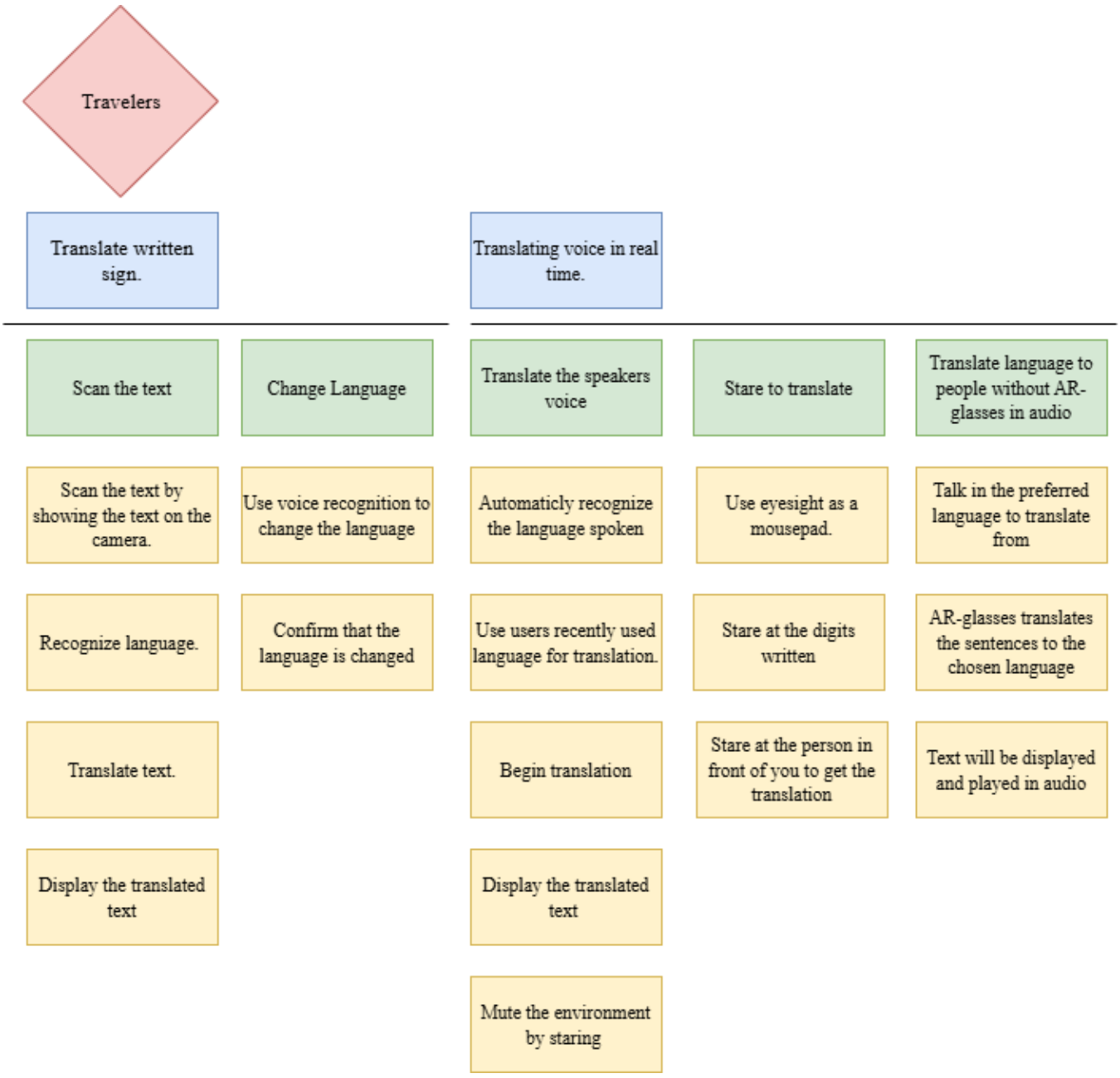

Figure 5: User story Map

D WIREFRAME (FINAL)

Figure 6: Final Storyboard

E VALIDATIONS

Validation documentation for iteration # 1

For Iteration 1, a combination of focus group feedback and heuristic evaluation was applied. The presentation of the landing page to peers and teaching assistants served as a focus group, where immediate feedback on usability, clarity, and market potential was collected. Constructive feedback highlighted interface improvements and communication clarity, while positive feedback validated the product's core concept.

Some positive feedback we received from peers and teaching assistants after the presentation of the landing page was the following:

- The landing page is simple and gives a clear message.
- The options for different languages instead of just a few international languages are great for including more users.
- The function of both text and audio translation is comprehensive for a minimum viable product.

The constructive feedback or problems with the idea and landing page were the following:

- What about if the system in the AR glasses has to translate dialogue in a noisy environment while many people are speaking at the same time?
- The landing page has these different problems:
  - Its not clear that the two people are talking to each other.
  - Its not clear if the message emphasizes our product efficiently.
  - Some flags could be added to visualize the different nationalities of the two people talking to each other.

Based on the feedback, we made a few changes to our idea and the landing page according to the constructive feedback we received. The following are the **solutions** we implemented so far:

- We focus on one-way dialogues by staring at the person in front of you, and in future phases, we would be focused on handling the noise in crowded places.
- We adjusted the landing page by editing the AI-made image where people are looking at each other and talking with open mouths.
- We removed the text nearly every language and added 50+ languages instead to not make it sound like we excluded a few languages, but we included many of them.
- We added flags to identify the language in the dialogue boxes.
- We highlight important text to specify AR glasses and real-time translation on the landing page.
- We added a QR code at the bottom of the landing page to let possible users know that they can read more about the VR glasses on the website.

Validation documentation for iteration # 2

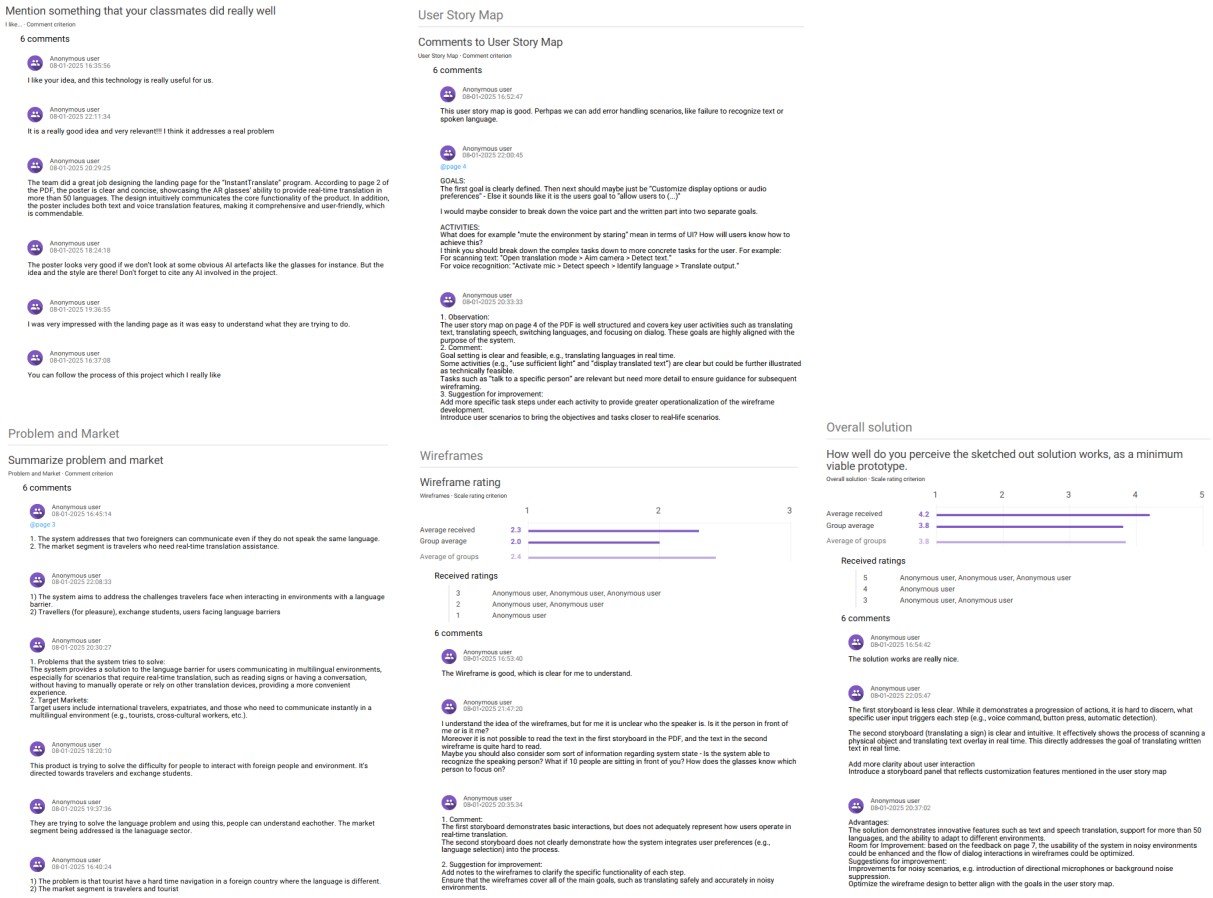

Figure 7: Feedback from FeedbackFruit.

Validation documentation for iteration # 3

For Iteration 3, a **think-aloud cognitive** walkthrough was implemented with two external users unfamiliar with the course. This method simulated real user behavior by guiding participants through critical tasks (language switching, text translation, and settings interaction) while verbalizing their thoughts. This approach identified usability issues, such as the need for clearer instructions and improved visual feedback during translation tasks. This iterative process allowed us to refine the user interface, enhancing intuitiveness and functionality.

Switch Languages and Engage in a Dialogue. The user must use the language switching feature, select the Danish language, and then engage in a short dialogue with the person in front, where the system will translate the conversation. The task tests the systems ability to switch between different languages and support a dialogue in the selected language. This is important to validate whether AR glasses can be used to facilitate communication between people who do not speak the same language.

Translate the Written Text on the Menu. The user focuses the AR glasses on the menu and reads it to detect the text. The system must recognize the text on the menu and display a translation on top of it. The task evaluates the ability of the product to recognize and translate text from items, such as menus in restaurants, which is essential for travelers.

Interact with the Settings. The user is asked to interact with the system settings using gesture-based input. The task involves finding the settings menu and navigating through different settings to evaluate if specific preferences are met. The task tests the usability and intuitiveness of the system user interface, especially the ability to navigate and customize settings.

Switch Languages and Engage in a Dialogue. Both users were able to easily find the language-switching feature, but one had difficulty reading the translated text. Both users suggested some kind of instruction before the system starts translating the dialogue. One user stated:

> *Maybe instead of letting the translation process start right away, it could be a good idea to add instructions at the start, just so I knew the following process.*

One user commented on the potential for use in international contexts, especially in countries using different characters than English.

Translate the Written Text on the Menu. The text recognition worked effectively for the menu, but one user expected the same handwriting from the menu to appear on the translated text. Users were impressed that handwritten text could be translated since they had different experiences from other apps that usually struggled with this. Feedback on the user interface pointed to a need for better visual feedback during scanning to confirm the text has been detected and is being translated. One of the users suggested:

> *I would like to see a marked indication of the text to make sure every word is detected before it starts to translate.*

Interact with the Settings. Both users initially expected to use eye-tracking movements to find the settings. One of the users tried to figure out where to find the battery percentage since it was not included in previous wireframes.

F LINK PROTOTYPE

Here the link of the executable prototype