Final Project Conceptualization and Evaluation Plan

Title : **SenseAI: Immersive Audio Descriptions for people who are blind or have low vision.**

Problem Statement:

The community of blind people often faces challenges in perceiving the richness, detail, and emotional essence of the world around them. Existing adaptability tool like screen readers or object recognition tools provide basic, utilitarian descriptions that lack depth and immersion. These descriptions focus mainly on function and identification but do not capture the nuances, beauty, and emotional resonance of the visual experience. As a result,blind individuals are often deprived of the full experience of the world—the ability to feel the grandeur of a sunset, the intricate patterns of artwork, or the emotive expressions of human faces in an evocative way.

This gap in sensory experience significantly affects an individual's quality of life, influencing not only their emotional health but also their ability to engage socially. The absence of visual cues makes it challenging for blind people to navigate social interactions, often leading to uncomfortable or awkward encounters. These experiences can foster feelings of stigma, shame, and social isolation, as blind or blind individuals may struggle with knowing how to connect effectively with sighted people.[1]

Research by Martins et al. (2019) [2] demonstrates that sensory disabilities, including visual impairment, can significantly impede social perception abilities. This finding underscores the notion that blind individuals may overlook subtle emotional cues in their surroundings, which restricts their social engagement. It highlights the necessity for developing advanced technologies and adaptive systems aimed at enhancing their capacity to interact with and relate to their environment in a more psychologically and emotionally rich manner.

Furthermore, a study by Chu and Chan (2022) [3] examines the heightened levels of loneliness experienced by blind individuals compared to their sighted peers. They find that this loneliness is often linked to difficulties in social interactions and limited community participation. This research emphasizes the need for solutions that go beyond functional assistance, addressing the psychological and emotional aspects of the blind.

Biases

Possible biases that the system might have:

Social Context:

1. **Cultural Bias:** Given that the system is designed to capture more nuanced descriptions, interpretations of these descriptions may vary significantly across cultures. For instance, take the color white: in many Western cultures, white symbolizes purity, and peace, and is often associated with weddings, embodying innocence and new beginnings. Conversely, in various East Asian cultures, white traditionally represents mourning and funerals, symbolizing death and loss. If the system is prompted to describe the color white, it may unintentionally produce suggestions that feel uncomfortable or inappropriate for users from cultures where white connotes sorrow or mourning. This bias arises from the system's assumption of a universal interpretation of color, overlooking cultural variations in meaning.

   Mitigation :
   a. Inclusivity in fine-tuning dataset
   b. Transparency about the statistics of the dataset so that users are aware of what to expect.

2. **Interpretation Bias**: The system's interpretation of information is inherently based on sighted perspectives, which may not always align with the experiences or needs of blind users. This could lead to descriptions that prioritize visual elements that are less relevant or meaningful to blind individuals. For example: if the system describes an object as "about the size of a basketball" or "the color of grass," it assumes that users can relate to these visual references, which might not be meaningful or relevant. This bias occurs when the system interprets and describes experiences through visual or metaphorical references that may not align with or be helpful for a blind user's experience of the world.

   Mitigation:
   a. We fine-tune the AI model using datasets that emphasize sensory experiences beyond visual elements, incorporating audio, tactile, and olfactory descriptions.
   b. Experiment with different large language models (LLMs) that tailor descriptions to the user's context, emphasizing relevant sensory information instead of visual cues. This approach simplifies the fine-tuning process.
   c. Have blind individuals as stakeholders.

Technological aspect:

1. **Data Representation Bias:** AI models can exhibit biases if they are trained on datasets that do not sufficiently capture the varied experiences of visually impaired individuals. When utilizing a pre-trained AI model and subsequently fine-tuning it for specific applications, the biases embedded in the original dataset may continue to influence the model's performance.
   Example: If a language model is trained on a dataset lacking diverse visually impaired experiences, it may still produce biased outputs after fine-tuning. For instance, it might describe a

visually impaired person solely as “struggling to navigate,” ignoring their interests or achievements, like a passion for music.

Mitigation:
Creating a well-balanced fine-tuning dataset in collaboration with blind individuals can effectively reduce bias propagation. As highlighted in the Towards Fairness article[4], conducting regular audits and involving blind individuals as stakeholders are essential strategies for managing and controlling these biases.

Harms

HARM TO PEOPLE (Blind Individuals)

1. Uncertainty and Unpredictability Risks:
    - Providing inaccurate or misleading descriptions that could lead to physical harm (e.g., misidentifying hazardous objects)
    - Creating confusion through inconsistent or contextually inappropriate descriptions
    - Causing emotional distress through poorly chosen descriptive language

HARM TO ORGANIZATION (developers and stakeholders):

1. Reputational Risk:
    - If the system fails to deliver accurate or appropriate descriptions, it may lead to negative user experiences, harming the organization’s reputation. Misleading descriptions or culturally insensitive content can result in public backlash or accusations of neglecting inclusivity in the product design.
    - Impact: This could reduce public trust, damage relationships with advocacy groups, and deter potential users, particularly within the visually impaired community.
    - Mitigation: Establishing an ongoing feedback loop with users and stakeholders to regularly update and improve the system based on real-world use cases and culturally sensitive data.

HARM TO ECOSYSTEM (Blind individuals community)

1. Over-reliance on AI for Sensory Perception:
    - Issue: Blind users may become overly reliant on AI-generated descriptions, potentially reducing their motivation or ability to develop alternative sensory perception skills.
    - Impact: This could affect users’ adaptive strategies and hinder their ability to navigate real-world scenarios independently, where technology may not always be available.

- Mitigation: Design the system as an assistive tool that complements rather than replaces users' sensory skills, encouraging users to develop personal strategies alongside AI assistance.

Feasibility

1. Timeline Considerations

- **Research and Development Phase** :(Already done) This includes gathering background on immersive audio technology, exploring dataset options, and defining initial system requirements.
- **Data Collection and Processing Phase and Model Development and Fine-Tuning** (~ 1 week):
  - Use open-source web text descriptions to query and generate a text-based description of the term using an RAG LLM system. Data can be collected through web scraping of websites like Wikipedia.
- **Evaluation and User Testing Phase** (~ 1 week): Usability testing
- **Documentation and Final Presentation** (~1 week): This final phase involves compiling findings, documenting development, and preparing project deliverables for assessment.

2. Required Resources

- **Hardware**: Standard hardware setups (e.g., laptops with good processing power or access discovery for AI model training) will suffice.
- **Software**: Necessary software includes Python (with libraries like PyTorch or TensorFlow for model development) and libraries for natural language processing (like Hugging Face's Transformers or ChatGPT). Additionally, using audio synthesis tools such as Text-to-Speech (TTS) systems (e.g., Google's TTS API or Microsoft Azure's Cognitive Services) will help convert textual descriptions into audio.
- **Data**: Obtaining an inclusive dataset is key. We will rely on open-access sources for general text data and collaborate with visually impaired users to ensure cultural relevance and inclusivity. This will include fine-tuning data and accessibility-related adjustments.

3. Required Skills

- **Natural Language Processing**: Familiarity with language models, fine-tuning, and generating contextually rich descriptions.
- **Audio Synthesis**: Skills in integrating TTS technology to translate text outputs into immersive audio descriptions.
- **User-Centered Design - UI**: Familiarity with usability testing and accessibility standards for visually impaired users. Classes have helped us get familiar with building websites using React and having a user-centered perspective while designing.

Evaluation Plan:

Evaluation Objectives: The parameters that we will be evaluating the project on are usability, accessibility, quality of answers and accuracy of text.

Methodology:

1. As said by Mallin et al.[5] ‘To provide a consistent positive experience to people with disability through AT it is necessary to distinguish the user's needs from the objective of the experience requirements’, we would have interviews with vision-impaired individuals to assess their needs while we develop the application.
2. After an initial phase of development, conducting tests based on usability and accessibility with blind individuals using surveys and one quality of answers and accuracy of text using surveys with non-vision impaired users will be done.

Participants:

Diversity in usability research participants is very important to make sure that the resulting technologies will be fit for the purpose across different user groups.[6] We plan to get in touch with envisioning access through the guest lecturer we had for one of the classes- Mrs Diane. The number of participants would be dependent on the response from Mrs. Diane but we aim to have 50 participants from diverse age ranges - 15 to 65 and different races and backgrounds. We would also like for them to be belonging to different gender identities. As the scope of the project is to just generate text in English, we would want the participants to be able to understand and speak English. Similarly, for the non-vision impaired surveys, we would like to reach out to our friends, family, relatives and classmates to ensure that we get a diverse population with different age ranges, cultures and gender identities.

Evaluation Process:

As part of the initial survey, we would present the premise of the application to the participants and ask them if this is something that could help them. We will also ask them what features they think a tool such as this should include and any guardrails or helpful tools we should include the make the tool easier to use and more accessible.

As part of the final survey, we will ask the users to use the tool in front of us and once they are done using the tool, we will ask them what they thought about the experience, whether they felt uncertainty or confusion while using it, how did they like the answers and then ask them to rate the tool from 1-5 based on usability, accessibility, and quality of answers.

For the non-vision impaired participants, we will simply show them generated text over 10 different prompts and ask them to rate the accuracy of the text on a scale of 1-5.

Success Criteria:

The success criteria for our tool will be based on the average scores of the surveys on usability, accessibility, quality of answers, and accuracy of text. If all these values are over 4 rating on average, the project an be considered to be successful.

References:

[1] L. Romo and M. Taussig, "Study Uncovers How Blind and Visually Impaired Individuals Navigate Social Challenges," NC State News, Aug. 23, 2022. [Online]. Available: https://news.ncsu.edu/2022/08/blind-visually-impaired-challenges/

[2] A. T. Martins, L. Faísca, H. Vieira, and G. Gonçalves, "Emotional recognition and empathy both in deaf and blind adults," Journal of Deaf Studies and Deaf Education, vol. 24, no. 2, pp. 119–127, 2019. doi: 10.1093/deafed/eny046

[3] H. Y. Chu and H. S. Chan, "Loneliness and Social Support among the Middle-Aged and Elderly People with Visual Impairment," International Journal of Environmental Research and Public Health, vol. 19, no. 21, pp. 14600, Nov. 2022. doi: 10.3390/ijerph192114600

[4] A. Guo, E. Kamar, J. Wortman Vaughan, H. Wallach, and M. Ringel Morris, "Toward fairness in AI for people with disabilities SBG@a research roadmap," SIGACCESS Accessibility and Computing, vol. 125, Article 2, pp. 1, Oct. 2019. doi: 10.1145/3386296.3386298

[5] S. S. V. Mallin and H. G. de Carvalho, "Assistive Technology and User-Centered Design: Emotion as Element for Innovation," Procedia Manufacturing, vol. 3, pp. 5570–5578, 2015. doi: 10.1016/j.promfg.2015.07.738

[6] S. Rutter, E. Zamani, J. McKenna-Aspell, and Y. Wang, "Embedding equality, diversity and inclusion in usability testing: recommendations and a research agenda," International Journal of Human-Computer Studies, pp. 103278, 2024. doi: 10.1016/j.ijhcs.2024.103278

[7] A. Walczak and L. Fryer, "Creative description: The impact of audio description style on presence in visually impaired audiences," British Journal of Visual Impairment, vol. 35, no. 1, pp. 6–17, 2017. doi: 10.1177/0264619616661603

[8] J. Martínez and R. Pardo, "Subjectivity and creativity versus audio description guidelines," in Advances in Accessibility and Inclusive Practices, pp. 45–58, Springer, 2023. doi: 10.1007/978-3-031-60049-4_3

[9] G. Vidal and M. Suarez, "Enhancing audio description: Inclusive cinematic experiences through sound design," Journal of Audiovisual Translation, vol. 4, no. 2, pp. 75–92, 2021. [Online]. Available: https://pdfs.semanticscholar.org/d4ec/e4850b35f7328154803154dc38c32fde4b4a.pdf

[10] A. Hättich and M. Schweizer, "I hear what you see: Effects of audio description used in a cinema on immersion and enjoyment in blind and visually impaired people," British Journal of Visual Impairment, vol. 38, no. 3, pp. 284–298, 2020. doi: 10.1177/0264619620911429

[11] "Seeing AI: Free app narrating world around you," Paths to Literacy, 2023. [Online]. Available: https://www.pathstoliteracy.org/resource/seeing-ai-free-app-narrating-world-around-you/

[12] "Audible Sight: A new method for creating audio descriptions," Perkins School for the Blind, 2023. [Online]. Available: https://www.perkins.org/resource/audible-sight-a-new-method-for-creating-audio-descriptions/

[13] "WorldScribe: Towards context-aware live visual descriptions," arXiv preprint, 2024. doi: 10.48550/arXiv.2408.06627

[14] "AI-powered software narrates surroundings for visually impaired in real time," Tech Xplore, 2024. [Online]. Available: https://techxplore.com/news/2024-10-ai-powered-software-narrates-visually.html

[15] "Audio description: Enhancing viewing for the visually impaired," Vanan Services Blog, 2022. [Online]. Available: https://vananservices.com/blog/the-power-of-audio-description-elevating-the-viewing-experience-for-the-visually-impaired/

[16] "Image Captioning with Transformers: Transforming visual content into audio for the visually impaired," GitHub Repository, 2023. [Online]. Available: https://github.com/sathishprasad/Image-Captioning-with-Transformers--Transforming-Visual-Content-into-Audio-for-the-Visually-Impaired

Language in this document is used in reference to : Be My Eyes “Inclusive Language” Guide