**1. Ethical Business Plan**

**A. Company Name**

PersonaAI

**B. Long Term Vision**

With the AI boom of 2022 following the novelty of ChatGPT, there has since been many tech companies which have incorporated these LLMs into their own products and services such as Google Gemini, Microsoft Copilot, Opera Aria, etc. Despite the number of AI companions that have come out, they mostly behave in homogenous and boring ways with little variation in their personalities. With this, we ask ourselves, how can we make AI more unique? How can we make AI more useful? How can we make AI more fun? Our goal is to give people the platform to create, share, and engage with their own unique AI personalities. We want to give them more memorable experiences with AI. We want people to make the most out of AI.

**C. Strategy with Ethical Impacts AND Ethical Safeguards**

Our company provides a personalized AI platform where users can create their own AI with desired personalities and traits, share, and interact with them. We want our service to be readily available and accessible to all, so our company will provide the necessary infrastructure to run the AI models server side rather than requiring hardware acceleration from the client side to run AI locally. Stakeholders in the business will be the individuals in the company itself, users, and potential investors.

**OKR’s**

(Bry):

1: ONE Objective and Key Result:

Conducting usability testing and implementing feedback is one of PersonAI's OKRs. We serve a diverse clientele, and our AI platform does not discriminate based on a person's income, gender, race, academic background, or language. We designed the platform to be accessible to all users. AI allows them to experiment with their creativity. Whether it's through surveys we send to our clients or if they have a persistent problem with our AI platform, we use customer input to improve the user interface and optimize the AI chatbot's performance. We can determine what has to be improved on our end to make it nearly effortless for our clients to utilize PersonAI to explore their creativity with AI by using user feedback. Increased client involvement and product viability are advantageous to investors and/or business partners. In addition to helping individuals use PersonAI to explore their creativity, they can also see what PersonAI is all about and raise any problems they believe need to be fixed. Whether it's a persistent problem they have or if they have suggestions for how we may improve PersonAI. At PersonAI, we make sure that we adhere to privacy and digital accessibility laws. The people at PersonAI who are in charge of it would be the only ones with access to any stored personal data. Unless required by law, we will not divulge any of our customers' information.

2: Metric(s) with Experimentation:

We truly reached out to our customers using our metrics and experimentation. We employ consumer surveys and A/B testing. We can ask our staff to take part in A/B testing. At least 200 of our staff will be divided into two groups. Group 2 will use our enhanced user interface (UI) version of our AI product, while Group 1 will use our current AI version. We then assign tasks like registering and making a profile. They will be asked to begin by developing their own AI chatbot and documenting what functions well and what requires improvement. We will next compare and contrast the participant remarks as a team to determine what needs to be improved or what they believe should be added to the chatbot to enhance the experience. Finally, we have customer surveys, where everyone is asked:

1. How easy is it to navigate through our platform? (on a scale of 1-5, 5 being very easy)
2. How intuitive was the design and layout?(on a scale of 1-5, 5 being very intuitive)
3. Did you encounter any difficulties while using the platform? (Yes/No, optional text box)
4. How quickly did the AI chatbot perform the task you wanted(scale of 1-5, 5 being super fast)
5. How satisfied are you with the overall speed and performance of PersonAI (scale of 1-5, 5 being very satisfied)
6. How satisfied are you with the overall experience? (Scale: 1-5, 5 being very satisfied)
7. On a scale of 1-5 how likely are you to recommend our platform to others? (with 5 being very likely)
8. What feature(s) do you find the most useful? (text box)
9. Is there anything we can do to improve your experience? (text box)
10. Did PersonAI help you explore your creativity with the help of AI? (scale of 1-5, 5 being very much)

We gather as a team to compare and contrast each respondent's responses to questions 1 through 9 on the customer surveys. Since questions 3, 4, 5, 8, and 9 are primarily concerned with what users are saying about our platform and what we can do to make it better, we pay close attention to them. From there, we can determine what we should actually concentrate on using the findings from the consumer surveys and the A/B testing.

3: Ethical Impact(s)/Issues(s):

| **Stakeholder** | **Financial** | **Privacy** | **Conflicting Interest** | **Violation of Rights** |
| --- | --- | --- | --- | --- |
| Customers | low | High | low | high |
| Product development team | low | mid | low | none |
| Customer Support Team | low | low | low | none |
| Investors & Business Partners | high | low | mid | low |
| Universities and Educational Orgs | low | low | low | low |
| Government and Policy Stakeholders | low | high | low | mid |

Customers: Assuming that payment information is not misused, there is minimal direct financial risk to customers. Analytics and usability testing may be utilized to improve the product using customer data. There is a chance of privacy infringement if their data is not appropriately anonymized or shared. The enhancement of our product is in line with the interests of the client. A breach of a customer's rights may result from a breach of their data privacy or from a failure to prioritize accessibility.

Product Development Team: The development procedures do not immediately pose a substantial financial risk. To guarantee that user data is maintained safe, the group should adhere to procedures. Their primary goal is to enhance the product without sacrificing other considerations. No rights are infringed upon as long as the team adheres to the law and moral principles.

Customer Support Team: helping users has no financial impact. While addressing user concerns, they ought to preserve consumers' privacy. No rights are infringed as long as they adhere to the company's policies.

Business partners and investors: improper handling of user data or unethical behavior may result in fines or reputational damage that reduces investor profits. They don't have a direct hand in handling client data. Profit may be their first priority, which might put pressure on them to compromise ethics, particularly when it comes to data consumption. Investors are at minimal risk since they don't directly violate user rights, but they might still be subject to legal action if the firm is shown to be in violation.

Universities and Educational Orgs.: Unless their partnerships result in extra liabilities, universities and educational organizations do not face a financial risk. Although they aren't in charge of it directly, they must be conscious of how user data is handled. They would be in line with ethical norms and place a greater emphasis on the wellbeing of students. If they suggest things that violate privacy regulations or accessibility requirements, they may be in danger.

Government and Policy Stakeholders: They don't have any financial risk, but if they don't enforce moral principles, the public may criticize them. If data privacy regulations are broken, they could be affected. Usually, they seek to safeguard user rights and the public interest. If they don't control or deal with user rights abuses, they could be complicit.

With these ethical impact(s)/issue(s) brought up here we can look at a court case where a lady named Mary Louis was denied tenancy to an apartment in 2021. In this court case we have Mary Louis and Monica Douglas v. SafeRent solutions LLC and Metropolitan Management Group, LLC. in this court case Louis and Douglas had filed a complaint alleging claims under the FHA and Massachusetts general laws. “Louis and Douglas, both Black women, are rental applicants who use federally funded housing vouchers to pay part of their rent. They allege they were denied housing due to their SafeRent scores, and algorithm based ‘lease performance risk score’” pg 2 crt doc. Based on the court case these ladies were denied their tenancy based on an algorithm score that was given not just to them but also to black and hispanic people. In the case its noted that someone's SafeRent scores are marketed to housing providers so that they are able to select “better renters”. Here we have the use of an AI algorithm being used on their customers to be biased and having conflicting interests. We also see that financial issues also played a factor in this. Louis and Douglas use these vouchers to help pay their rent which is stated in the case. The case ended up closing with a federal judge approving a settlement and the company behind the algorithm agreed to pay over $2.2 million and also look over parts of its screening products that were alleged to be discriminatory. With PersonAI we aren't looking at your race/ethnic origin or your income were focused solely on the creativity of are users and the safety of your data and privacy. We use our users data whether that be age or race/ethic background to help us be inclusive with demographics and respect the data and privacy of our users.

4: Ethical Safeguards:

We will utilize a demographically diverse sample for user testing in order to prevent biased data collecting and guarantee that product enhancements represent the experiences of a varied user base. To make sure the product is enhancing the experience for all users, this involves gathering input from a range of age groups, racial/ethnic origins, economic levels, and technological proficiencies. Who will be involved? To make this happen we will be using ethical and context experts. Ethical experts will help create an inclusive criteria for data sampling and context experts will work with data scientists and product managers to develop appropriate demographic sampling strategies. We can find these experts through networking or do freelance and see if they would be willing to join our team at PersonAI. How you would implement the safeguard(s)? We would use minimum demographic samples and demographic data tracking. For demographic samples we would establish thresholds for a minimum number of responses that are necessary from each demographic category. For example we could have 500 users from each racial/ethnic group and 500 users from each age group. For demographic data tracking we could track the demographic breakdown of participants to ensure that there is a diverse representation. Can you measure its effectiveness and if so how?

We would use data analysis to analyze feedback and to see if it is representative of the overall user population and whether improvements lead to better user satisfaction across all demographics. We would also use feedback bias metrics we could analyze data from various user groups on a regular basis to look for feedback bias like comparing and contrasting the satisfaction levels of users across the various age groups.

5: References:

J. Buolamwini, “Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products – MIT Media Lab,” *MIT Media Lab*, 2019. <https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/>

(For this reference I was able to download a pdf version of it through the website.)

[1]*MARY LOUIS and MONICA DOUGLAS, on behalfofthemselves and similarly situated persons, and COMMUNITY ACTION AGENCY OF SOMERVILLE, INC, v. SAFERENT SOLUTIONS, LLC, and METROPOLITAN MANAGEMENT GROUP, LLC*, vol. Case No. 22cvI 0800-AK. Available: <https://www.justice.gov/crt/case-document/file/1562776/dl>

Sairish: Integrating the voice customization with emotion recognition features is a key aspect of enhancing PersonAI, ensuring a seamless and engaging user experience. This integration will allow users to interact with the AI using realistic and personalized voices, creating a natural and immersive interaction. Additionally, the AI will adapt its responses to align with the user’s emotional tone, ensuring that each interaction feels authentic and tailored to the user’s mood. By enabling emotional context to analyze speech patterns, our platform can create more human-like interactions, allowing the AI to understand the user’s emotional state and generate responses accordingly.

2: Metric(s) with Experimentation

To measure the success of the OKR, there are 4 metrics that aligns with the objective of enhancing the platform through voice customization and emotional recognition. First, we well focus on the adoption rate which will measure the number of active users using the new features within the first few months of its release.This metric will be measured by tracking how many users use the new voice customization with emotion recognition features, and calculate the percentage of users who adopt these capabilities. Second, user engagement will also be tracked to better understand how frequently users interact with these features. It will be measured by monitoring their interaction with the voice and emotion features over a period of time, and displaying the data analytically. Third, user satisfaction will determine how positively users perceive the new features. We will conduct data through user feedback and survey forms to evaluate their overall satisfaction and impact on the user experience. Fourth, the overall performance will be measured to confirm that all features are functioning. This includes the response time, errors that occur, and stability of the system.

3: Ethical Impact(s)/Issue(s)

Scenario 1 - Privacy

There are several ethical issues that may arise including privacy and data protection. The use of personalizing the voices will involve collecting, processing, and analyzing user information which could lead to consequences. Similar issues arose in the Facebook-Cambridge Analytica scandal where Facebook’s data was misused without user consent and affected many people.

Scenario 2 - Bias

Another issue could be the bias in the AI’s algorithms. For instance, it might struggle to correctly interpret emotional signs from different demographic groups, resulting in inaccurate responses because if it’s focused on one specific group, it won’t be fair for other users.

Scenario 3 - Emotional Manipulation

Another issue is that the emotion recognition feature might influence users in ways that manipulate their feelings. For instance, AI could intentionally change its responses to match the user’s mood which might make them feel like they’re being controlled by their emotional state.

| Stakeholder | Financial Risk | Privacy Risk | Conflicting Interest Risk |
| --- | --- | --- | --- |
| Company | High | Low | Mid |
| Customer | Low | High | Mid |
| Developer | Low | Low | Low |

Analysis of Ethical Impact Risk

Company: The financial risk is high due to the potential damage to its reputation from the violations and its misuse of data. The privacy risk is low, assuming that all safeguards are implemented in the right way. However, the conflicting interest risk is in the mid-level, as the platform prioritizes user engagement while adhering to the principles of trust and consent.

Customer: The financial risk is low since users don’t risk losing any money from the features, although the risks of data if not used properly. The privacy risk is high because they’re unaware of how their data is processed or if it’s not stored securely. The conflicting interest risk is mid due to the AI’s responses being

Developer: The financial risk is low but it could likely increase based on the violations and misuse of technology. The risk is low for privacy because the code is reliable and data is handled at a good measure, but issues may grow if these measures are not implemented effectively before the feature is launched. The conflicting interest risk is low for developers because they aren’t involved with the business and marketing aspects of the company, but might be asked to develop new features that drive user engagement.

4: Ethical Safeguards

Ethical experts would be involved in helping design this safeguard because they are essential in ensuring that the platform adheres to guidelines and doesn’t take advantage of users emotionally. To implement safeguards for these features, there are several steps like consenting, ethical reviews, and bias checks. First, we will introduce a process where users will be informed about the data that is collected, how it's being used, and the length of time. Next, ethical reviews will be conducted by AI/ML developers, psychologists, and ethical experts to assess how the data is being used and ensure no manipulations take place. Additionally, to prevent unfairness, bias checks will make sure that the features don’t favor a specific group.

5: References

[1]

N. Al Mazrouei, “Emotion AI: Transforming Human-Machine Interaction,” *Trendsresearch.org*, Feb. 17, 2025. <https://trendsresearch.org/insight/emotion-ai-transforming-human-machine-interaction/?srsltid=AfmBOopqbJw8eWwMdBROmhifSiHxexDPLq7XluXcCxr1gTb3aWIwm6av> (accessed Mar. 12, 2025).

[2]

I. ur Rehman, “Facebook-Cambridge Analytica data harvesting: What you need to know,” University of Nebraska - Lincoln, 2019. Available: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=5833&context=libphilprac>

Richard: PersonAI will increase the usability of AI chat interactions by 200% (double). PersonAI is a platform for users to create their own personal AI. We want to give users the best experience possible and that means making sure our AI model adheres to the social expectations of our users just as if they were speaking with a real character or person. As such, it is paramount that the AI infrastructure which PersonAI operates under can keep up with user requests as they user our platform. Key primary stakeholders would include our company as we are providing cloud compute infrastructure for our AI, our users who depend on our infrastructure, and investors/business partners who have an interest in the success of PersonAI. In order to improve the usability of our AI, we will need it to retain more context memory and have reduced response time. We can improve context memory by implementing Long Short-Term (LSTM) Networks alongside an improved Attention Mechanism. LSTMs have memory cells that can maintain information over long periods which will allow them to better remember and understand important context during conversations [2]. Attention mechanisms will enhance generated responses by allowing the LLM to dynamically focus on different parts of user input and weigh each token differently on their level of importance while generating an appropriate response [2]. As for response time, we can simply upgrade our AI infrastructure by buying and adding more GPU compute power [1]; the more GPUs, the more of said GPUs we can allocate to handle each user request.

2: Metric(s) with Experimentation:

To measure the improvement in usability in PersonAI LLM, we first must collect data on certain aspects of LLM behavior which each have their own metrics. First is to test the accuracy of the new model with LSTM where an A/B test experiment will be conducted. This is the performance test of the LSTM using a control model (without attention) and test model (with attention) where each model will be tested on whether they can give appropriate responses after a certain duration of time in a conversation; they will be scored out of a number of total responses. For example, telling each model a series of context information, then asking it to recall a piece of information during the conversation to see if each model can recall information and generating relevant responses. If a model fails to generate a relevant response, it will receive a point deduction; later the results will be converted into a percentage.

The next aspect to measure is our AI API response time between the client sending input, and server side processing and sending back output. This is easy enough as we can simply log and measure the average time taken for user queries to receive generated responses before infrastructure changes, then compare them with newly upgraded infrastructure to see the difference in response time.

To relate these two different metrics of accuracy and response time into usability, we first need to consider that the two metrics generated, response time and percent accuracy, are two different units in different scales. We’ll need to normalize the metrics so one metric doesn’t disproportionately affect the usability score too much:

*NormalizedResponseTime = (ResponseTime – MinimumResponseTime)/(MaximumResponseTime – MinimumResponseTime)*

Since accuracy is already a percentage, we can simply divide by 100 to get a normalized accuracy metric *Accuracy / 100 = Normalized Accuracy*. After all that, we can throw the new normalized numbers for response time and accuracy into a weighted sum formula such as:

*Usability = (Weight1 \* ResponseTime) + (Weight2 \* Accuracy)*

The weights are for here to determine the importance of each metric to the overall usability.

3: Ethical Impact(s)/Issue(s)

With the improvements to the usability of our AI, we can expect that users will enjoy the improved believability or immersion of using their favorite personalized AI that hopefully match their expectations. However, it is important to note that in continuously pursuing better AI that behaves or responds more believably, there will be those at risk of being deluded by AI character constructs; particularly children as they are more impressionable at a young age and have difficulty distinguishing reality from fantasy. One notable case was 14 year old Sewell Setzer III who made news headlines for committing suicide with the apparent involvement of a sexualized AI bot which was created by users on Character.AI in likeness of Game of Thrones character Daenerys Targaryen [3]. Below is an expected ethical impact risks table for such a issue when this OKR is implemented.

A screenshot of a computer screen

AI-generated content may be incorrect.

Analysis of Ethical Impact Risk:

Company/Users: There is generally no financial risk for users as PersonAI does will not take money from users directly, although there is great financial risk to the company as their operations will depend on the success of the platform and how many people use it. Users may be at high privacy risk for having PersonAI computing done on the cloud instead of locally on their machines.

Company/Investors/Business partners: Investors/business partners may have an interests in the wellbeing of the company whether that is because of equity financing or for interest in wanting to keep a business partner (think hardware vendors working with company, researchers, etc).

Users/Investors/Business Partners: There is generally little overlap between these two as they do not directly interact with each other.

4: Ethical Safeguards.

In order to better protect users who may be at risk of mental health problems or from self-harm, we can implement a system to identify certain behaviors and prohibit them from use. For example, if a user creates an AI personality that likely to engage in inappropriate behavior that entices users to self-harm or suicide, we can prevent this from occurring.

References

[1] C. Bronsdon, “Understanding latency in AI: What it is and how it works,” Galileo AI, https://www.galileo.ai/blog/understanding-latency-in-ai-what-it-is-and-how-it-works (accessed Mar. 11, 2025).

[2] M. Chauhan, “A simple overview of RNN, LSTM and attention mechanism | by Manu Chauhan | the startup | medium,” Medium, https://medium.com/swlh/a-simple-overview-of-rnn-lstm-and-attention-mechanism-9e844763d07b (accessed Mar. 12, 2025).

[3] K. Lang, “TikTok algorithm guide 2025: How to get your videos on fyps,” Buffer, https://buffer.com/resources/tiktok-algorithm/ (accessed Mar. 3, 2025).