**Data and Technology Risk Assessment**

| **Risk Area** | **Risk Definition** - *The project will/could fail if:* | **Proposed Mitigation Strategy** |
| --- | --- | --- |
| Data sources | Listed Below are our data sources:  Handheld Scanner   * Data of different body types - to guess the waist size, bust size, and hips size. An example would be the visual-body-to-BMI dataset. * Data to get the other measurements such as the neck, sleeve size, etc   Clothing Recommender   * The particular User’s purchase and viewing history of fashion apps like Amazon and SHEIN * All user’s viewing and purchasing trend data - to be used to gauge patterns within regions, countries and economic backgrounds * Skin Color Compatibility dataset which helps suggest the best-colored clothes that match well with different skin colors * The best solution for our needs would be the ModaNet dataset. Based on eBay's Paperdoll dataset], ModaNet is a sizable new fashion dataset. * It contains 55,176 high-quality, completely annotated photos of street fashion, making it the largest fashion dataset available for semantic segmentation and object recognition.   The datasets related to the user's behavior in fashion apps are mostly not public.  In the publicly available data, the split between different races, genders, etc is not clear in many cases | * Work with multinational companies like amazon and market the usefulness of our recommendation system and the sales improvement they can achieve if they are willing to share their data with us. * Then we need to make a data exchange deal with them * Conduct better surveys and collect more features from existing datasets |
| Privacy concerns | * Data related to physical appearance is generally very sensitive and personal * Features like weight, height, dimentions etc can be misused by other agencies for many illegal and sketchy activities | * Explicit consent must be taken from the user * The default option for the user must be ‘We do not want to share data’ * Only if the user agrees to share should data be collected * User should also be given the option to chose if they should get recommendations based on generalisations using data from other people * We will have a strong legal team consisting of privacy experts |
| Data format | * Data should be be unbiased, securely collected * Should be of reasonable price if it is bought * Most of the features we need must be available lik ethe weight, various sizes | * Before starting on learning the data, it should undergo proper sanity tests * An expert level data pre processing and cleaning should take place * If any necessary features are missing, we should think about changing the dataset |
| Data analysis | * For each feature in the dataset, proper correlation graphs should be plotted by the data analyst to guage significance * Professional analysis must be done to judge the quality of the clothing recommendation predictions * The accuracy of the hand held scanner’s output must be studied in detail * The fit quality of th AR Mirror must also be analysed | * We will ensure that wwe hire expert analysts * There would be multiple points during the project where an external analyst will drop in to provide insights and corrections from a different perspective |
| Insights | * We expect all the three products to have hight explainability * If a user asks why he was recommended a product, we should be able to say answers like “this matches your skin tone perfectly + is a perfect fit for your waist size” | * For high explainability we might have to compromise on performance * The balance will be key |
| Model replication | * We will be updating our data on a regular basis * Every other week a new fashion app come up, so if we want to sell our product to thwm we should be able to utilise their new data and replicate what we have already done on existing data * Retraining will be a costly exercise if the feature set is different | * Whenerver new data come in, we should re-run the models * We should set up a completely independent pipeline for testing out new datasets - as there might be changes in the list of features * The model we make will be one with high replicability |
| Data infrastructure | * Since we are dealing with multiple data sources with entirely different types of features (text, image, numbers, etc) without proper organization it will be near to impossible to manage them * We also have three separate but connected products and we will probably have to exchange data and talk between the products | * A well-organized hierarchy and infrastructure should be set up for each product before starting the projects * Regular checks should be set up up to review the data infrastructure |
| Partnerships and synergies | * Since we have three different but interconnected products as part of the entire project, we are likely to get into various blocks in the pipeline * A blocker in one product will adversely effect the other two products also * Collaboration between different departments will get complicated in one department works on say all three products | * We should establish good partnerships with the fashion shopping apps and maintain good relations with them. This is crucial for marketing and getting relevant data * A very fluid company hierarchy will be set up where the three different products will be known to everyone in every department * They will know if doing one change in the handheld devices would create an issue in the recommendation software, which would in effect ruin the AR mirror experience   We will engage in partnerships with   | Fashion Apps | For getting data and selling product collaborations | | --- | --- | | Marketing consultants | To know which apps to create deals with | | AI ethicists | To make sure everything we do stays ethical | | Law experts | Privacy policies | |

Credit: NEXT Canada