- **Mr Smith** is a junior government official whose job entails reviewing documents to identify sensitive information that should not be available to the public. He believes he would be able to perform his job more efficiently if he was provided with a tool that could automatically identify names, phone numbers and addresses in the documents he was reviewing. He has basic computer skills and is looking for a web-based tool that doesn't require excessive training to understand.
- **Mrs Johnson** is a senior government official in charge of coordinating a team of document sensitivity reviewers. She is under pressure from her senior supervisors to reduce staff budget spending on his department, however her team of staff is struggling to keep up with the current workload of document reviews. She recognises that current staff must complete document reviews quicker to keep up with the workload. She believes her staff would complete document reviews more efficiently if they had software available to them that could automatically make an initial judgement as to whether a document should be deemed sensitive or not.
- **Mr O'Shea** is a member of parliament who is currently under investigation for numerous disciplinary issues. He accepts the consequences of his actions however he is concerned that the government issued documents related to his offences will contain sensitive, personal information that is not currently available to the public. He is worried that the reviewers assigned documents relating to him may not correctly identify them as sensitive. He believes they should be assisted by an application that can automatically recognise personal information, rather than having each document manually reviewed.

My view for the project is a web application designed to assist sensitivity reviewers in detecting and flagging sensitive documents. The homepage of the web app will feature background context for the motivation of the project and the desired use, as well as in depth instructions as to how to efficiently use the provided tools. Users will be able to upload a corpus of documents, with each document being individually analysed using named entity recognition tools. Each entity detected in a document will have its 'entity abstract' – a brief description of itself – scraped from a corresponding website and displayed available for the user to hover over and view. The process of entity recognition and web scraping may be computationally expensive so I will develop a database for the storage of entity abstracts, entity instances and document texts. This will allow the web app to perform one off analysis of each document when it is uploaded and refer back to stored data for future references.

Once documents are uploaded and processed, users will be able to navigate to document or corpus analytic pages. The document analytics page will provide in depth analysis such as where entities occur in the document, what types of entities occur the most frequently and filtering by types of entity. The corpus analytics page will provide in depth analysis on the set of documents uploaded as a whole – Which entities occur most frequently in a document for a given entity and analysis of conditional probability of which entities are most likely to appear in a document, given it is deemed sensitive and vice versa.

https://www.figma.com/file/HhZTX2XQglS0nDlXdQi19G/L4-Project-Wireframes?node-id=0%3A1