**Proof of Concept Scope and Future Considerations**

**Sheffield FOI Insight Tool**

**Team Members**

**FOI AI Team**

|  |
| --- |
| Abokor Ahmed |
| Lasse |
| Peter Karakostas |
| Richard |

**Proof of Concept Declaration**

This project is delivered as a Proof of Concept (PoC).

A PoC is an early demonstration designed to test the feasibility and value of this proposed solution. It is not intended to be a production-grade system but instead serves to:

* Validate the technical approach.
* Showcase key functionality.
* Identify opportunities and limitations.
* Inform future development decisions.
* By building this prototype, we demonstrate that it is technically feasible to:
* Scrape Freedom of Information (FOI) request data from public sources.
* Apply Natural Language Processing (NLP) to extract structured insights.
* Present summarised data through a searchable, interactive web interface.

**Current Scope and Limitations**

The current implementation operates under the following constraints:

**Data Sample Size:**

While approximately 500 FOI requests have been successfully scraped, only 25 requests have been processed through Azure OpenAI for metadata extraction and summarisation.

This limited scope is due to time, cost, and processing constraints in the PoC phase.

**Manual Interventions:**

The current pipeline requires some manual validation and handling of the AI-processed data before web publishing.

**Model Configuration:**

The AI prompts and extraction methods are in early-stage configuration.

Further research is needed to optimise accuracy, reliability, and consistency of the NLP outputs.

**Future Considerations for Robust Implementation**

**Model Refinement and Validation**

* Conduct further training, testing, and refinement of the NLP models to ensure they produce meaningful, reliable results at scale.
* Collaborate with subject matter experts to validate topic categorisation and summarisation accuracy.
* GDPR and Data Protection Compliance
* Review GDPR requirements to ensure:
* Personal data (names) in FOI requests is handled lawfully and ethically.
* Data subjects' privacy is respected in both processing and public presentation.
* Identify sensitive content (e.g., personal names, addresses) and implement redaction or filtering where appropriate.

**Ethical and Legal Considerations**

* Explore ethical implications of automated summarisation, accountability etc.
* Consider licensing and usage rights of data sourced from third-party platforms (e.g., WhatDoTheyKnow).
* System Scalability and Automation
* Expand the processing pipeline to support all scraped FOIs.
* Automate the end-to-end flow, reducing manual intervention.
* Evaluate infrastructure scalability for handling larger datasets or multiple councils.

**Conclusion**

This Proof of Concept has successfully demonstrated the technical viability of automating FOI data analysis and publishing. However, further research, model refinement, ethical review, and legal validation are essential before advancing to a production-ready system. The learnings from this PoC will inform the next phase of development if this project continuous.