

1 NLP/IR for e-discovery

In response to FOIA requests, legal teams in governments must find needle-in-a-haystack information in vast document sets under extreme time pressure. Optimize any part of this 'high-recall' pipeline, from processing to retrieval and active learning.

2 Helping citizens formulate FOIA requests without exposing sensitive data

Citizens often don't know what to ask for. Build LLM-powered, conversational, or traditional IR systems that help them specify a request from government data, and study how to prevent leaking personal or other sensitive information.

3 Delete or archive? Automatic e-mail classification and explanation using LLMs

According to the "Archiefwet" government officials' e-mail must be archived for FOIA purposes, or deleted. Test whether BERT models and small, local LLMs can make that call, and whether their explanations are plausible.

4 Proactive disclosure: matching open government documents to citizen information needs

Governments should publish information on their own initiative. But what information? Match newly disclosed government documents, to the publics who need to know about them, and evaluate whether they actually reach them.

5 Extracting and consolidating compliance requirements from legal text

Dutch government organisations face overlapping standards and regulations such as ISO 27001, BIO2 and NIS2. Use LLMs to extract and consolidate requirements, tested against an expert-made gold standard.

6 Citizen-relevant information omission in LLM summarization of long OpenGov documents

Summaries increasingly mediate access to long government documents, such as minutes. Study which citizen-relevant information LLMs drop, and whether opposition and minority positions are omitted more.

Interested?

Scan the QR code to send an expression of interest.

