Artificial Intelligence and Law
https://doi.org/10.1007/s10506-023-09383-6

ORIGINAL RESEARCH

Decision support for detecting sensitive text in government records

Anonymous submission

Karl Branting[1] · Bradford Brown[1] · Chris Giannella[1] · James Van Guilder[1] · Jeff Harrold[1] · Sarah Howell[1] · Jason R. Baron[2]

Accepted: 7 November 2023

Abstract
Freedom of information laws promote transparency by permitting individuals and organizations to obtain government documents. However, exemptions from disclosure are necessary to protect privacy and to permit government officials to deliberate freely. Deliberative language is often the most challenging and burdensome exemption to detect, leading to high processing costs and delays in responding to open-records requests. This paper describes a novel deliberative-language detection model trained on a new annotated training set. The deliberative-language detection model is a component of a decision-support system for open-records requests under the US Freedom of Information Act, the *FOIA Assistant*, that ingests documents responsive to an open-records requests, suggests passages likely to be subject to deliberative language, privacy, or other exemptions, and assists analysts in rapidly redacting suggested passages. The tool's interface is based on extensive human-factors and usability studies with analysts and is currently in operational testing by multiple US federal agencies.

Keywords Artificial intelligence and law · Freedom of information law · Machine learning · Human language technology · Human factors analysis · Human–computer interface

1 Introduction

Transparency is vital for representative democracy. Freedom of information laws, such as those described in Sect. 2 below, permit individuals and organizations to obtain government documents, but exemptions to disclosure are necessary to protect privacy and to permit government officials to deliberate freely. Responding to requests under these laws often burdens agency personnel with the tedious and

Extended author information available on the last page of the article

Published online: 10 December 2023

error-prone task of manual identification and redaction of exempt text. Identification of deliberative text is often particularly challenging.

This paper describes the design and implementation of an automated tool, the *FOIA Assistant*, to assist agency personnel in complying with the US Freedom of Information Act (FOIA). The FOIA Assistant ingests documents responsive to a FOIA request, suggests passages that are likely to be exempt, and assists FOIA analysts in efficiently accepting or rejecting suggested redactions. A key technical innovation of the FOIA Assistant is a novel model for detecting a particularly challenging type of exempt text, deliberative language. The model uses BERT (Devlin et al. 2019) fine-tuned on a new annotated data set developed with a FOIA expert.[1] The interface of the FOIA Assistant is based on extensive human-factors and usability studies with FOIA analysts. The FOIA Assistant is currently under evaluation by multiple US federal agencies.

The next section sets forth the nature and scope of open-records laws and describes how the FOIA typifies the challenges that often arise under these laws of balancing disclosure with protection of sensitive information. Section 3 surveys relevant related work on automated detection of sensitive text. Section 4 describes development of a new annotated data set for deliberative language-the form of exempt text that is often the most challenging to detect-and the implementation and evaluation of a new deliberative-language detection model based on that corpus. Section 4 also describes the FOIA Assistant's approach to detecting sensitive personal information. Section 5 describes the human factors analysis performed to identify the requirements for decision support for FOIA analysts and the interface design and functionality that satisfies those requirements. The system-level implementation of the FOIA Assistant is described in Sect. 6. The final section proposes future directions for decision support for detecting sensitive passages in documents subject to the FOIA and other open-records laws.

2 Freedom of information requirements and exemptions

Over 70 nations have adopted some form of freedom of information laws allowing requesters to obtain records of their government (International FOI Laws 2023; Wikipedia: Freedom of Information Laws by Country 2023). The US FOIA, which was enacted on July 4, 1966 (Freedom of Information Act 1966), provides a right of access to records of executive branch agencies. Each US state also has some form of freedom of information law (State Freedom of Information Laws 2023). As subsequently amended, the FOIA includes express recognition of the right of requestors to obtain records in electronic formats, e.g., email and word processing documents in their electronic form (Electronic Freedom of Information Act 1996).

The FOIA establishes a presumption of government openness: individual requestors are entitled to all records of executive branch agencies except those

[1] A BERT-based classifier proved to be the best-performing among several alternative classifiers that we compared in experiments described below in Sect. 4.1.

records or portions of records that fall within any of nine exempt categories (and three additional narrow exceptions) (Freedom of Information Act, as amended 2023). The US Supreme Court has held that government agencies should construe the scope of any exempt categories narrowly (FBI v. Abramson 1982). The primary focus of this paper is identification of deliberations of government officials on matters of official policy (FOIA Exemption 5) and personal information that would invade the privacy of named individuals (FOIA Exemptions 6 and 7(c)). Together, these exemptions are responsible for the majority of exemption applications; exempt deliberative language is often particularly challenging for agency staff to identify.

FOIA Exemption 5 allows agencies to withhold documents subject to several privileges, including attorney-client, attorney work product, and the "deliberative process privilege." The latter applies to documents that reflect advisory opinions, recommendations, proposals, suggestions, and deliberations "comprising a part of a process by which governmental decisions and policies are formulated" (Judicial Watch v. State Department 2018; NLRB v. Sears 1975). As the US Supreme Court recently recognized, "[t]he privilege is rooted in the obvious realization that officials will not communicate candidly among themselves if each remark is a potential item of discovery and front page news" (U.S. Fish and Wildlife Service v. Sierra Club Inc 2021). To be within the scope of this privilege, a document must first be inter-agency or intra-agency in nature, i.e., both sent and received by employees of the executive branch (in the U.S., FOIA does not apply to the legislative or judicial branches of government). For documents meeting this threshold condition, documents must then be both "predecisional" and "deliberative." In the usual case, documents within the privilege are written by lower-level staff for consideration by final decision makers "prior to consummation of the agency's decision-making process" (U.S. Fish and Wildlife Service v. Sierra Club Inc 2021).

Freedom of Information laws in many countries outside the US have parallel provisions protecting deliberations of government officials. Examples include:

- Canada. Four provisions referencing various consultations and deliberations involving federal-provincial matters; international affairs; where directors, officers, or employees of a government institution, or the minister of the Crown or the staff of the minister participate; and records of the Council regarding deliberations (Canada, Access to Information Act 1985).
- India. Exemption for disclosure of cabinet papers including records of deliberations of the Council of Ministers, Secretaries and other officers (India, Right to Information Act 2005)
- Israel. Exemptions for policies still being formed; negotiations of certain kinds; internal discussions; words spoken in the course of an internal inquiry; opinions, drafts, advice, and recommendations given for purposes of decision-making (Israel, Freedom of Information Act 1998).
- South Africa. Exemptions for documents submitted to the Cabinet for consideration or proposed by a Minister of Government, including drafts; an official record of any deliberation or decision of the Cabinet, including drafts; and documents the disclosure of which would involve the disclosure of any deliberation of

decision of the cabinet (South Africa, Draft Model Freedom of Information Law 1999).
– United Kingdom. Four provisions covering information relating to the formulation and development of government policy; communications between ministers and any information relating to those communications; decisions about whether to request legal advice; and information relating to the operation of ministerial private offices (UK, Freedom of Information Act 2000).

Detection of deliberative language exempt from disclosure is thus a task confronting agencies across the democratic world.

Exemption 6 covers "personnel and medical and similar files" when material within those files, if released "would constitute a clearly unwarranted invasion of personal privacy." The US Supreme Court has held that the term "similar files" is to be interpreted broadly, and that all information applying to an individual qualifies as within the exemption's scope (US Department of State v. Washington Post Co 1982). Such material may include what is commonly defined as personally identifiable information (PII) (NIST Guide 2010) or other forms of free-form text on matters that would invade individual privacy. Under Exemption 6, agencies are required to engage in a balancing test to determine the propriety of withholding such text, weighing the extent to which a substantial privacy interest exists against the requester's asserted public interest in disclosure (Washington Post Co. v. HHS 1982). While courts have held that a strong presumption exists in favor of disclosure (Consumers' Checkbook Ctr 2009), agencies routinely withhold PII across a wide spectrum of circumstances.

Using similar language, Exemption 7(c) authorizes the withholding of material relating to law enforcement investigations which "could reasonably be expected to constitute an unwarranted invasion of privacy." This exemption, in contrast to Exemption 6, is worded without using the modifier "clearly" before "unwarranted." This is intended to acknowledge an individual's strong privacy interest in not being associated with criminal activities (NARA v. Favish 2004). Similar exemptions are also recognized in numerous foreign statutes (International FOI Laws 2023).

Agencies cannot simply withhold documents that contain exempt text, but must provide the portions of a document that are "reasonably segregable ... to any person requesting such record after deletion of the portions which are exempt" (Freedom of Information Act Exemptions and Exceptions 2023). Factual material in documents are presumptively outside the scope of the privilege (Heffernan v. Azar 2018). This promotes transparency by preventing agencies from withholding an entire document simply because one line or one page is exempt (Freedom of Information Act Exemptions and Exceptions 2023). However, the requirement of identifying and providing all non-exempt portions of documents imposes a significant burden on agency personnel to closely analyze each document that is responsive to a FOIA request.

The burden for government agencies of compliance with open-records requirements leads to frustrations and delays for requestors. In the US, for example, 928,300 federal FOIA requests were filed in 2022, at a cost of $523 million in processing costs by government agencies and $39 million in litigation costs triggered by the inability of agencies to comply in a timely fashion. As shown in Fig. 1, the

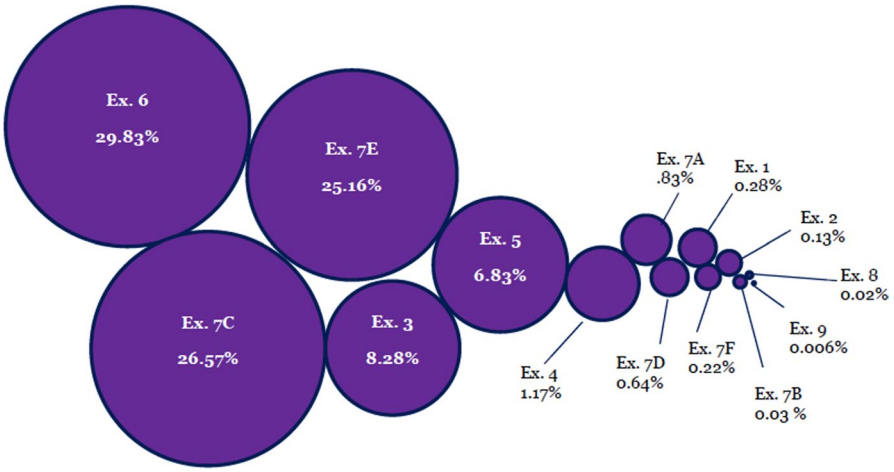

Fig. 1 Breakdown of exemption usage under the US FOIA (source: Summary of Annual FOIA Reports for Fiscal Year 2022). Exemptions 6 and 7(c) involve sensitive personal information, and Exemption 5 covers deliberative language. If the graph were scaled by time and effort on the part of FOIA analysts, Exemption 5 would be the largest circle

exemptions most frequently applied by US federal agencies are those involving sensitive personal information (Exemptions 6 and 7(c)). However, FOIA experts report that Exemption 5 (deliberative language) is typically far more time consuming and requires more expertise than the other exemptions. Automated tools to identify exempt deliberative language and personal information have the potential to improve handling of freedom of information requests both in the US and in the numerous other countries with open-records laws (Summary of Annual FOIA Reports for Fiscal Year 2022).

3 Related work

Identification of documents and passages as exempt from disclosure under freedom of information laws can be operationalized computationally as a text classification problem. Text classification has been addressed over the course of many years using a wide range of methods (Kowsari et al. 2019) including, most recently, Deep Learning approaches (Minaee et al. 2021).

Most of the prior work on automatic detection of privileged documents and passages has focused on sensitive personal information (which corresponds under the US FOIA to detecting text subject to Exemptions 6 and 7(c), above). For example, Graham McDonald and colleagues used government records whose sensitivities had been identified by government assessors as training data for supervised text classification (McDonald et al. 2014). Subsequent work by this group used text classification approaches to identify documents containing sensitive international-relations or personal sensitivities (Exemptions 27 and 40, respectively) under U.K. open-records loaw (McDonald et al. 2017). A user study of the benefits of automatic sensitive

classification prediction demonstrated empirically that even a moderate level of classifier accuracy (Balanced Accuracy of 0.7) could significantly improve both mean reviewer accuracy and mean reviewing speed (McDonald et al. 2020).

Sensitive information detection is important in information retrieval contexts in which relevance must be balanced against sensitivity (Sayed and Oard 2019; Iqbal et al. 2021). Convolutional Neural Networks have been applied for redaction of entire documents, e.g., Chhatwal et al. (2020), but identifying the minimal exempt spans of document text, which is necessary to satisfy the presumption of government openness, requires detecting segment boundaries. Named entity recognition (NER) is the dominant approach to detecting sensitive private information expressed in arbitrary text spans, because NER sequence models are trained to detect named-entity boundaries (Savova et al. 2010; Pearson et al. 2021).

There has been recent work on classifying more complex sensitive information, such as detection of biased language (Sheng et al. 2019) and distinguishing descriptions of criminal incidents (which might include personal details of victims or unsubstantiated accusations) from political events, which are typically public (Narvala et al. 2022). In developing the FOIA Assistant, we limited ourselves to sensitive information consisting of named entities, since detecting such sensitive information is more tractable computationally and was identified as a higher priority by the agency FOIA subject matter experts whom we interviewed.

In contrast to the abundance of prior work on sensitive personal information detection, there has been little published research on identification of deliberative language in the context of the FOIA prior to Baron et al. (2022). As described in greater detail below, Baron et al. annotated each paragraph of a corpus of presidential records from the Clinton White House as to whether it was within the scope of the deliberative process privilege. Initial results applying Support Vector Machine and Logistic Regression classifiers to a term frequency vector representation established that deliberative language was detectable to some degree using these machine learning techniques. Our work extends that of Baron et al. (2022) in four ways.:

1. We expanded the feature spaces of the simple term frequency-based classifiers by including linguistically more complex features.
2. We implemented an additional classifier not requiring explicit feature vector construction but instead utilizing text representations generated through the BERT Devlin et al. (2019) large language model.
3. We annotated the Clinton corpus at the sentence-level to address limitations of the original paragraph-level annotations. As such, we address a different classification problem.
4. We integrated a sentence-level deliberative language classifier into a practical decision support system.

4 Sensitive content identification as a text classification task

The process of identifying sensitive content can be formalized as a task of classifying responsive-document passages as instances or non-instances of each category of exempt text. Individual exemptions generally must be identified separately to permit agencies to justify individual redaction decisions. This section discusses in detail our novel deliberative language classification model and summarizes our approach to sensitive personal information identification.

4.1 Deliberative language classification

A key contribution of this project is development of a set of new annotations to the Baron et al. (2022) FOIA corpus that focus on deliberative language at the sentence level. These new annotations were essential to the development and evaluation of the deliberative language classifier incorporated into the FOIA Assistant.

4.1.1 The original Clinton corpus annotations

Baron et al. collected a corpus of files from the Clinton Presidential Library (2023) by searching with keywords "Elena Kagan" and "Cynthia Rice."[2] They manually culled the resulting files and organized them into batches: K1, K2, K3, K5 (the Kagan files) and R4 (the Rice files). Each batch contains files covering one or more topics with no single topic represented across multiple batches. Baron et al. annotated these files by assigning to each paragraph one of three labels: D1, for paragraphs exempt because they were within the scope of the deliberative process privilege; T0, for trivially non-exempt paragraphs (e.g. file header information); and D0, for all other non-exempt paragraphs. We ignored the T0 paragraphs because we judged them not to be useful for training a content-based classifier. The resulting corpus of paragraphs labeled as D1 or D0 is a valuable resource but has two limitations that complicate its use in building accurate classifiers.

First, some paragraphs labeled D1 do not contain any sentences that are deliberative per se (i.e., sentences like recommendations, opinions, suppositions, or choices that have a deliberative character irrespective of context). Instead, the D1 annotation of such paragraphs was justified by the larger context of the document in which the paragraphs appear. For example, the paragraph below appears as one of a series enumerated paragraphs immediately preceded by "The president could:". Because of this introductory remark, all paragraphs in the series were labeled D1. However, the paragraph below contains only sentences factual in nature and, therefore, is not deliberative per se.

[2] In the Clinton administration, Elena Kagan served as Deputy Assistant to the President for Domestic Policy and Deputy Director of the Domestic Policy Council, among other positions. Cynthia Rice held the title of Special Assistant to the President for Domestic Policy.

> "a. TIMSS on-line challenge.
> Parents will be able—beginning at back-to-school time this fall to download a math and science quiz from the internet, give it to their children, and get a rough sense of what their children need to know in math and science and how they are doing compared to their peers around the world."

Second, some paragraphs contain sentences that are deliberative per se but were labeled D0 because they occur in documents that were not between members of the executive branch, i.e., they are not "intra/inter-agency" (IIA). As an example, the paragraph below was labeled D0 because it appears in a letter from an external organization ("NOW Legal Defense") sent to the Clinton White House Domestic Policy Office. The letter was included in batch K1. The first sentence is deliberative per se as it expresses an opinion ("...we believe it is essential...").

> "In light of the additional research and data about the occurrence of violence in the lives of welfare families, we believe it is essential that the states have all the guidance and support that they need to address this problem and craft workable solutions. Currently, many states are hesitant to elect the FVO for fear of economic sanctions. If the imprecise wording of the Family Violence Option in last year's welfare bill is the stumbling block to a lucid interpretation of this option, then the answer is this Congressional technical clarification. It is imperative that this Administration, with its reputation as a friend to battered women, step forward and support this clarification. We hope that you will look favorably upon our request for your support of S. 671 and use the powers of your office to secure this endorsement."

To avoid these limitations and make the detection of deliberative language a more tractable text classification task, we adopted a new annotation scheme. The primary objective of the new scheme was to assign deliberative or non-deliberative labels to passages of text based only on the content and structure of the passage itself irrespective of the broader context. We chose to identify sentences in IIA documents whose deliberative character depends only on the text of each sentence when considered in isolation, independent of other sentences in the same document. Our label for such sentences is "AD," meaning "Always Deliberative."

4.1.2 New annotations

The new annotations were made through the following process. For each paragraph labeled D1, each sentence was examined and assigned label "AD" if its text was deliberative in isolation, irrespective of context, and was otherwise labeled "Non-AD." For each paragraph labeled D0, if the paragraph appeared in a non-IIA document, the paragraph was dropped; otherwise, all sentences in the paragraph were labeled "Non-AD." This was justified by the fact that paragraphs in IIA documents would not have been originally labeled as D0 if any of their sentences were deliberative per se. Below are examples of AD sentences and Non-AD sentences.

Table 1 Counts of sentences that were labeled AD and non-AD

Batch	AD	Non-AD	Percentage AD
K1	270	795	25
K2	411	528	44
K3	400	849	32
K5	84	964	8
R4	210	499	30

AD. "You could also announce that you will expand AmeriCorps to include a new child-care corps."
AD. "So HHS wants to throw into the technical mix the possibility that Congress could clarify this issue, and they may raise it at our conference call tomorrow."
AD. "I am hopeful that we may be in a position to announce at least the mayor segment when the President meets with the USCM winter meeting attendees for breakfast in the WH on January 30."

Non-AD. "A quick survey of the programs identified above indicates that up to 20,000 prisoners may be receiving benefits improperly."
Non-AD. "1) SSI One-Month Gap—We communicated to HCFA that the one-month gap policy should be made administratively through an All States Letter to the Medicaid Directors as well."
Non-AD. "Spoke to Jim Dobbins at NSC on the status of the proposed SWB process."

Table 1 sets forth the counts of AD and Non-AD sentences in each batch annotated according to our new scheme. The class imbalance between AD and Non-AD sentences in batches K1, K2, K3, and R4 was modest: between 25 and 44% of sentences were labeled AD. However, in batch K5, the class imbalance was much more pronounced: just 8% of the sentences were labeled AD. For this reason, we ignored the K5 batch in the experiments described below.

In summary, we developed a new set of annotations at the sentence level for the Clinton corpus that corresponds to the task that we wish the Exemption 5 detector of FOIA Assistant to perform: identifying individual deliberative sentences. This corpus is freely available to researchers at https://github.com/cmgiannella/FOIA-SENTENCE-DATA.

4.1.3 Classifier methodologies

Our goal was to build a classifier that, given the text of a sentence, predicts whether the sentence should have an AD or Non-AD label. We implemented and compared three classifiers. The first two were extensions of two of the classifiers used in Baron et al. (2022): a Support Vector Machine, and a Logistic Regression classifier using simple word-count based features. Our extensions consisted of modifying the feature space to include the linguistically more complex features described below. The

resulting two classifiers are denoted *LR* and *SVM*. The feature space modification proceeded as follows.

First we modified the simple word count features using the spaCy (2023) named entity recognizer as follows. Before counts were computed, words in the text that were part of a named entity were replaced by normalized strings "<ET>" where ET denotes the name of the entity type, e.g., "<PERSON>". Next, we added 12 additional features (each computed at the sentence level): the number of modal words (e.g., could, would, etc.), adverbs, adjectives, nouns, comparators (adverbs ending in "er"), progressive aspect verbs, perfect aspect verbs, past tense verbs, present tense verbs, first person pronoun subjects, strongly subjective words, and an indicator based on an overall sentence subjectivity classification. The first ten of these were straightforward to calculate from an application of spaCy to the sentence. For the eleventh feature, we utilized Sentiwordnet (Baccianella et al. 2010), a system that assigns a subjectivity score between zero and one to pairs of words and their part-of-speech-tags. We deemed words whose score exceeded 0.9 to be strongly subjective. For the last feature, we used a corpus of sentences with manually assigned "subjective" or "objective" labels (Cornell Movie Review Data 2023) and trained a Bi-RNN classifier using an open-source implementation (Fractalego 2020). If the classifier assigned a "subjective" label to a sentence, the sentence subjectivity indicator was one, otherwise zero.

The last classifier we implemented operates directly on sentence text without requiring the manual specification of domain-specific features. We followed a common approach and added a logistic regression layer (with dropout) on top of the pooled output from the BERT transformer (Devlin et al. 2019). We employed the simple baseline strategy for optimization described by Mosbach et al. (2021) with dropout 0.1. We implemented the classifier in Python using Keras (2023) and the "small_bert," uncased, L-2, H-512, A-8 transformer model. We denote this classifier as *BERT*.

4.1.4 Experiments

As discussed earlier, we ignored batch K5. Using the remaining batches, we carried out the following experiments to evaluate the classifiers (following the design in Baron et al. 2022).

1. Train on the union of K1, K2, K3 and test on R4. This experiment is denoted $K123 = R4$.
2. Train K1, K2 and test on K3—denoted $K12 = K3$.
3. Train K1, K3 and test on K2—denoted $K13 = K2$.
4. Train K2, K3 and test on K1—denoted $K23 = K1$.
5. 10-fold cross-validation on the union of the sentences in K1, K2, K3, and R4—denoted *10FCV*.[3]

[3] Cross-validation was performed on the union of sentences irrespective of batch boundaries or document boundaries within batches. This is consistent with Baron et al. (2022) wherein, among other experiments, cross-validation was performed on the union of paragraphs irrespective of other boundaries.

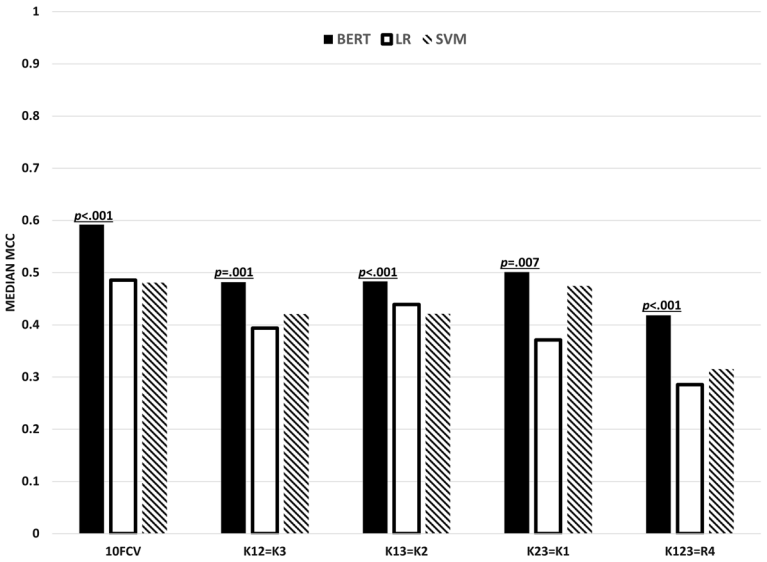

Fig. 2 Median MCC scores of BERT, LR, and SVM. The median is calculated over 30 trials for *SVM* and *LR*, 18 for *BERT*

For the *SVM* and *LR* classifiers, we dropped all words appearing only once in the training data and tuned the hyper-parameters using the same grid search as Baron et al. (2022), except that we used the lemmatizer in spaCy instead of the Porter stemmer. For *LR*, we used L1 regularization.

For each classifier and each experiment, we carry out repeated trials owing to the stochastic nature of the training optimization algorithms: 30 trials for *SVM* and *LR*, 18 trials for *BERT*,[4] In the results below, we report precision, recall, and Matthew's Correlation Coefficient (MCC).[5] We do not report F1 score or accuracy since, in our view, MCC is preferred over those measures.[6]

4.1.5 Results

Figures 2 and 3 depict the results of all experiments. For each experiment, Mood's median test (Conover 1999) was conducted on the results of the classifiers with the highest two medians. The *p*-values from these tests are depicted in the figures above the median bar of the top classifier in each experiment. As an example, for the *10FCV* experiment, *BERT* had the largest median MCC with *LR* second and

[4] The much greater computation time of *BERT* compared to *SVM* and *LR* motivated the running of fewer trials.

[5] MCC assumes a value between negative one and one (Matthews 1975). The extremes indicate a perfect negative and positive agreement, respectively, between the ground truth and classifier labels. Zero indicates no agreement.

[6] Chicco and Jurman (2020) argue that MCC should be preferred over F1 and accuracy since MCC more effectively takes into account "the ratio of positive and negative elements".

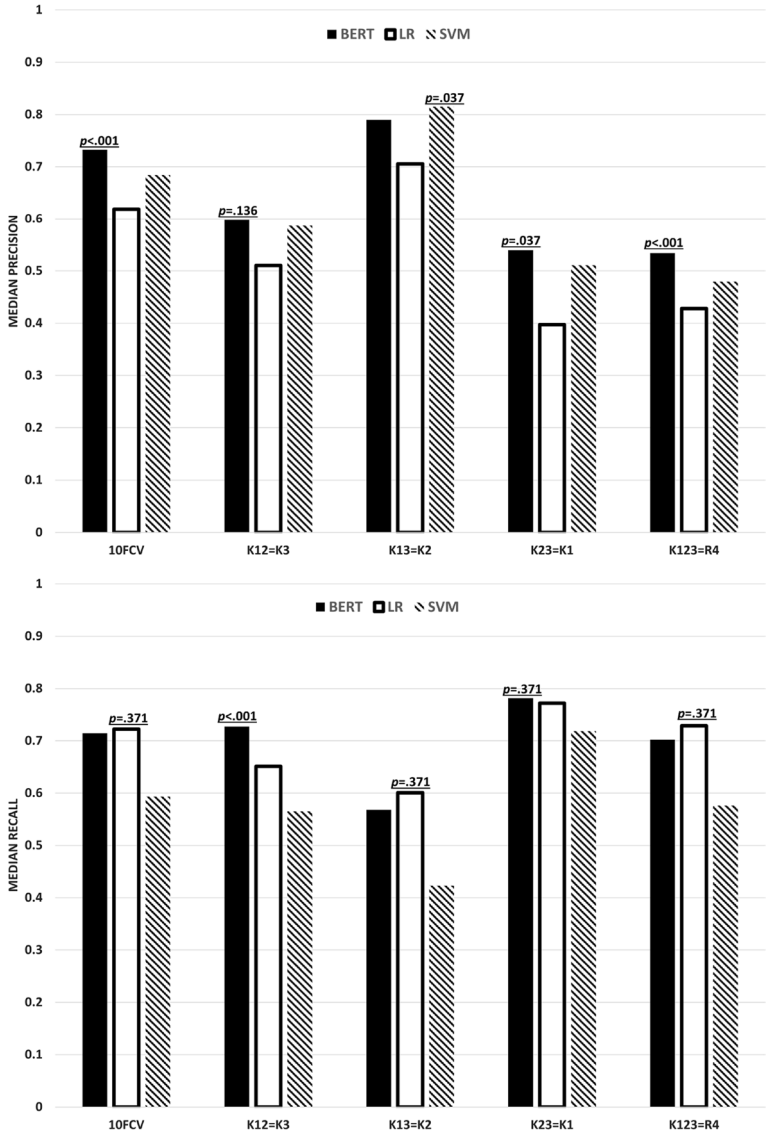

Fig. 3 Median Precision (top) and Recall (bottom) of BERT, LR, and SVM. The median is calculated over 30 trials for *SVM* and *LR*, 18 for *BERT*

the hypothesis that these two medians are in fact not different can be rejected with $p < .001$.

As seen in the figures, *BERT* is generally the best performing classifier. In terms of MCC, *BERT* outperforms all other classifiers in all experiments ($p \leq .007$). In terms of precision, *BERT* outperforms all other classifiers in three of five experiments ($p \leq .037$), is outperformed by *SVM* in one experiment ($p \leq .037$), and has a

statistically insignificant difference ($p = .136$) with *SVM* in one experiment. Finally, in terms of recall, *BERT* outperforms all other classifiers in one of five experiments ($p < .001$) while the difference between *BERT* and *LR* (the next best classifier in terms of recall) on the remaining four experiments is not statistically significant ($p = .371$). In view of its superior performance, *BERT*, trained on all batches except K5, was the model used in the FOIA Assistant to identify deliberative sentences. As discussed in the related work section, above, McDonald et al. (2020) show that using automated exempt text classification to assist human reviewers on a similar task can significantly improve reviewer speed and accuracy when the automated classifier achieves a Balanced Accuracy of 0.7. The median Balanced Accuracy of *BERT* in all our experiments is between 0.72 and 0.8, providing support for the hypothesis that our deliberative language classifier could improve human reviewer speed and accuracy. We leave a formal evaluation of this hypothesis, along the lines of McDonald et al. (2020), to future work.

4.2 Sensitive personnel information detection

The primary research focus of the FOIA Assistant project has been deliberative text identification because of the novelty, difficulty, and importance of this task. However, FOIA analysts must also identify and redact sensitive personal information, so the FOIA Assistant provides assistance with this task as well. Personally Identifiable information (PII) is typically expressed in relatively short text segments. The categories of PII typically of interest to agencies governed by the US FOIA, discussed above, include names, phone numbers, Social Security Numbers (SSNs), and email addresses. These categories overlap the sets of entities that are targeted by Named Entity Recognition (NER) systems, which typically use sequence-learning algorithms (such as conditional random fields) to find the most probable assignment of entity labels to a text.

We use the spaCy library (SpaCy 2023) for NER identification, with post-processing to improve the detection of names. This post-processing includes using name lists to find more mentions of names and pattern matching to find phone numbers and SSNs that are in unusual contexts or that deviate slightly from the standard format.

5 The FOIA assistant

Machine-learning models for detection of exempt text can improve the efficiency and accuracy of agency staff only to the extent that the models can be integrated into analysts' workflows. To determine how to accomplish this integration, we performed a three-stage human factors analysis of analysts' work processes, consisting of cognitive task analysis, requirements development, and collaborative interface design. This process was conducted in collaboration with a US federal agency that provided three subject matter experts with extensive experience reviewing records requested under FOIA. The subject matter experts participated in a series of interviews and

workshops designed to create a cognitive task model of analysts' reasoning and decisions. As part of this agreement, we obtained records previously requested and released by the agency, including both the original record and redacted version. The interface of the FOIA Assistant was incrementally updated to add and improve functionality based on this elicitation process.

5.1 Human factors analysis

A series of interviews and workshops were conducted with the subject matter experts (SMEs) to determine how they approached the FOIA review process and what functional capabilities were required to improve their ability to accomplish this process. The three subject matter experts had a combined 29 years of experience with the FOIA.

5.1.1 Cognitive task analysis

In the first set of interviews, the SMEs participated in individual interviews and responded to questions pertaining to their backgrounds, experience, and general process reviewing records. As part of the initial interview, SMEs were asked to review an unredacted record that had been reviewed by the agency. Each SME provided context and rationale for each redaction. The interviewer probed with additional questions the reasons that the SME redacted each specific passage in the document.

Analysis of the interview data revealed several key elements informed analyst's decisions, including what type of record was requested, who requested the records, and whether the requested records pertain to an open investigation. The analysis also revealed contextual nuances when redacting PII (e.g., titles are included in redactions, not all addresses are considered PII, email domains are excluded from redactions of personal emails, etc.).[7] These key elements informed an initial series of feature requirements, which served as the basis for the subsequent workshop.

5.1.2 Requirements development

Once the FOIA review process was well understood, a series of workshops was held to identify the features that would be most helpful for analysts and to develop an interface design incorporating those features. The first workshop used a prioritization matrix (Nayak and D'Souza 2019) to understand what features would be most useful to an analyst. Data from this workshop revealed that analysts were not receptive to decision support that could be construed as attempting to automate their professional judgment. Instead, they wished for features that would improve their ability to exercise that judgment and increase awareness of sensitive information in the records.

[7] We note that the nuances of decisions identified by the SME's from one agency might differ from those of FOIA analysts at another agency.

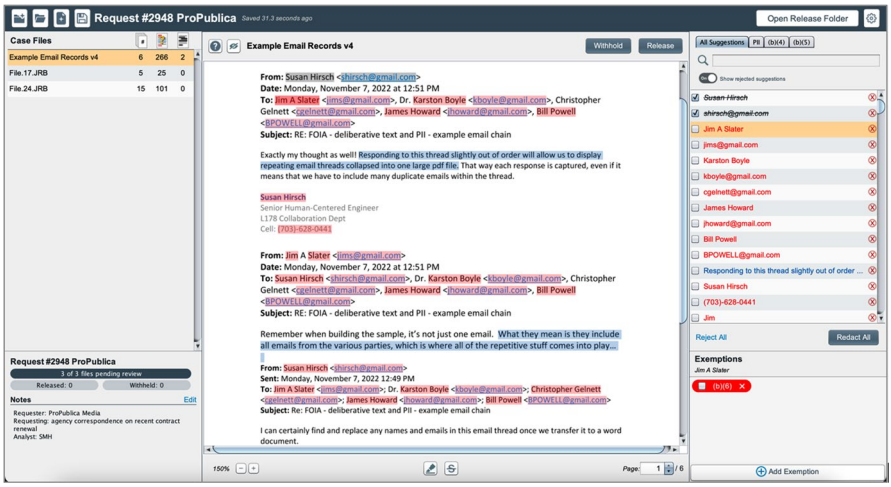

Fig. 4 The interface of the FOIA Assistant showing a panel for case files on the left, document text with color-coded suggested redactions in the middle panel, and tabs for bulk sorting and accepting, or rejecting suggested redactions on the right panel. (Color figure online)

5.1.3 Collaborative interface design

The subsequent workshops focused on development of a user-interface design emphasizing these features. We conducted a participatory design session in which the SMEs collaboratively created an initial set of low-fidelity mockups showing how these features might be presented in a user interface. This workshop was facilitated as a focus group. The designs produced from the workshops were incorporated into the FOIA Assistant prototype.

Several group workshops with the SMEs were held to refine the requirements for the AI/ML features detecting text that could be exempt under Exemptions 5 and 6.[8] The first involved a group card sort (Righi et al. 2013) to identify what types of statements are typically exempt under Exemption 5. In this card sort, participants categorized deliberative statements as: Likely Exempt, Requires More Information to be Considered Exempt, or Likely Not Exempt. The interviewer asked probing questions, and the SMEs elaborated on why or why not a statement could be exempt.

Similarly, two of the SMEs participated in a card sort that identified and ranked all the types of PII that could be exempt under Exemption 6. Participants were given an initial set of standard PII types (e.g., names, emails, etc.), and participants were then invited to add additional PII types based on their experience (e.g., job duty and title, biometric data, drivers licenses, etc.). These were ranked by relative sensitivity, which was defined by the participants as how unique the data is to an individual and

[8] Since text subject to Exemption 7(c) is almost always also subject to Exemption 6, we restrict automatic detection to the latter, leaving it to the analyst to add 7(c) to 6 if appropriate.

Fig. 5 Menu for choosing whether to accept or reject a suggested redaction

how easily retrievable that data may be. Both card sorting activities for Exemptions 5 and 6 informed the design for the AI/ML features for detecting and suggesting PII and deliberative passages as exempt.

5.2 Interface design and functionality

The interface of the FOIA Assistant was designed to implement the requirements formulated through the human factors analysis. As shown in Fig. 4, users interact with the tool through three panels. The left panel displays the case currently under review by the user (which corresponds to an individual FOIA request) and, for each document in the case, the number of pages, suggestions (i.e., passages flagged as potentially exempt), and redactions (i.e., suggestions that have been accepted) in that document. The middle panel displays the text of the document with suggestions highlighted in colors that indicate the type of sensitive text, e.g., red for PII, blue for deliberative language, and green for dollar amounts (which can be sensitive under Exemption 4, which covers privileged commercial or financial information). These suggestions are intended to be reviewed and accepted by the user before any redaction is performed in the released file; analysts are free to accept, reject, or ignore any suggestion.

The interface presents multiple affordances for an analyst to accept or reject suggestions, depending on analyst's preferences. For example, analysts can review the text in the middle panel and select any highlighted region. A right click displays a menu to accept or reject that specific suggestion (Fig. 5), or the user can use the Spacebar key as a shortcut for accepting the current suggestion. Alternatively, users can refer to the right panel, which lists all of the suggestions, each with a corresponding checkbox (Fig. 6A). Selecting the checkbox accepts the suggestion and applies the corresponding redaction. Clicking the `X' to the right of the text rejects the suggestion, removing the highlighted region from the document panel.

Using the Redact All feature in conjunction with the available filters enables users to rapidly apply redactions to repeated text, such as PII. The suggestion panel provides filtered views of the suggestions, permitting the analyst to view lists of suggestions associated with each exemption type. An analyst can perform a keyword search to reduce the displayed list to only matching results, as shown in Fig. 6B). Clicking Redact All accepts every suggested redaction displayed in the list.

If an analyst wishes to redact a passage under an exemption not currently implemented (i.e., an exemption other than 4, 5, or 6) or redact text not suggested for redaction by the tool's currently implemented models (i.e., because of a false

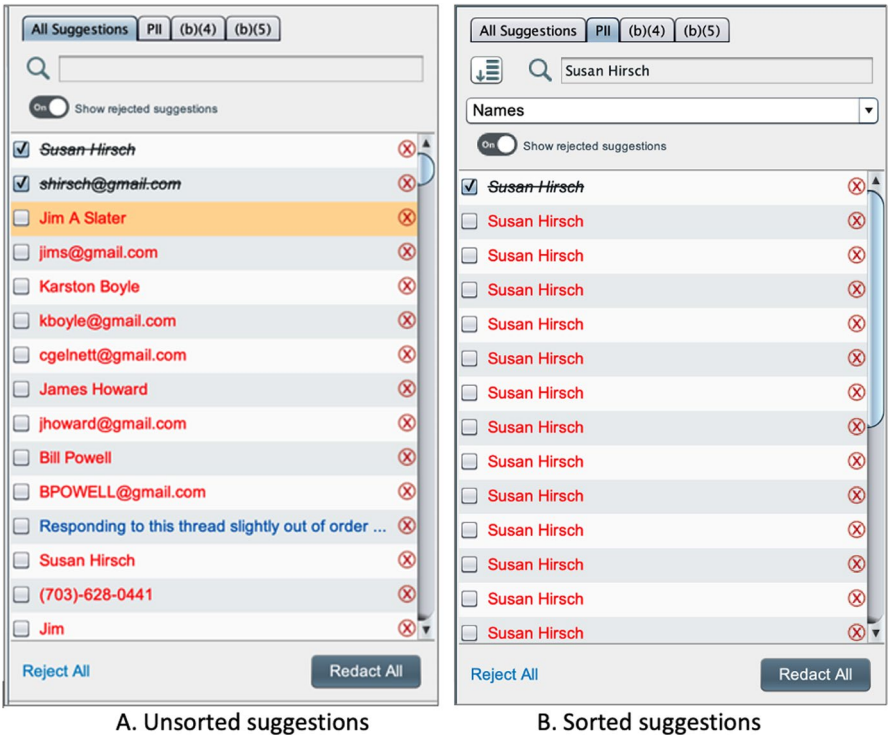

Fig. 6 Suggested redactions displayed in the right panel of the FOIA Assistant, which enables bulk redaction decisions on repeated texts. The left side shows suggestions in order of appearance in the text, and the right side shows suggestions after alphabetic sorting

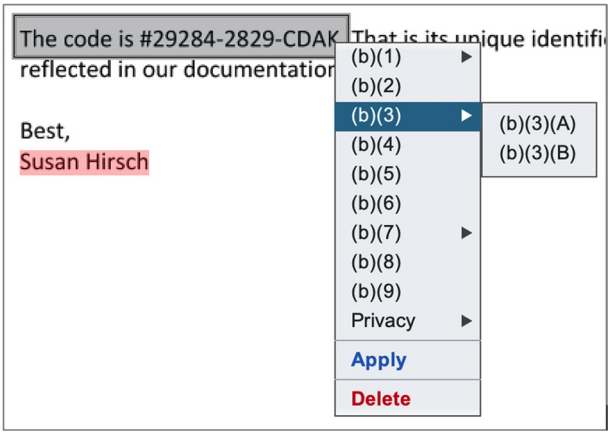

Fig. 7 Ad hoc redaction menu

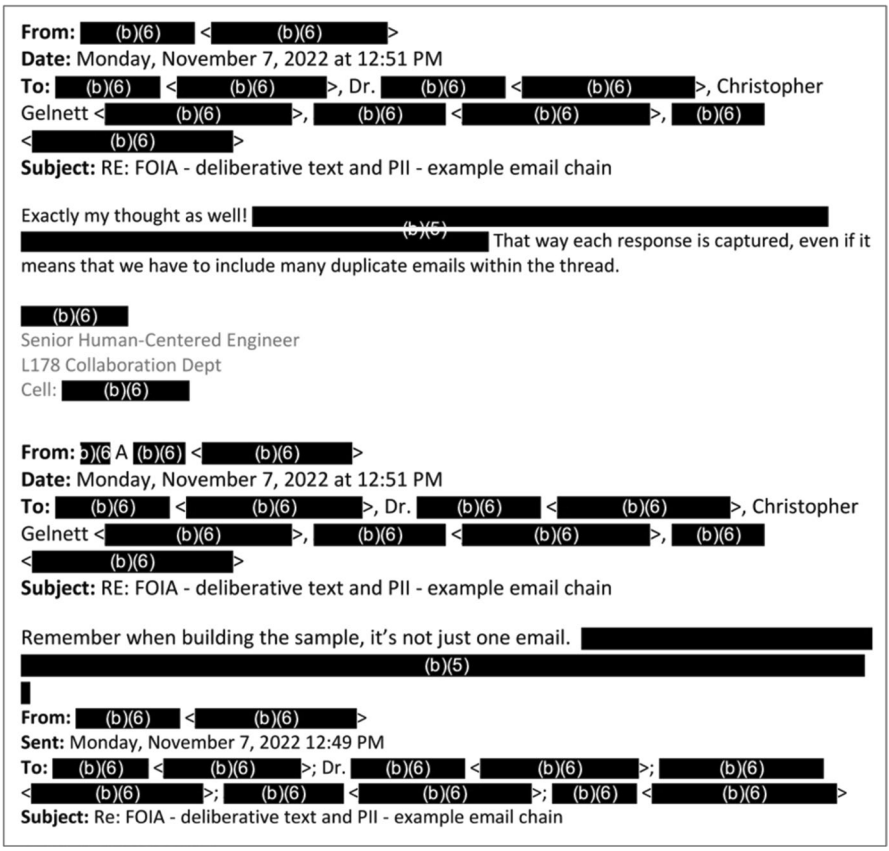

Fig. 8 A redacted file ready for release

negative by one of the models), the analyst can create an ad hoc redaction using a drawing tool available in the middle panel (see Fig. 7). The ad hoc redaction tool, shown in Fig. 7, enables the user to draw a box around any desired content for selection and redaction under any appropriate FOIA exemption or Privacy Act exemption.

The drawing tool on the right provides two types of obfuscation: strikethrough and obscure. The strikethrough can be applied to text that can be released but is no longer valid (e.g., marking a document as Classified in the margins). The obfuscation tool could be used to obscure a wet signature or other sensitive content that doesn't necessarily require an exemption justification.

Once the file has been reviewed and all necessary redactions applied, the analyst can release the file by clicking Release at the upper right of the middle panel (see Fig. 4). The tool also provides the option to withhold a file in full if necessary. Releasing or withholding a file will update the Case Files table, marking them accordingly (R for released files, W for withheld files). Files selected to be withheld

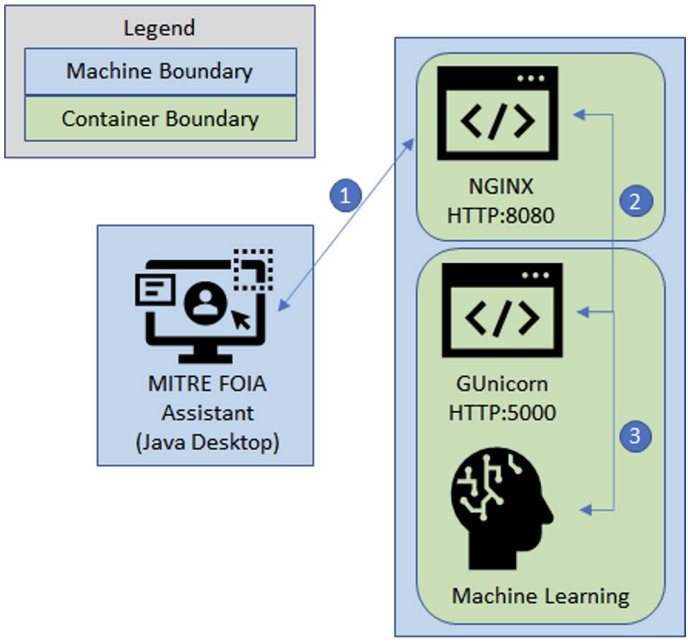

Fig. 9 The FOIA Assistant's client–server architecture

are withheld in full and therefore do not require an alternate redacted version to be created. If a file is released with redactions applied, a new PDF file is generated. To access the released version of the file, the user can double click the file name in the Case Files list or access the folder in their directory by clicking Open Release Folder. Figure 8 shows an example of a released file with redactions.

6 System implementation

6.1 System architecture

The FOIA Assistant is implemented in the client–server architecture depicted in Fig. 9. The Java Desktop application interacts with the file system (ingesting PDF documents, extracting text, and writing the annotated and redacted versions of document), invokes the backend service to obtain suggestions from the machine-learning models, and implements the user interface functionality described above. The server centralizes the machine-learning models. The client sends documents to the server in batches of at most 3 documents at a time to permit analysts to start working on documents with suggestions without having to wait for the full set of documents to be processed. The server has a modular design that can accommodate additional models to enable the FOIA Assistant to be customized to agencies needing other types of sensitive text.

6.2 Document processing steps

Each individual document in a case is processed through the following steps:

1. The FOIA Assistant desktop application reads the original PDF document and performs text extraction, including calculating position and size information for each character extracted from the document.
2. A simplified JSON representation of the document, containing only the extracted text, is sent to the FOIA Exemption suggestion service (Step 1 in the diagram above).
3. The NGINX web server acts as a reverse proxy to the GUnicorn web server for REST calls. Its primary purpose is to efficiently serve static content, such as documentation (Step 2).
4. The FOIA Exemption suggestion service contained in the GUnicorn web server uses a variety of techniques, including custom artificial intelligence models, to annotate the text and return a richly annotated form of the document (Step 3).
5. The FOIA Assistant uses the resultant annotated form of the document to display suggestions for redaction to the analyst on a rendering of the original PDF document.

6.3 PDF extraction

The FOIA Assistant currently handles only documents in PDF format because PDF is widely used across US government agencies (PDF in Government 2023) and because many agencies convert documents into PDF format at an early state of document analysis and redaction.

The accuracy of the models for detecting deliberative language and PII (see Sect. 4, above) depends on accurate extraction of text from the native document format. It is particularly important to recover the sentence order of words in the document, both because the deliberative language model classifies text at the sentence level and because the spaCy's NER model depends on an accurate sequential context for segmentation (determining the span of an entity) and labeling (the label of a given span may depend on the labels of nearby spans).

Extraction of text from a PDF document (Step 1 in Sect. 6.2, above) to accurately recover word order is challenging due to the nature of the representation of the text within the PDF document and the fact that PDF representations can come in several internal formats. The current implementation of the FOIA Assistant is designed to handle native text and embedded OCR results in PDF documents; handling image based content and other document formats is future work. A custom PDF extraction engine was developed (described at a high level below) to meet two critical requirements of the project. First, to develop models for deliberative language it was desirable to use existing annotated documents that comprise a natural dataset for the task. Annotations in the existing documents were maintained as geometric shapes within the document. These shapes needed to be accurately aligned with the underlying

Concerning (2), while the same gain occurs from validating either a true positive (TP) or a true negative (TN) – this gain is 0, since the human annotator will not change their labels – the gain that occurs from validating a false positive (FP) or a false negative (FN) may be, as shown in [2], different. When the two gains are different, the utility-theoretic approach tested in this paper:

- is indeed different from an approach based on (1) only, which we describe as the *purely probabilistic approach* (indeed, the two approaches instead coincide when the two gains are the same), and

Fig. 10 An excerpt from a PDF document

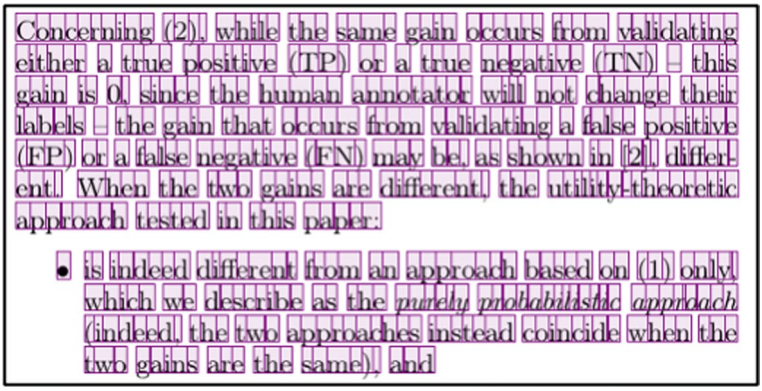

Fig. 11 PDF document excerpt with segment bounds

text to determine which text was being redacted. Secondly, the output of the tool needed to maintain visual fidelity with the original documents. This required that the suggestions and markup processes be represented on a rendering of the original document in which the text processed by the system maintains its original positioning and sizing.

Several extraction engines were reviewed before deciding to create a custom extraction process. Most existing text extraction engines, such as TIKA[9] and pdfminer,[10] do not maintain the requisite visual information as their use cases do not require this information. Additionally, many of the engines had errors in text extraction on more complex layouts such as multi-column documents. Most, for instance, interleave text fragments from multiple columns, rendering the output incomprehensible and introducing an unacceptable error rate in the suggestion engine. Some engines also subtly modify the text to normalize it for application specific purposes.

[9] https://tika.apache.org/.

[10] https://pypi.org/project/pdfminer/.

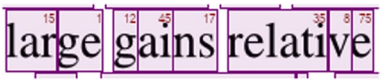

Fig. 12 Native PDF text segmentation

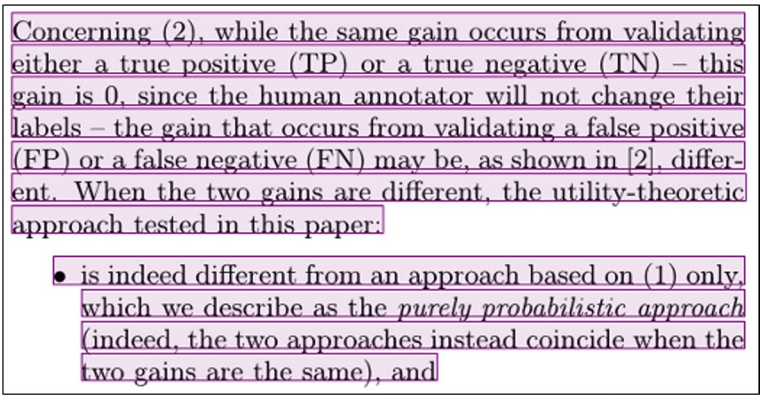

Fig. 13 Line-based resegmentation

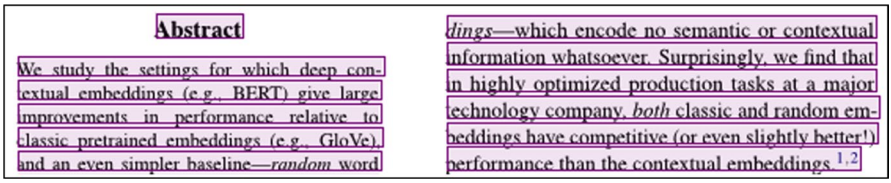

Fig. 14 Multi-column layout

Examples of this included removing hyphenation in words that spanned multiple lines. This normalization process can introduce errors by overcorrecting and misaligning the text to its visual counterpart.

The first step performed by the custom text extractor is reading the internal format of the PDF document and identifying segments of the document that are text-based using PDFBox (Apache PDFBox 2023). Each text character on the screen is represented as a glyph within an embedded font. The embedded font information is used to map each glyph to its equivalent Unicode representation while maintaining its visual boundary information. Text characters are often represented in groups within the document, although these groups are not based on text constructs such as words, phrases, or sentences. For known glyphs without a valid Unicode mapping, OCR is used, per-character, to attempt to determine the Unicode character that is being represented by the glyph. The result of converting PDF input into glyphs is illustrated in Figs. 10 and 11.

Once the segment and glyph boundaries are extracted, the next step is assembling the independent text snippets into groups useful for text analytics. As illustrated by the PDF fragment shown in Fig. 12, the text segments within the PDF document

are not in general represented in the way a text-based representation would group them. The segments are often only a few characters long and do not represent words, phrases, or sentences. The red numbers in the top right of each of the segments in Fig. 12 represent their order in the underlying PDF, showing that the segments are not in general sequenced in top-down, left-to-right order. Instead, the grouping and ordering of the characters within the PDF is based on such factors as optimizing the rendering or editing of the document. Note that whitespace between the characters is often not represented by the PDF document (because representing and rendering whitespace takes space and time and often does not affect the visual display).

The text extraction process groups the characters into segments on visual "lines" based on the glyph/segment rotations and boundaries. This involves potentially rotating, grouping, and comparing boundary proximities and overlaps. The spacing between words is inferred based on heuristics of presumed space sizes for the glyphs as they are assembled. This interpretation of whitespace can introduce some errors into the text extraction process, as embedded font information is often unreliable. Figure 13 shows the results of the initial line-based resegmentation. Figure 14 illustrates that larger gaps prevent assembling groups of text that are on the same visual line into a single segment. This permits downstream analysis to group them within more complicated text layouts, such as the two-column example shown in Fig. 14.

The current implementation of the FOIA Assistant doesn't fully recover the human read-order of PDF documents, but is sufficient for our current sentence-level text analytics. Extending the current work to group the line-based segments into paragraph blocks, based on the justification of the segments, is future work.

7 Conclusions and future work

This paper has presented a new deliberative-language detection corpus and model embedded in a decision-support system for open-records requests under the US Freedom of Information Act. This system, the *FOIA Assistant*, ingests documents responsive to an open-records requests, suggests passages likely to be subject to be exemption under the deliberative language or privacy exemptions, and assists analysts in rapidly redacting suggested passages. The FOIA Assistant is currently in operational testing in multiple US federal agencies.

We hope that our new annotations of the Baron et al. (2022) deliberative language corpus will prove useful for other researchers and will be an exemplar for additional corpus-based work on deliberative language detection in other jurisdictions and nations. It would be extremely beneficial to the community for other teams to contribute deliberative language corpora derived from other genres of government documents.

A number of significant future tasks remain. The FOIA Assistant automates suggestions for Exemptions 5 and 6, which are frequent in all US agencies and relatively uniform across agencies. However, some agencies would require additional exemptions models. Specialized to their own needs and practices. As described

above, the modular design of the FOIA Assistant architecture facilitates the addition of new models.[11] Thus, we anticipate that the tool may be customized for individual agencies.

Development of the FOIA Assistant was premised on the observation that detection of sensitive language to be withheld from disclosure is a problem common to federal and state agencies across the US and in democratic nations across the world. This project has demonstrated that machine learning models for passage classification can be combined with human factors analysis of analysts' functional requirements to produce a decision support system that can improve analysts' speed, accuracy, and consistency. Such decision support systems promise to improve agencies' transparency and responsiveness to open-records requests.

Acknowledgements The MITRE Corporation is a not-for-profit company, chartered in the public interest. This document is approved for Public Release; Distribution Unlimited. Case Number 23-1731. © 2023 The MITRE Corporation and patent pending. The authors wish to acknowledge the contributions of Alex Yeh to the sensitive personal information detecting component of the FOIA Assistant.

Author Contributions All authors contributed to the conception and design of the system described in the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest. There is and has been no financial relationship between any author and any organization of relevance to this work. Further, no author is currently in any negotiations regarding future paid employment with any organization of relevance. The manuscript has not been submitted or published anywhere else, nor will it be submitted elsewhere until completion of the editorial process. All authors have approved the manuscript for submission and consent to publication should this submission be successful. All interviews and workshops were conducted in accordance with procedures approved by the the authors' Institutional Review Board.

References

Apache PDFBox 2.0.27 (2023) https://pdfbox.apache.org/. Accessed 30 January 2023

Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources and evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta

[11] For demonstration purposes, a module was implemented that identifies a narrow category of Exemption 4 text consisting of dollar amounts. In general, the criterion for confidential or proprietary information or trade secrets text subject to Exemption 4 is specific to individual agencies.

Baron JR, Sayed MF, Oard DW (2022) Providing more efficient access to government records: a use case involving application of machine learning to improve FOIA review for the deliberative process privilege. J Comput Cult Herit 15(1):1–19

Canada, Access to Information Act (1995) §§14(a), 15(1), 21(1)(b) & 69(1)(c)

Chhatwal R, Keeling R, Gronvall P, Huber-Fliflet N, Zhang J, Zhao H (2020) CNN application in detection of privileged documents in legal document review. In: 2020 IEEE international conference on big data (big data), pp 1485–1492. https://doi.org/10.1109/BigData50022.2020.9378182

Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21(1):1–13

Clinton Digital Library (2023) https://clinton.presidentiallibraries.us/. Accessed 1 December 2023

Conover WJ (1999) Practical nonparametric statistics, 3rd edn. Wiley, New York

Consumers' Checkbook v.HHS (2009) For the study of Servs. v. HHS. 554 F.2d 1046, 1057 (D.C. Cir. 2009) (FOIA's "presumption favoring disclosure ... is at its zenith under Exemption 6")

Cornell Movie Review Data (2023) https://www.cs.cornell.edu/people/pabo/movie-review-data/. Accessed 31 January 2023

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL. Association for Computational Linguistics, pp 4171–4186

Electronic Freedom of Information Act Amendments of 1996 (1996) Pub. L. 104-231, amending 5 U.S.C. 552(f)(2)(a) (defining "record" as including in an "electronic format")

FBI v. Abramson (1982) 465 U.S. 615, 630

Freedom of Information Act (1966) Pub. L. 89-487, codified at 5 U.S.C. 552(b)

Fractalego (2020) Subjectivity detection with Bi-RNNs. https://github.com/fractalego/subjectivity_classifier

Freedom of Information Act as amended (2023) 5 U.S.C. 552(b)(1)-(9) and (c)(1)-(3). Accessed 1 December 2023

Heffernan v. Azar (2018) 317 F.Supp.3d 94, 125 (D.D.C.2018)

India Right to Information Act (2005) §8, Clause I

International FOI Laws (2023) National Freedom of Information Coalition. https://www.nfoic.org/international-foi-laws/. Accessed 30 November 2023

Iqbal M, Shilton K, Sayed M, Oard D, Rivera J, Cox W (2021) Search with discretion: value sensitive design of training data for information retrieval. Proc ACM Hum–Comput Interact 5(CSCW1):1–20

Israel Freedom of Information Act (1998) §9.2

Judicial Watch v. State Department (2018) 349 F.Supp.3d 1, 7 (D.D.C. 2018)

Keras (2023) https://keras.io/. Accessed 31 January 2023

Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D (2019) Text classification algorithms: a survey. Information 10(4):150

Matthews BW (1975) Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochim Biophys Acta (BBA) Protein Struct 405(2):442–451

McDonald G, Macdonald C, Ounis I (2017) Enhancing sensitivity classification with semantic features using word embeddings. In: Jose, J.M., Hauff, C., Altingövde, I.S., Song, D., Albakour, D., Watt, S.N.K., Tait, J. (eds.) Advances in information retrieval—39th European conference on IR Research, ECIR 2017, Aberdeen, UK, April 8–13, 2017, proceedings, lecture notes in computer science, vol 10193, pp 450–463

McDonald G, Macdonald C, Ounis I (2020) How sensitivity classification effectiveness impacts reviewers in technology-assisted sensitivity review. ACM Trans Inf Syst 39(1):1–34

McDonald G, Macdonald C, Ounis I, Gollins T (2014) Towards a classifier for digital sensitivity review. In: de Rijke M, Kenter T, de Vries AP, Zhai C, de Jong F, Radinsky K, Hofmann K (eds) Advances in information retrieval—36th European conference on IR Research, ECIR 2014. Amsterdam, The Netherlands, April 13–16, 2014. proceedings, lecture notes in computer science, vol 8416. Springer, pp. 500–506

Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J (2021) Deep learning-based text classification: a comprehensive review. ACM Comput Surv 54(3):1–40

Mosbach M, Andriushchenko M, Klakow D (2021) On the stability of fine-tuning BERT: misconceptions, explanations, and strong baselines. In: Proceedings of the 9th international conference on learning representations, ICLR

NARA v. Favish (2004) 541 U.S. 157, 165–66
Narvala H, Mcdonald G, Ounis I (2022) The role of latent semantic categories and clustering in enhancing the efficiency of human sensitivity review. In: ACM SIGIR conference on human information interaction and retrieval, pp 56–66
NLRB v. Sears (1975) Roebuck and Co. 421 U.S. 132, 150
Nayak V, D'Souza R (2019) Comparison of multi-criteria decision making methods used in requirement engineering. CiiT Int J Artif Intell Syst Mach Learn 11(5):92–96
NIST Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (2010) Special Publication 800-122. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf
Pearson C, Seliya N, Dave R (2021) Named entity recognition in unstructured medical text documents. In: 2021 International conference on electrical, computer and energy technologies (ICECET), pp 1–6
PDF in Government (2023) https://www.talkingpdf.org/pdf-in-government/. Accessed 4 February 2023. ("There's not a federal agency that does not use PDF," says Greg Pisocky, Business Development Manager for Adobe Systems. 'Acrobat software and Adobe PDF are key technologies in some capacity at all branches and levels of government, the military and virtually every agency')
Righi C, James J, Beasley M, Day DL, Fox JE, Gieber J, Howe C, Ruby L (2013) Card sort analysis best practices. J Usability Stud 8(3):69–89
Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG (2010) Mayo clinical text analysis and knowledge extraction system (cTAKES). JAMIA 17(5):507–513
Sayed MF, Oard DW (2019) Jointly Modeling Relevance and Sensitivity for Search among Sensitive Content. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, SIGIR. Association for Computing Machinery, pp 615–624
Sheng E, Chang KW, Natarajan P, Peng N (2019) The woman worked as a babysitter: on biases in language generation. EMNLP/IJCNLP 1:3405–3410
South Africa, Draft Model Freedom of Information Law (Adopted in 1999) §25(1)(a)-(d)
SpaCy (2023) Industrial-strength natural language process in python. https://spacy.io/. Accessed 19 January 2023
State Freedom of Information Laws (2023) National Freedom of Information Coalition. https://www.nfoic.org/state-freedom-of-information-laws/. Accessed 4 February 2023
Summary of Annual FOIA Reports for Fiscal Year (2022) US Department of Justice. https://www.justice.gov/oip/page/file/1581856/download. Accessed 14 May 2023
UK Freedom of Information Act (2000) §35(1a)-(1d). See generally, information commission's office, government policy (section 35). https://ico.org.uk/media/for-organisations/documents/1200/government-policy-foi-section-35-guidance.pdf. Accessed 4 February 2023
US Department of State v. Washington Post Co (1982) 456 U.S. 595, 602
U.S. Fish and Wildlife Service v. Sierra Club Inc (2021) 141 Supreme Court 777
Washington Post Co. v. HHS (1982) 690 F.2d 252, 261 (D.C. Cir. 1982)
Wikipedia: Freedom of Information Laws by Country (2023) https://en.wikipedia.org/wiki/Freedom_of_information_laws_by_country. Accessed 1 February 2023

Authors and Affiliations

Karl Branting[1] · Bradford Brown[1] · Chris Giannella[1] · James Van Guilder[1] · Jeff Harrold[1] · Sarah Howell[1] · Jason R. Baron[2]

✉ Karl Branting
lbranting@mitre.org

Bradford Brown
bcbrown@mitre.org

Chris Giannella
cgiannella@mitre.org

James Van Guilder
jamesv@mitre.org

Jeff Harrold
jharrold@mitre.org

Sarah Howell
showell@mitre.org

Jason R. Baron
jrbaron@umd.edu

[1] The MITRE Corporation, McLean, VA, USA

[2] University of Maryland, College Park, MD, USA