

PNNL-35927

# Testing Artificial Intelligence Tools to Streamline the Public Comment-Review Process for NEPA Environmental Reviews

April 2024

Mike Parker  
Dan Nally  
Anurag Acharya  
Gihan Panapitiya

Ann Miracle  
Hung Phan  
Sai Munikoti  
Sameera Horawalavithana

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY  
*operated by*  
BATTELLE  
*for the*  
UNITED STATES DEPARTMENT OF ENERGY  
*under Contract DE-AC05-76RL01830*

Printed in the United States of America

Available to DOE and DOE contractors from the  
Office of Scientific and Technical Information,  
P.O. Box 62, Oak Ridge, TN 37831-0062;  
ph: (865) 576-8401  
fax: (865) 576-5728  
email: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)

Available to the public from the National Technical Information Service  
5301 Shawnee Rd., Alexandria, VA 22312  
ph: (800) 553-NTIS (6847)  
email: [orders@ntis.gov](mailto:orders@ntis.gov) <<https://www.ntis.gov/about>>  
Online ordering: <http://www.ntis.gov>

## Acronyms and Abbreviations

|      |                                       |
|------|---------------------------------------|
| AI   | artificial intelligence               |
| DOE  | U.S. Department of Energy             |
| LLM  | large language models                 |
| NEPA | National Environmental Policy Act     |
| PNNL | Pacific Northwest National Laboratory |
| RAG  | retrieval-augmented generation        |

## Terminology

|                           |  |
|---------------------------|--|
| Bracketing                | Dividing narrative content into topic-focused sections, informed by categorization where applicable.   |
| Categorization            | Assigning subject matter categories to comments.   |
| Comment                   | A bracketed section of correspondence, typically one or more sentences of sequential text that express a distinct idea, that includes subject matter-specific context and requests that should be vetted by a subject matter expert. |
| Correspondence            | Public meeting transcripts, emails, mailed letters, electronic submissions to web-based comment platforms such as regulations.gov.   |
| Narrative content         | NEPA-relevant content of a piece of correspondence that does not include headers, footers, salutations, and system-generated information.  |
| Non-comment               | A bracketed section of correspondence that is not a comment (i.e., it contains neither context nor a specific request).  |
| Subject matter categories | Specific sections used in a NEPA scoping document, generally including categories specific to the specific action and to the NEPA process itself.  |

## Contents

|  |    |
|--|----|
| Acronyms and Abbreviations .....   | ii |
| Terminology .....  | ii |
| 1.0 Introduction .....   | 1  |
| 2.0 Current Processes and Possible AI-Augmented Solutions to Processing Public<br>Comments ..... | 2  |
| 2.1 Current Human-Driven Process .....   | 2  |
| 2.2 AI Solutions .....   | 3  |
| 2.2.1 Prompting .....  | 3  |
| 2.2.2 Instruction Tuning .....   | 6  |
| 3.0 Current Limitations and Next Steps .....   | 8  |
| 3.1 Current Limitations .....  | 8  |
| 3.2 Next Steps .....   | 8  |

## Figures

|           |   |   |
|-----------|---|---|
| Figure 1. | Conceptual AI-Assisted Workflow for Public Comment Review .....   | 2 |
| Figure 2. | Possible AI-Integration into Public Comment Processing .....  | 3 |
| Figure 3. | Example Prompt Used to Test the Ability of LLMs to Bracket and<br>Categorize Public Comments .....  | 4 |
| Figure 4. | Alternate Prompt Used to Bracket and Categorize Correspondence .....  | 5 |
| Figure 5. | Prompt Template Used for Instruction Tuning .....   | 6 |
| Figure 6. | (a) Comment/Non-comment Classification Accuracy and (b) Set 3<br>Sentence Classification Accuracies When Only Set Two Documents<br>Were Used to Train the LLM ..... | 6 |
| Figure 7. | Examples of Comment Filtration Using Instruction Tuning .....   | 7 |
| Figure 8. | Example of Future Tool Used to Suggest Subject Matter Categories .....  | 8 |

## Tables

|   |   |
|---|---|
| Table 1. Favorable LLM Response to a Comment Bracketing and Categorization Prompt ..... | 4 |
| Table 2. Example of Successful Bracketing and Categorization .....                      | 5 |
| Table 3. Example of Unsuccessful Bracketing and Categorization .....                    | 5 |

## 1.0 Introduction

The National Environmental Policy Act (NEPA) and its implementing regulations require federal agencies to evaluate foreseeable environmental impacts of a proposed action and consider reasonable alternatives when making decisions to authorize, fund, or conduct such actions. For proposed actions that may significantly affect the quality of the human environment, such as construction of a new transmission line across public lands, agencies must prepare a detailed environmental impact statement and fulfill procedural requirements to inform and gather input from the public. Thoughtfully directed and controlled application of artificial intelligence (AI) could improve outcomes and efficiency of reviews conducted under NEPA and similar permitting practices.

At the beginning of a NEPA review when an agency announces that it intends to prepare an environmental impact statement, there is an open call for all interested parties, which typically include other federal agencies; state, local, and tribal governments; industry representatives, and the general public, to submit comments regarding the proposed action. This comment period is referred to as public scoping and must be a minimum of 30 days. Agency notices announcing the dates of the scoping period provide specific instructions on how to submit comments. Common pathways for agencies to receive comments are through electronic submissions to web-based comment platforms such as regulations.gov, an email inbox, mailed letters, or oral or written statements at public meetings. Correspondence documents and meeting transcriptions containing public comments are typically received or download as text, Microsoft Word, or portable document format (PDF) files.

Processing and reviewing each piece of correspondence and identifying and categorizing discrete comments within for further review by subject matter experts is a tedious and a time-consuming task that is prone to human error and inconsistencies. In this report, we relate our early experiences testing AI applications enabled by large language models (LLMs) to automate, with human supervision and review, select steps in the comment-review process. Figure 1 illustrates a conceptualized workflow for reviewing public comments and identifies specific tasks that could be performed by AI. Continued research and experimentation aim to encapsulate these AI-assisted comment-review capabilities and combine them with other capabilities, such as comment summarization and report generation, into a software application with an intuitive user interface.

Subsequent sections of this document describe current processes and possible AI-augmented solutions to processing public comments (Section 2.0) as well as current limitations and next steps (Section 3.0)

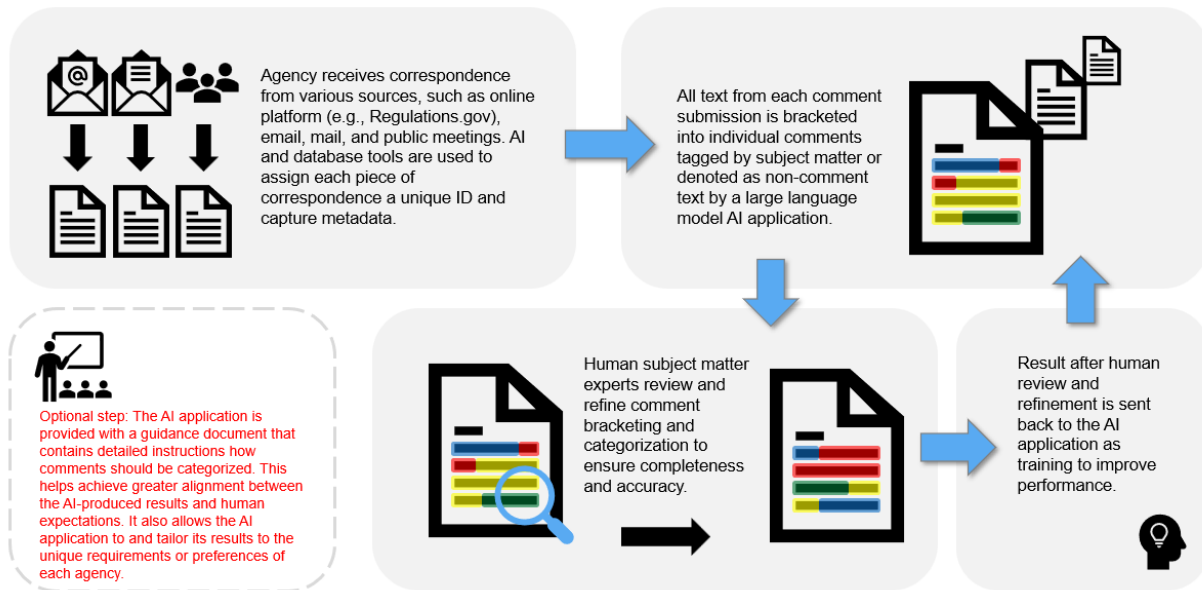


Figure 1. Conceptual AI-Assisted Workflow for Public Comment Review

## 2.0 Current Processes and Possible AI-Augmented Solutions to Processing Public Comments

In this section, we provide a representative human-driven process and some insight on how AI-augmented tools can expedite steps in that process.

### 2.1 Current Human-Driven Process

Agencies performing NEPA reviews must track and address comments from other agencies, industry, tribal nations, and the public. The current process, without using AI or LLMs, includes the following steps, which can be human- or database-driven. These tasks ensure that each piece of correspondence is tracked and that each comment in each piece of correspondence is bracketed to a specific subject matter area. A project manager and others trained in comment analysis typically perform these steps:

1. Assign each piece of correspondence a unique identifier (e.g., a document number).
2. Capture metadata (e.g., author name, date of submission, author email) for each piece of correspondence.
3. Conduct an initial review of all narrative content (including tables, footnotes, and images) and bracket and categorize portions of that content as comments or non-comments.
4. Assign each comment a unique identifier that is associated with the correspondence identifier assigned in Step 1.
5. Further categorize each comment by subject matter area.

## 2.2 AI Solutions

LLMs can understand context and, further, extract specific content from provided text (e.g., a PDF document). To that end, an AI-augmented tool should have the ability to identify and extract metadata, bracket and extract comments and non-comments, and provide an initial categorization of comments into subject matter categories. Further, a system that integrates AI with database functionality should be able to assign unique identifiers to each piece of correspondence and to each comment. Figure 2 shows where AI could integrate into the initial steps of comment processing.

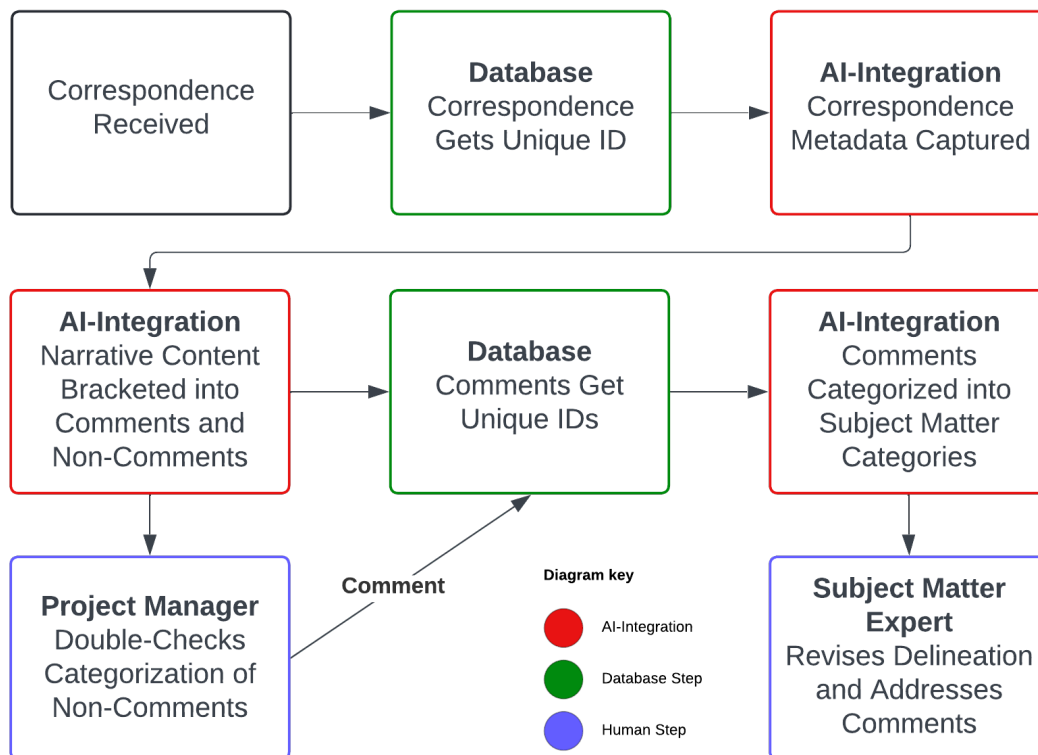


Figure 2. Possible AI Integration into Public Comment Processing

### 2.2.1 Prompting

To provide the context of PDF documents to the AI models, we leverage a technique called retrieval-augmented generation (RAG)<sup>1,2</sup>. In a RAG setup, a model takes as context a document or a set of documents and, using the documents as the reference and generates a response for the provided prompt. We started with zero-shot prompting of the RAG-based AI system to bracket and categorize exact quotes as comments from provided correspondence. In a zero-shot scenario, the AI model is trained to recognize and categorize content without prior exposure to examples. The prompt shown in Figure 3 provides NEPA-specific categories used to categorize, and later summarize, public comments received during a public scoping period.

<sup>1</sup> Gao, Yunfan, et al. "Retrieval-augmented generation for large language models: A survey." arXiv preprint arXiv:2312.10997 (2023).

<sup>2</sup> <https://research.ibm.com/blog/retrieval-augmented-generation-RAG>.

Prompt: This set of correspondence contains quotes that will be used to populate a NEPA scoping summary report. That report will be organized by resources areas (e.g., Environmental Justice, NEPA Policy, Air Quality). With that in mind, I need this correspondence to be broken into discrete delineated quotes by NEPA-relevant resource areas. All narrative content should be delineated, and if a quote does not fit any category, list that quote at the beginning under the heading "Not Applicable". The content should be formatted as a table where column 1 would have a relevant section title, column 2 the name of the commenter (if known), and column 3 would contain the quote from the correspondence that fits in that section.

The table should be sorted in the order comments appear in the correspondence.

The scoping summary report we're writing includes the following sections: B. Public Comments B.1 Comments on the Resource Areas B.1.1 Postulated Accidents B.1.2 Alternative - No-Action B.1.3 Alternatives - Replacement Power B.1.4 Environmental Justice B.1.5 Geologic Environment B.1.6 Greenhouse Gas Emissions and Climate Change B.1.7 Historic and Cultural Resources B.1.8 Human Health - Microbiological Hazard B.1.9 Human Health - Radiological B.1.10 Hydrology B.1.11 Meteorology and Air Quality B.1.12 Socioeconomics B.1.13 Uranium Fuel Cycle B.1.14 Waste Management B.2 Comments that are Peripheral or are not Significant (10 CFR XX.29(a)(3)) B.2.1 General Comment: comments general in nature that do not convey significant information for environmental review. B.2.2 Review Process NEPA, 10 CFR XX, and 10 CFR XX B.2.3 Emergency Preparedness B.2.4 Peripheral or Outside of Scope Miscellaneous B.2.5 Aging Management B.2.6 Safety B.2.7 Security and Terrorism B.2.8 Energy Costs B.2.9 Need for Power

Figure 3. Example Prompt Used to Test the Ability of LLMs to Bracket and Categorize Public Comments

Table 1 shows a favorable response to a prompt similar to the one in Figure 3 used on a short piece of correspondence. The response is bracketed into discrete quotes, categorized by NEPA-relevant resource areas, and formatted as a table with the relevant categorization in column 1, the name of the commenter in column 2, and the quote in column 3. The table is sorted by the order the comments appear in the correspondence.

Table 1. Favorable LLM Response to a Comment Bracketing and Categorization Prompt

| Section Title                                       | Commenter <sup>(a)</sup> | Quote <sup>(a)</sup>  |
|---|--------------------------|---|
| B.1.14 Waste Management                             | John Doe                 | I am opposed to furthering any useful development activity of material until methods are found that are sound, safe, and a definitive solution in dealing with the residue that might result from its use. Neglecting this issue is not near insanity; it is insanity.                |
| B.1.2 Alternative – No-Action                       | John Doe                 | I urge you to reject the application. To protect public health, safety, security, and financial health, the project should retire on or before 2080.  |
| B.2.2 Review Process NEPA, 10 CFR XX, and 10 CFR XX | John Doe                 | I request that in-person scoping meetings be held, as originally planned, when Covid risks subside. I request that a hearing opportunity be allowed for those seeking to intervene and that the deadline for requesting a hearing be postponed for 90 days past the current deadline. |
| B.2.2 Review Process NEPA, 10 CFR XX, and 10 CFR XX | John Doe                 | Both an afternoon and evening public meeting were scheduled, but then postponed at the last moment, without adequate public notice. Four hours of in-person meeting time were reduced to only two hours online.   |

(a) content has been manually sanitized and redacted.

Various prompts have been developed to test the performance of AI against manually bracketed and categorized correspondence. The following notes detail one instance of success and one of failure. In each case, the prompt in Figure 4 was used on a short piece of correspondence (i.e., less than three pages).



Prompt: This set of correspondence contains quotes that will be used to populate a NEPA scoping summary report. That report will be organized by resource areas (e.g., Environmental Justice, NEPA Policy, Air Quality). With that in mind, I need this correspondence to be broken into discrete delineated quotes by NEPA-relevant resource areas. If a quote does not fit any category, list that quote at the beginning under the heading "Not Applicable". The content should be formatted as a table where column 1 would have a relevant section title and column 2 would contain the quote from the correspondence that fits in that section.

Figure 4. Alternate Prompt Used to Bracket and Categorize Correspondence

### **Successful Bracketing and Categorization**

Several different LLMs (i.e., GPT-4 Turbo Preview, Claude 3-Sonnet, Claude-3 Haiku, Google Gemini-1.5, Claude-2, and Llama-2) were able to use the prompt in Figure 4 to match the human-bracketing and categorization with high accuracy. This included identifying those comments which were specific to the NEPA process, and which were related to hydrology, geology, and aging management. What the models did miss (and the prompt did not address) was the ability to recognize subject matter categories that fall outside the scope of the NEPA process or scope of project, which can be nuanced based on project specifics.

Table 2. Example of Successful Bracketing and Categorization

| Section                  | Quote <sup>(a)</sup>  |
|--------------------------|---|
| Public Meetings/Hearings | "The Date XX, 20XX in-person meeting was canceled in favor of an online meeting, therefore, we are requesting an in person in Redacted County so residents and other concerned citizens can directly interact with agency staff. There were multiple technical issues with the presentation and many speakers were not able to unmute and speak." |
| Public Meetings/Hearings | "We request a formal hearing due to the aging nature of the plant for safety reasons."  |
| Comment Period Extension | "We request an extension for comment period for the XX/XX/20XX Federal Register Notice."  |
| Safety – Plant Aging     | "There are too many safety issues concerning cracks in the foundation, cracks in pipes, the dome, the spillway, and "loss of material," on the dam."  |

(a) content has been manually sanitized and redacted.

### **Unsuccessful Bracketing and Categorization**

All available LLMs failed to identify the narrative content of another document, resulting in all bracketed content containing metadata or non-relevant sections of the correspondence.

Table 3. Example of Unsuccessful Bracketing and Categorization

| Category                 | Verbatim Quote <sup>(a)</sup>   |
|--------------------------|---|
| Contact Information      | "Please send our office an electronic copy of the Draft EIS when it is electronically filed with the Office of Federal Activities using the following link: <a href="https://www.agency.gov/nepa/e-nepa-guide-registration-and-preparing-eis-electronic-submission">https://www.agency.gov/nepa/e-nepa-guide-registration-and-preparing-eis-electronic-submission</a> . If you have any questions, please contact.... |
| Signatory                | "Digitally signed by ..... Date: 2023.01.23   |
| Email Correspondence     | "From: Doe, John John.Doe@xxxx.gov Sent: Tuesday, January 24, 2023 9:14 AM To: XXXXXXXXXX Resource Cc: Jack Doe; Jason Doe Subject: [External_Sender] Agency Scoping Comment Letter for the Power Plant Attachments: Agency Scoping Comments for Power Plant.pdf"   |
| Mail Envelope Properties | "Subject: [External_Sender] Agency Scoping Comment Letter for the Power Plant Sent Date: 1/24/2023 9:13:51 AM Received Date: 1/24/2023 9:14:13 AM From: Doe, John Created By: John.Doe@agency.gov Recipient": "Jack Doe" Jack.Doe@agency.gov "Jason Doe" Doe.Jason@agency.gov "EIS Resource" EIS.Resource@agency.gov  |

(a) content has been manually sanitized and redacted.

In this case, the LLM was unsuccessful in matching the human-bracketing and categorization with high accuracy despite the similarity of the correspondence document to others that were successfully processed. One potential explanation is that the RAG mechanism is sensitive to

noise present in the data. For example, correspondence letters contain various amounts of text unrelated to the narrative content (e.g., contact information, personally identifiable information, and other administrative document identification). We are exploring ways to augment the data curation and fine-tune the models to ensure the semantics of the documents are better captured in the embedding space.

### 2.2.2 Instruction Tuning

To better align the LLM’s reasoning ability with context corresponding to NEPA comments, we fine-tuned an LLM to identify whether the content in each correspondence document are comments corresponding to a subject matter category or not. We used sentences containing comments in three sets of documents: Set 1, Set 2, and Set 3. All these sentences were separated into training, validation, and testing folds using the stratified splitting technique. During training, each sentence in a correspondence document was fed to the LLM using the prompt template in Figure 5. Then, the trained model was used to generate a response, given the input text.

|   |  |
|---|--|
| <b>Instruction:</b>   |  |
| Scoping is the gathering and analysis of information that a state agency will use to establish the breadth, or scope, of environmental review of a proposed project. The public can submit scoping comments in writing to the contact person listed in the Scoping Notice. Identify whether the given sentence is a scoping comment or not. |  |
| <b>Input:</b>   |  |
| <The sentence to be classified>   |  |
| <b>Response:</b>  |  |

Figure 5. Prompt Template Used for Instruction Tuning

We varied the training dataset size and composition of training and testing folds. In one setting, inference data were selected exclusively from Set 3 documents, and the training data comprised sentences from Set 1 and Set 2 documents (orange curve in Figure 6). In the other setting, we used some Set 3 sentences to train the model as well. We observed a significant increase in the accuracy when sentences similar to those in the inference set were considered to tune the LLM. Based on this observation we tuned a model using only the Set 2 text, which have a similar context to the Set 3 text. As shown in Figure 6(b) comment identification accuracy is similar to case where the training data consists of Set 1, Set 2, and Set 3 text.

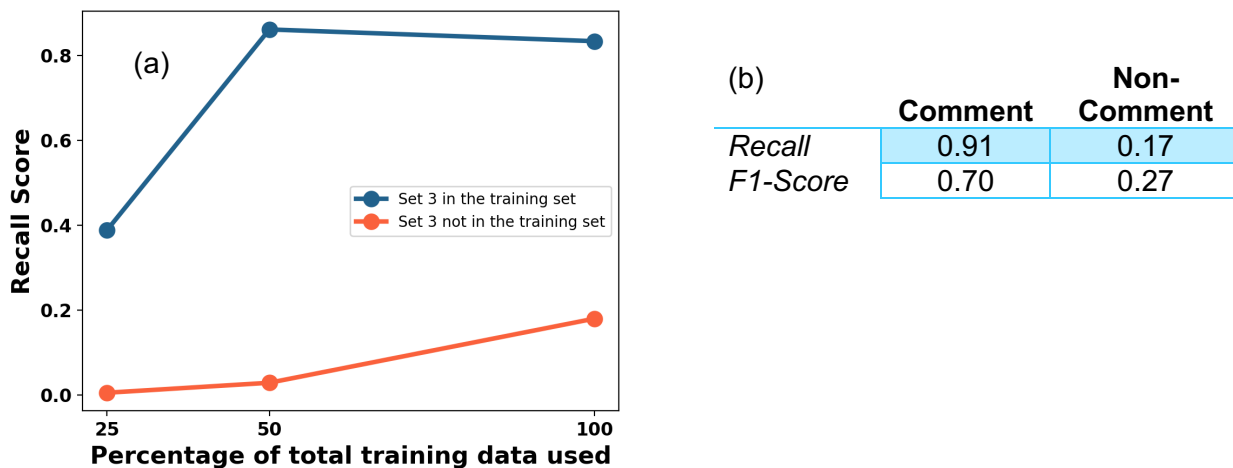


Figure 6. (a) Comment/Non-Comment Classification Accuracy and (b) Set 3 Sentence Classification Accuracies When Only Set 2 Documents Were Used to Train the LLM

These high accuracies enable these models to be effectively employed for comment bracketing (See Figure 7).

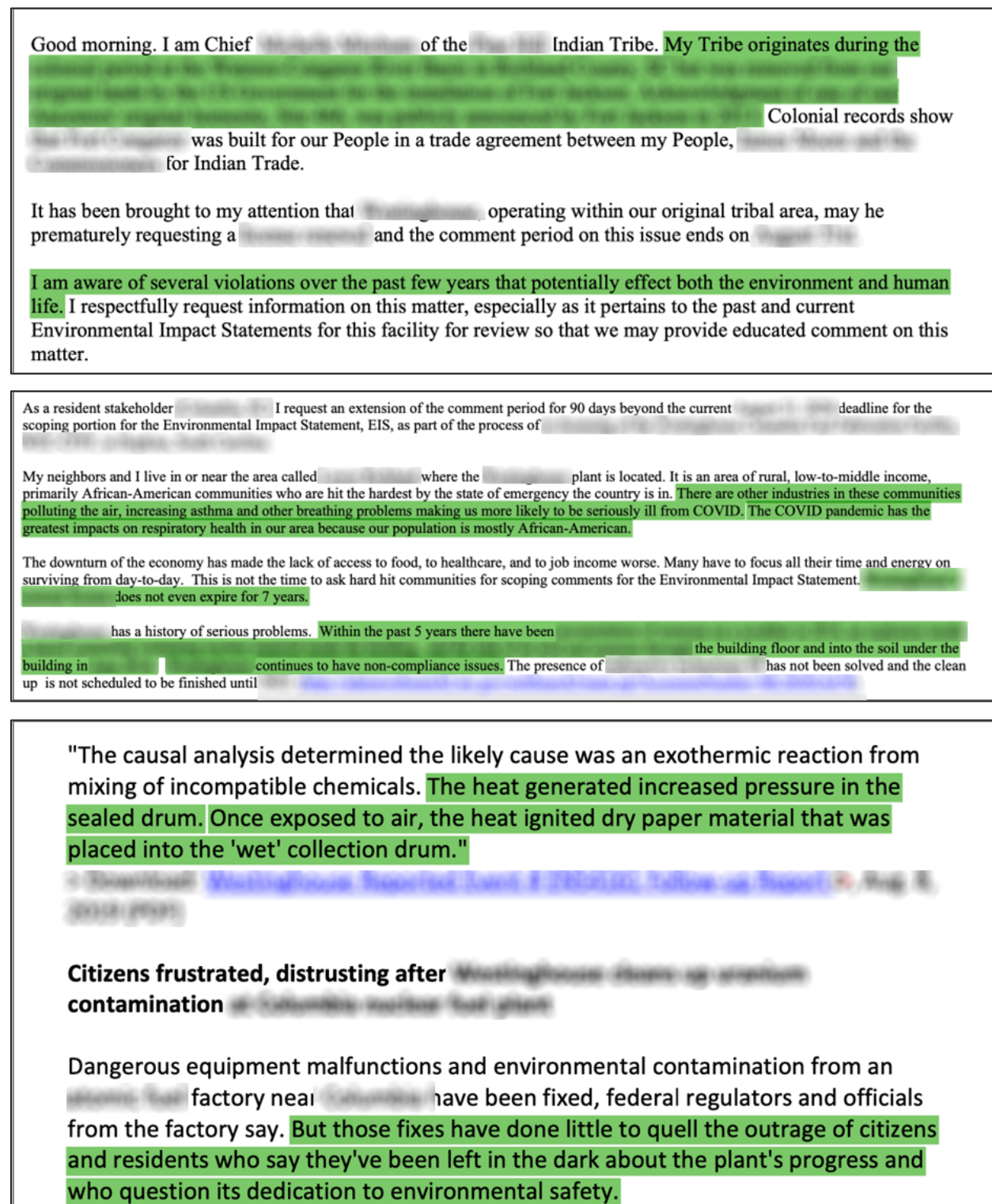


Figure 7. Examples of Comment Bracketing Using Instruction Tuning. The text has been sanitized and redacted.

## 3.0 Current Limitations and Next Steps

Based on initial approaches, AI-augmented solutions hold promise for bracketing and categorizing comments. As we proceed, we are working to mitigate current limitations and to create AI-augmented tools for NEPA workflows.

### 3.1 Current Limitations

There are several obstacles that must be overcome to enable effective AI integration.

1. The ability of AI to consistently bracket all narrative content for categorization. Failing to recognize narrative content means a relevant comment may not be discovered and addressed. This can be due to a particular file type or encoding, the length and complexity of a piece of correspondence, AI misinterpretation of context, or a combination of factors.
2. The ability of AI to provide initial categorization relevant to the specific NEPA review. Differences in the types of proposed actions, agency decision-making processes, agency requirements and preferred practices, relevant environmental issues, and myriad other factors mean that subject matter categories will vary by project. The AI integration must be flexible enough to accommodate agency-specific needs.

### 3.2 Next Steps

Our next step is to incorporate the ability to suggest the subject matter category for each identified comment. Figure 8 depicts the expected behavior of the application, where the AI agent brackets the comment text and provides suggested subject matter categories. The human comment processor can accept or reject the suggestions. In case the AI agent fails to identify a comment, the human expert can click on a sentence to mark it as a comment. These capabilities ensure full capturing of the comments. Additionally, this human expert's feedback will be used to refine models' responses continuously improving user experience in a human-AI synergistic manner.

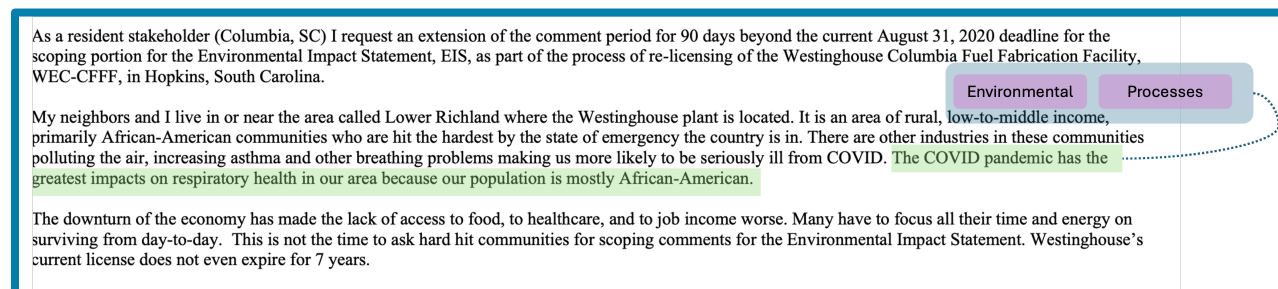


Figure 8. Example of Future Tool Used to Suggest Subject Matter Categories

# **Pacific Northwest National Laboratory**

902 Battelle Boulevard  
P.O. Box 999  
Richland, WA 99354  
1-888-375-PNNL (7665)

***[www.pnnl.gov](http://www.pnnl.gov)***