

## **Artificial Intelligence/Machine** Learning Approaches to **Improve Permitting Outcomes** and Efficiency

Dr. Sai Munikoti Co-PI, Policy AI



Pacific

PNNL is operated by Battelle for the U.S. Department of Energy





## **Overview**

- Generative AI in Environmental **Review and Permitting Process**
- **Technical Use Cases:** 
  - ✓ AI-Driven Document Search
  - AI Knowledge Assistant  $\checkmark$
- Case Study: Public Scoping **Comment Processing**
- Al Assurance: Towards Trustworthy and Explainable AI Systems
- Community Outreach Events: Kaggle NEPA Challenge



(PI)



Sameera Horawalavithana

Sai Munikoti (Co-PI)



Shivam Sharma







Anurag Acharya

Ann Miracle



Sridevi Wagle

Anastasia Bernat



Derek Lilienthal

Mike Parker







Phan Hung

**Gihan Panapitiya** 



**Dan Nally** 



Mahantesh Halappanavar



Sarthak Chaturvedi



Karl Pazdernik





# Final decision document







## **Review/Permitting** process



Domain Expert



# **Al-Driven Document Search**



- No centralized database
- Search only based on titles

- Search based on entire body content
- Get important fields along with search results
- Accompanied vector database seamlessly connects with AI copilots

# Unstructured Not AI ready



# **Al-Driven Document Search**

### PolicyAI NEPA Search V0.01 **PolicyAI NEPA Search V0.01** Radioactive waste X Press enter to add more Radioactive waste X Nuclear Fuel Management X Press enter to add more Search Search **Total Documents: 174 Total Documents: 13** 1. Hanford Site Tank Closure and Waste Management Project Richland Benton County WA 1. Versatile Test Reactor Environmental Impact Statement See Document See Document 2. Remediation of Area IV and the Northern Buffer Zone of the Santa Susana Field Laboratory 2. Construction and Demonstration of a Prototype Mobile Microreactor See Document See Document Document Status: Final Document Status: Final EPA Comment Letter Date: 02/04/2019 EPA Comment Letter Date: None Federal Register Date: 12/28/2018 Federal Register Date: 02/25/2022 Agency: Department of Energy Agency: Department of Defense State: ca ; california State: id ; idaho #Search Hits (Pages)/Total #Pages: 54/156 #Search Hits (Pages)/Total #Pages: 33/338

## Keyword/Concept/Phrase search

Compound search



Multi-Document

Search Docs

Summary

# **AI-Driven Search and Knowledge Assistant**

## **PolicyAI NEPA Search V0.01**

Radioactive Waste X Nuclear Fuel Management X Press enter to add more

Search

### **Total Documents: 7**

The search results provide information on the management of radioactive waste and nuclear fuel at the Savannah River Site, Los Alamos National Laboratory, and the Browns Ferry and Sequoyah Nuclear Plants:

At the Savannah River Site, the alternatives considered would generate various types of radioactive waste, including transuranic waste, low-level radioactive waste, mixed low-level radioactive waste, and radioactive TSCA (PCB) and asbestos waste. The impacts on waste management from the different alternatives are discussed, with the new facility alternative generally having the largest impacts from increased waste generation. The search results also cover the management of spent nuclear fuel and the impacts on naval spent nuclear fuel management. [1][2]At Los Alamos National Laboratory, the search results indicate that the site generates transuranic waste, low-level radioactive waste, and mixed low-level radioactive waste. The impacts on waste management from the different alternatives are discussed.[3]For the Browns Ferry and Sequoyah Nuclear Plants, the search results cover the management of used nuclear fuel assemblies generated from the use of mixed oxide fuel. The impacts on waste management from the use of mixed oxide fuel compared to low-enriched uranium fuel are discussed.[4]

## → 1. <u>Versatile Test Reactor Environmental Impact Statement</u>

### See Document

Document Status: Final

EPA Comment Letter Date: 06/21/2022

Federal Register Date: 05/20/2022

Agency: Department of Energy

Citation

**Fields** 



# Case Study: Public Scoping Comment Processing





## Comment Processing Existing Workflow

Agency receives correspondence from various sources, such as online platform (e.g., Regulations.gov), email, mail, and public meetings.





All text from each comment doc is manually bracketed into individual comments, tagged by subject matter category.



## Tedious

## Time/Resource intensive



## Summary report Draft EIS report



# **Comment Processing**

Human + AI workflow



Agency receives correspondence from various sources, such as online platform (e.g., Regulations.gov), email, mail, and public meetings.





Al bracket text from each comment doc into individual comments and categorize into subject matter categories.

Human experts review and refine comment bracketing and categorization to ensure completeness and accuracy



Result after human review and refinement is sent back to the AI model for training to improve performance



### Al powered by environmental review/ permitting knowledge corpus





## **Comment Identification/Categorization** Prompting

PROMPT Question + Instruction + Role + Examples interaction to Frozen Al model PDF Prompt

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.



## Idea of Prompting: manipulate your accomplish the task



## **Comment Identification/Categorization** Prompting

Successful	Section	Quote <sup>(a)</sup>
Instance Specific NEPA terms related to hydrology, geology, and aging.	Public Meetings/Hearings	"The Date XX, 20XX in-person meeting was canceled in favor of a are requesting an in person in Redacted County so residents and directly interact with agency staff. There were multiple technical is many speakers were not able to unmute and speak."
		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
	<b>Comment Period Extension</b>	"We request an extension for comment period for the XX/XX/20X
	Safety – Plant Aging	"There are too many safety issues concerning cracks in the found the spillway, and "loss of material," on the dam."
	(a) content has been manually	sanitized and redacted.

Unsuccessful	Category	Verbatim Quote <sup>(a)</sup>
instance containing metadata or non- relevant sections	Contact Information	"Please send our office an electronic copy of the Draft EIS when Office of Federal Activities using the following link: https://www.ag registration-and-preparing-eis-electronic-submission. If you have
	Signatory	"Digitally signed by Date: 2023.01.23
	Email Correspondence	"From: Doe, John John.Doe@xxxx.gov Sent: Tuesday, January 2 XXXXXXXX Resource Cc: Jack Doe; Jason Doe Subject: [Exter Comment Letter for the Power Plant Attachments: Agency Scopin Plant.pdf"



an online meeting, therefore, we other concerned citizens can ssues with the presentation and

for safety reasons." X Federal Register Notice." dation, cracks in pipes, the dome,

it is electronically filed with the gency.gov/nepa/e-nepa-guideany questions, please contact....

24, 2023 9:14 AM To: ernal\_Sender] Agency Scoping ng Comments for Power



# **Comment Identification/Categorization**

**Prompting Demo** 

## Scoping Copilot #1 (v0.015)

cooping copilor = (release)	00	Q ~ ¥ 1 014	- +
Select LLM Model			
openal.gpt-4-turbo-preview	© ~		
Search by Projects			
	*		
Search by Titles			
	~		
<u>6</u>			
Ask me a question about NEPA's documents!		Migratory Bird Permit	: Authorizing
0		Dear	
what are the concerns raised regarding migratory birds in this comment letter ?		Register on ( The response to advance to adv	rule authorizi
		The mission of the wildlife, parks, and recreational re-	itering and st sources of Mo
The concerns mentioned in the comment letter about the impact of the proposed construction project on	migratory birds include		
the potential for one-time losses of migratory birds and temporary impacts to their habitats. The letter sug specific permits should be used to address these impacts. Additionally, there is a recognition of the need for	gests that project- or habitat improvement		
Your question			







# **Comment Identification/Categorization**

Instruction Finetuning



**Idea of Instruction Tuning:** Adapt (Train) you model to follow instruction

## **Objective: Comment identification**

Data: Correspondence from three projects

Task: Sentence is comment or not

Label: Sentence level human annotation

AI Model: Llama 7B, Mistral 7B





# **Comment Identification/Categorization**

Instruction Finetuning

### Instruction:

Prompt

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.

Input: <The sentence to be classified> Response:

## Inference

- Acceptable Comment/Non-Comment identification accuracy (>85%)
- Generalize to new projects
- Require certain level of similarity between train and test documents





## **Comment Summarization** Human Oversight

The AI application prepares an extractive summary of all comments within each category to concisely convey the key concerns, questions, and information provided by the public.

matter experts summaries for accuracy and completeness.

Result after human review and refinement is sent back to the AI application as training to improve

performance.





# **Multi-Document Comment Summarization**

## Example prompt for summarizing scoping comments

The prompt requests that the AI application interpret and compose a concise summary of public comments as categorized in the first processing step for inclusion in a scoping summary report. The table below categorizes public scoping comments with the scoping summary report section name in column 1, comments in column 2, and the comment number in column 3. For each section, prepare a concise narrative summary of the key concerns, questions, and information conveyed in the comments.

The content should be formatted as a report, with sections listed as headings in the order they appear in the table.

Section Title	Comment	Com
B.1.1 Postulated Accidents	Severe accidents must not be discounted simply because agency assumptions deem them unlikely. Plausible worst case scenarios mandate evaluation and transparent disclosure to the public.	38-17
B.1.2 Alternative – No- Action	We would also ask the agency to consider alternatives including no renewal, renewal and a renewal of less time.	37-7
Etc.	Etc.	Etc.

ment Number



# **Multi-Document Comment Summarization**

## Example response to prompt from previous slide

The response shows scoping summary report sections drafted by the AI application.

Here is the narrative summary of the public scoping comments formatted as a report.

B.1 Comments on the Resource Areas

**B.1.1 Postulated Accidents** 

Comment Summary (Accidents – SAMAs): Commenters expressed concerns regarding accidents at the energy plant. One commenter requested that the plausible worst-case scenarios be disclosed to the public. Another commenter expressed concern about accidents releasing airborne toxins.

Comments: (38-17) (76-2)

B.1.2 Alternative – No – Action

Comment Summary: A commenter requested that the agency consider not renewing the operating license or consider a shorter period for operating license renewal for the energy plant: (37-7) (37-9)

Comment Summary (No - Action – Decommissioning): Commenters expressed concerns about decommissioning including the impacts of decommissioning, particularly the economic impacts, costs for decommissioning, and waste management.

Comments: (29-13-4) (38-25) (51-2)...



# **Comment Processing: Challenges & Next steps**

## Challenges

- ✓ Variation across doc types/complexity, AI misinterpretation of context
- Subjective comment categories due to different agency requirements and preferences, type of proposed actions, etc.

## Ongoing work

- ✓ Incorporate comment category identification model that accommodate agency specific needs
- ✓ Incorporation of advanced AI techniques : Multiple AI agents (Crew-AI)

My neighbors and I live in a rural area which is hit...

As a resident stakeholder I request an extension of the comment period for 90 days beyond the current deadline. The downturn of the economy has made the lack of access to food, to healthcare, and to job income worse. This is not the time to ask hard hit communities for scoping comments on the environmental impact statement.



Example of Future Tool: AI agent brackets the comment text and provides suggested subject matter categories

**NEPA Process** 

**Human Health** 



## Al Assurance: Towards Trustworthy and Explainable Al Systems







## Evaluate copilots for their domain Knowledge

## Approach

- Generate synthetic question-answers from NEPA documents using human-in-the loop method
- Test Frontier LLMs (GPT-4/Claude Sonnet/Gemini 1.5 Pro) on these questions
- Evaluate the performance of these models on this task under various experimental conditions



### **Data Sources**



## **Human-in-the-Loop Evaluation**



### 1. Gold Passages Selection

Experts choose portions of EIS documents containing high-quality content across beginning, middle, and end sections



### 2. Prompt Design

Experts design prompt templates to generate different types of questions from the gold passages

## NEPAQUaD1.0

1548 QAs 10 Question types





## 3. Question-Answer Generation

Use LLM to generate several question-answers based on the designed prompts





## **Human-in-the-Loop Evaluation**



NEPAQuAD1.0





### 4. Response Generation

Questions, alongside relevant context, are passed to the LLMs to generate response. Varying types of context are used.

5
Ċ
С
tr
е

Context	Input Prompt to AI model
None	Question (Zero context)
PDF	Question + corresponding PDF (entire document) from where Question is generated
Gold Passages	Question + corresponding passage from where Question is generated
Silver Passages	Question + passages retrieved based on semantic similarity (RAG)

## RAG is the preferred way to fuse NEPA knowledge into AI model

### 5.Response Evaluation

Generated responses are compared with ground ruth answer using to evaluate performance.



# **Community Outreach**



Kaggle competition: "LLM for Environmental Review"

Dataset: NEPAQuAD1.0

Task: Develop LLM for Question Answering

Start Date: May 15, 2024.

End Date: June 30, 2024









# Thank you



