NEPA Project Analysis: Decarbonization Technology Environmental Reviews

National Environmental Policy Act Text Corpus (NEPATEC) 2.0 Analysis

Your Name

2026-05-14

Table of contents

1 Project Overview

This document is a good reference for how to contextualize this project, the project goals and data, and to clearly understand the project deliverables and timeline.

1.1 Project context: National Environmental Policy Act (NEPA)

Federal environmental permitting—and, relatedly, the National Environmental Policy Act (NEPA)—has often been blamed as a core contributor to delays in infrastructure deployment. However, research to date on NEPA has been hindered by federal agency data management practices: NEPA documents are scattered across numerous agency databases, often in machine-unreadable formats, and typically lack basic metadata and other identifiers. As such, researchers have had to craft bespoke subsets of NEPA data from which to glean insights—a time-consuming task that has limited the information available about NEPA's effectiveness and paved the way for a national permitting reform conversation rife with anecdotes and cherry-picked information.

The newly released National Environmental Policy Act Text Corpus (NEPATEC) 2.0 dataset from the Pacific Northwest National Laboratory's (PNNL's) PermitAI project has the potential to add new, more comprehensive evidence into this conversation. PNNL has built and released a comprehensive dataset of past environmental reviews and permitting documents containing millions of pages and billions of words in machine-readable JSON format. The database contains more than 120,000 NEPA documents[1] from 60,000 projects prepared by more than 60 different agencies. Each document contains metadata for (as applicable):

- Lead agency
- Category
- Type of review
- Name of project
- Location
- Project sponsor
- Project sector
- Project type (a subset of project sector)
- Type of document
- Document title
- Agency or contractor responsible for preparing the document
- Categorical exclusion category
- Summary of the proposed action

> **Note**
>
> The National Environmental Policy Act Text Corpus (NEPATEC) 2.0 can be accessed here on huggingface.

[1] categorical exclusions, draft and final environmental assessments, draft and final environmental impact statements, records of decision, findings of no significant impact, and other supporting documentation

1.2 Data structure

This shows the original data structure of the project in its original json format. The code in this repo manipulates this raw data.


```
{
    "project": {
        "project_ID": "UNIQUE PROJECT ID FOR PUBLIC VERSION",
        "project_title": {
            "value": ""
        },
        "project_sector": {
            "value": ""
        },
        "project_type": {
            "value": ""
        },
        "project_description": {
            "value": ""
        },
        "project_sponsor": {
            "value": ""
        },
        "location": {
            "value": ""
        }
    },
    "process": {
        "process_family": {
            "value": ""
        },
        "process_type": {
            "value": ""
        },
        "lead_agency": {
            "value": ""
        }
    },
    "documents": [
        {
            "metadata": {
                "document_metadata":{
```

```
            "document_ID": {
                "value": "UNIQUE DOC/FILE ID FOR PUBLIC VERSION"
            },
            "document_type": {
                "value": ""
            },
            "document_title": {
                "value": ""
            },
            "prepared_by": {
                "value": ""
            },
            "ce_category": {
                "value": ""
            }
        },
        "file_metadata":{
            "file_ID": {
                "value": "UNIQUE DOC/FILE ID FOR PUBLIC VERSION"
            },
            "file_name": {
                "value": "PDF NAME"
            },
            "section_or_volume_title": {
                "value": ""
            },
            "main_document": {
                "value": ""
            },
            "total_pages": {
                "value": ""
            },
            "file_provider": {
                "value": ""
            }
        }
    },
    "pages": [
        {
            "page number": 1,
            "page text": "PAGE 1 TEXT"
```

```
                },
                {
                    "page number": 2,
                    "page text": "PAGE 2 TEXT"
                }
            ]
        }
    ]
}
```


2 Project deliverables

For this project, we want to create tables, figures, and maps that help us learn about the data and answer the following questions:

2.1 Phase 1

Table 1: Phase 1 Deliverable Timeline

Deliverable	Due_Date
1. Decarbonization Technology Projects	Jan 23, 2026
2. Programmatic Reviews	Feb 6, 2026
3. CE vs EA vs EIS	Jan 23, 2026
4. Geography	Feb 6, 2026
5. Pages Over Time	Feb 27, 2026
6. Technology-Specific	Feb 27, 2026

1. **Data on number of decarbonization technology projects within the dataset**: number of projects broken down by technology (e.g., offshore and onshore wind, solar, geothermal, nuclear), lead agency, and location
2. **Data on programmatic and tiered reviews**: how many tiered reviews are there compared to total and are they completed faster
3. **Data on how many decarbonization technology projects have been categorically excluded vs. have required environmental assessments and environmental impact statements**
    - Broken out by number of projects, generation capacity, and change over time

4. **Data on geography/project location**: whether projects are multi-state or multi-agency
5. **Number of pages over time**, including pre- and post- Fiscal Responsibility Act of 2023 (FRA), which set page limit requirements
6. **Technology-specific inquiries**
    - Transmission lines: length of lines from project summary correlated with timelines, location, etc.
    - Geothermal: timelines of environmental reviews for different phases of the same project
    - Carbon and hydrogen pipelines (if available): length of pipelines from project summary correlated with timelines, location, etc.; compare to natural gas pipelines

2.2 Phase 2

Phase 2 of the project is still not settled and we'll need to scope it based on what we find and do in Phase 1. Nevertheless, these are the envisioned deliverables:

1. **Reasons why NEPA was triggered** (e.g., federal land, federal funding) for different types of projects
2. **Determinations of significance across resource areas**; factors that contribute to a determination of "significant impact"
    - Starting with mitigated FONSIs
3. **Differences and similarities between NEPA reviews** for fossil fuel and decarbonization technology projects, as well as linear and non-linear projects—application of categorical exclusions, timelines, geography, etc.
4. **Timelines for categorical exclusions, environmental assessments, and environmental impact statements**, including segmentation by years (e.g., pre- and post-FRA [which set timelines for reviews], different CEQ NEPA regulations, agency, and type of project)
    - Timeline outliers could then be investigated through a case study approach to identify contributing factors, including whether NEPA was a cause of delay or not
    - May need to cross-reference with the Notice of Intents in the Federal Register using their API to get the start date
5. **Technical support for new regulatory categorical exclusion development**: identifying patterns in FONSIs

3 Project deliverable timelines

Table 2: Project Timeline and Deliverables

Meeting	Date	Deliverables
Kickoff	Jan 9, 2026	(0) Build database
1	Jan 23, 2026	(1) Decarbonization technology projects(3) CE vs EA vs EIS
2	Feb 6, 2026	(2) Reviews(4) Geography
3	Feb 27, 2026	(5) Pages(6) Technology
4	Mar 6, 2026	Present all findings

4 Project Structure

This project is organized into the following directories:

- **`code/`**: Python scripts for data processing and analysis pipelines
- **`data/`**: Raw and processed data files (parquet format) organized by review type (EA, EIS, CE)
- **`notebooks/`**: Jupyter notebooks for exploratory analysis
- **`reports/`**: Generated reports and deliverable documents
- **`output/`**: Analysis outputs including tables, figures, and visualizations
- **`notes/`**: Internal project documentation and working notes
- **`literature/`**: Reference materials and background documents

5 Definitions: What Qualifies as Decarbonization Technology?

5.1 Background

This section defines the universe of decarbonization technology projects analyzed across all deliverables of the NEPA Decarbonization Technology Analysis. It describes the classification criteria, exclusion rules, and refinements applied to identify decarbonization technology projects within the publicly released NEPATEC 2.0 database — a comprehensive record of federal environmental review activity under the National Environmental Policy Act (NEPA).

Understanding this classification framework is foundational: every project count, timeline, agency breakdown, and geographic pattern reported in subsequent deliverables flows from the decisions documented here.

5.2 The NEPA Universe

Figure 1 gives a sense of the total number of projects in the NEPATEC 2.0 database by NEPA review process type.

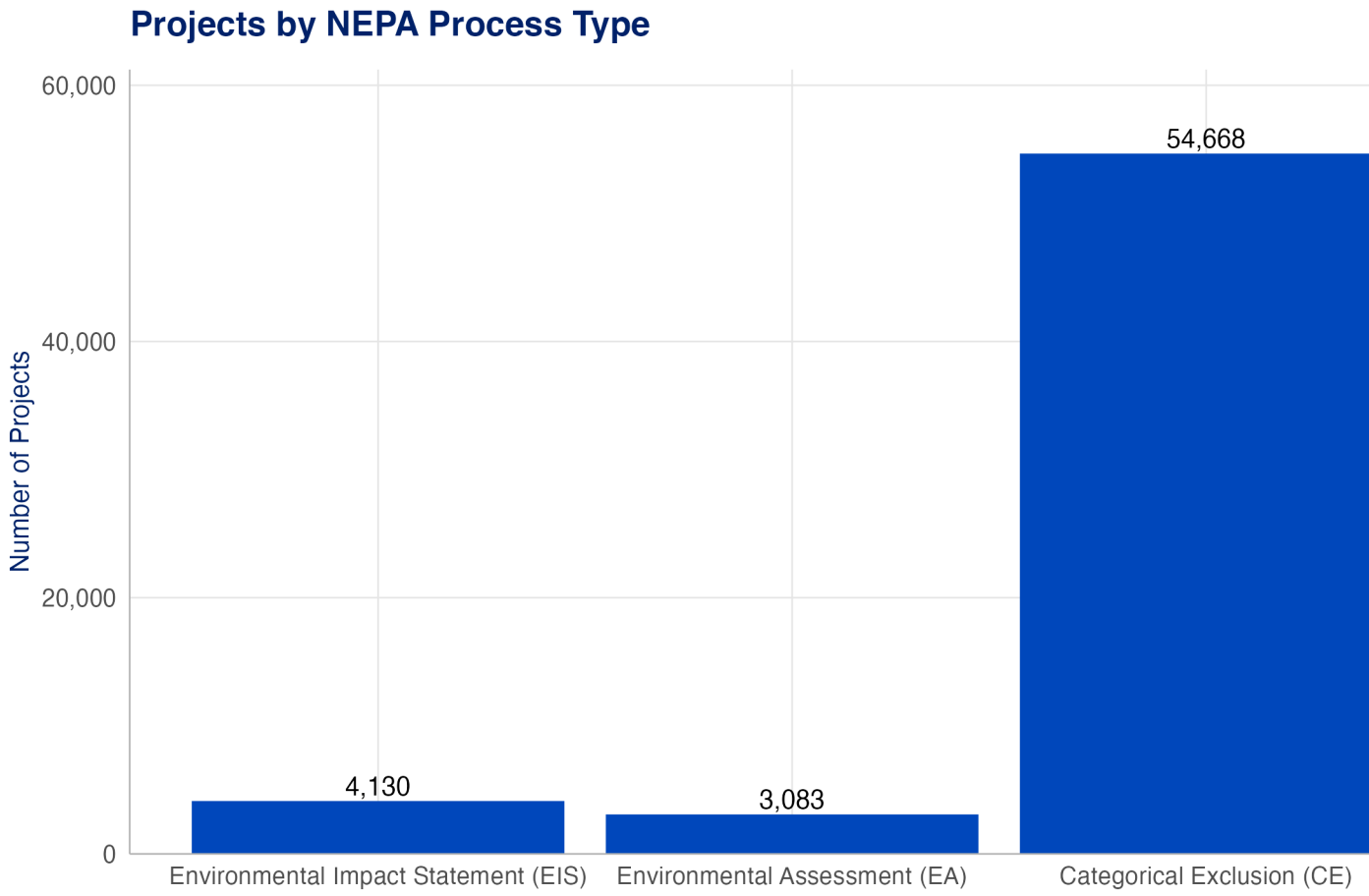

Figure 1: Distribution of all projects by review process in the NEPA database.

Figure 2 shows the decarbonization technology subset, totaling approximately **25,000 decarbonization technology projects** extracted from the NEPA database.

5.3 What Qualifies as Decarbonization Technology?

Table 3 enumerates the CATF-defined 14 decarbonization technology and 5 fossil fuel categories used to identify decarbonization technology vs fossil fuel energy projects in the NEPA database.

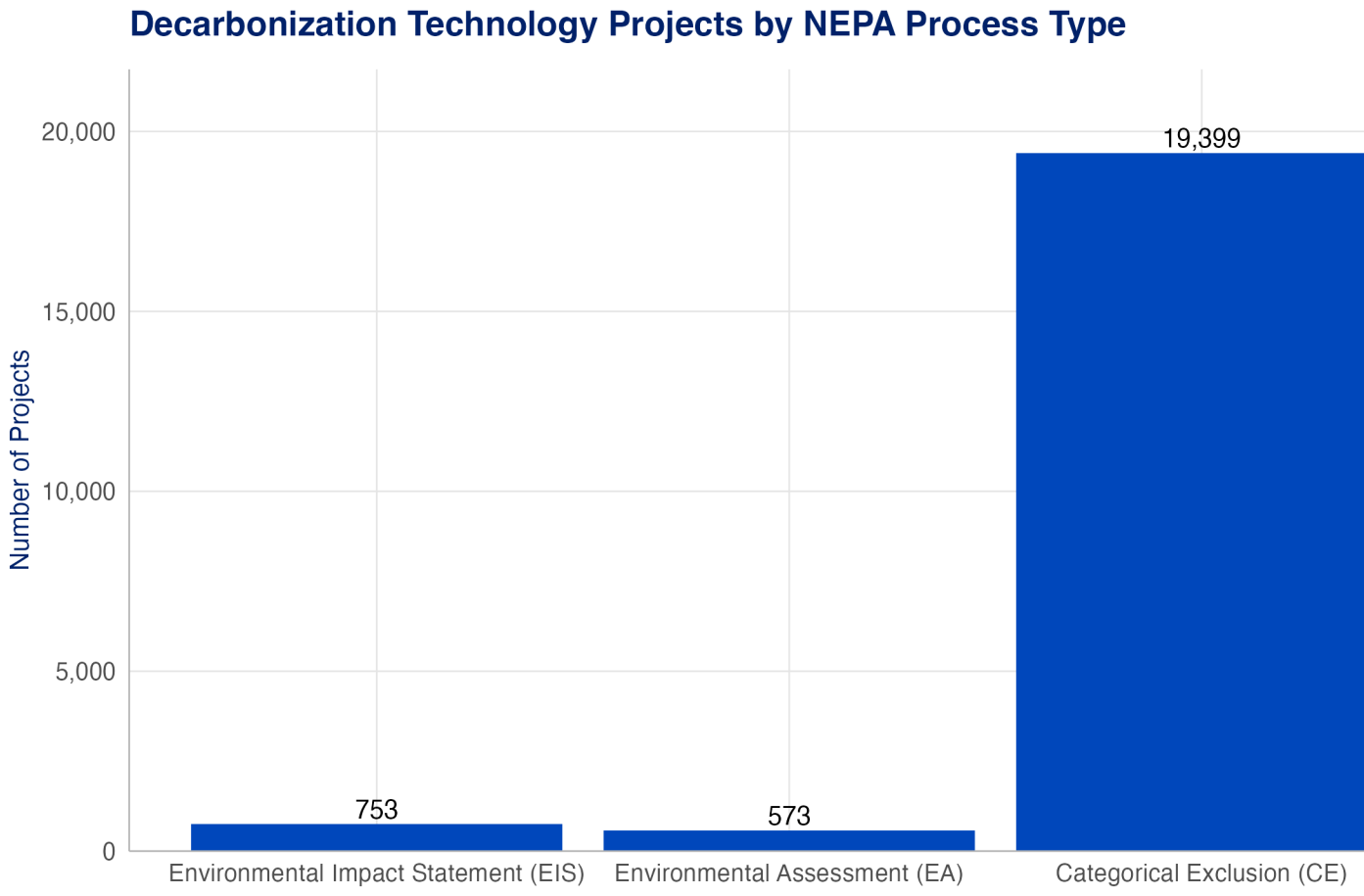

Figure 2: Distribution of decarbonization technology projects by review process in the NEPA database.

While projects often have multiple tags, a "decarbonization technology" project is identified as having at least one of the 14 tags *AND* no fossil fuel tags.

Table 3: Project type tags used to classify clean vs. fossil energy projects

Decarbonization Technology Tags	Fossil Energy Tags
Carbon Capture and Sequestration	Conventional Energy Production - Coal
Conventional Energy Production - Nuclear	Conventional Energy Production - Land-based Oil & Gas
Conventional Energy Production - Other	Conventional Energy Production - Offshore Oil and Gas
Electricity Transmission	Conventional Energy Production - Rural Energy
Nuclear Technology	Pipelines
Renewable Energy Production*[1]	
Utilities (electricity, gas, telecommunications)	

[1]Includes Biomass, Energy Storage, Geothermal, Hydrokinetic, Hydropower, Solar, Wind (Offshore & Onshore), and Other

Figure 3 shows the overall distribution of projects in the NEPATEC 2.0 database by energy type. Decarbonization technology projects comprise roughly 37% of all projects in the database, while fossil fuel projects represent 18%, and other (non-energy) projects make up 45% of the total. This breakdown helps contextualize the decarbonization technology subset analyzed in subsequent deliverables within the broader universe of federal environmental reviews.

About **10% of projects have only decarbonization technology tags**, but most (about 90%) have at least 1 decarbonization technology tag plus some other combination of tags.

5.3.1 Refining Decarbonization Technology

After reviewing a table of all co-occurring project types, three refinements were applied to improve the precision of the "decarbonization technology" classification:

- **Utilities + non-energy: 1,623 projects** tagged *ONLY* as Utilities *AND* with 1 or more of the following non-energy tags (e.g., broadband, waste management, land development) were excluded, since these likely reflect utility-adjacent infrastructure rather than energy generation.
- **Military and Defense + Nuclear: 481 projects** tagged as "Nuclear Energy" and "Military and Defense" were excluded from the decarbonization technology category. A full table can be viewed here. The majority of these projects are led by the Department of Energy rather than the Department of Defense, suggesting they involve weapons-related nuclear activities rather than energy production.

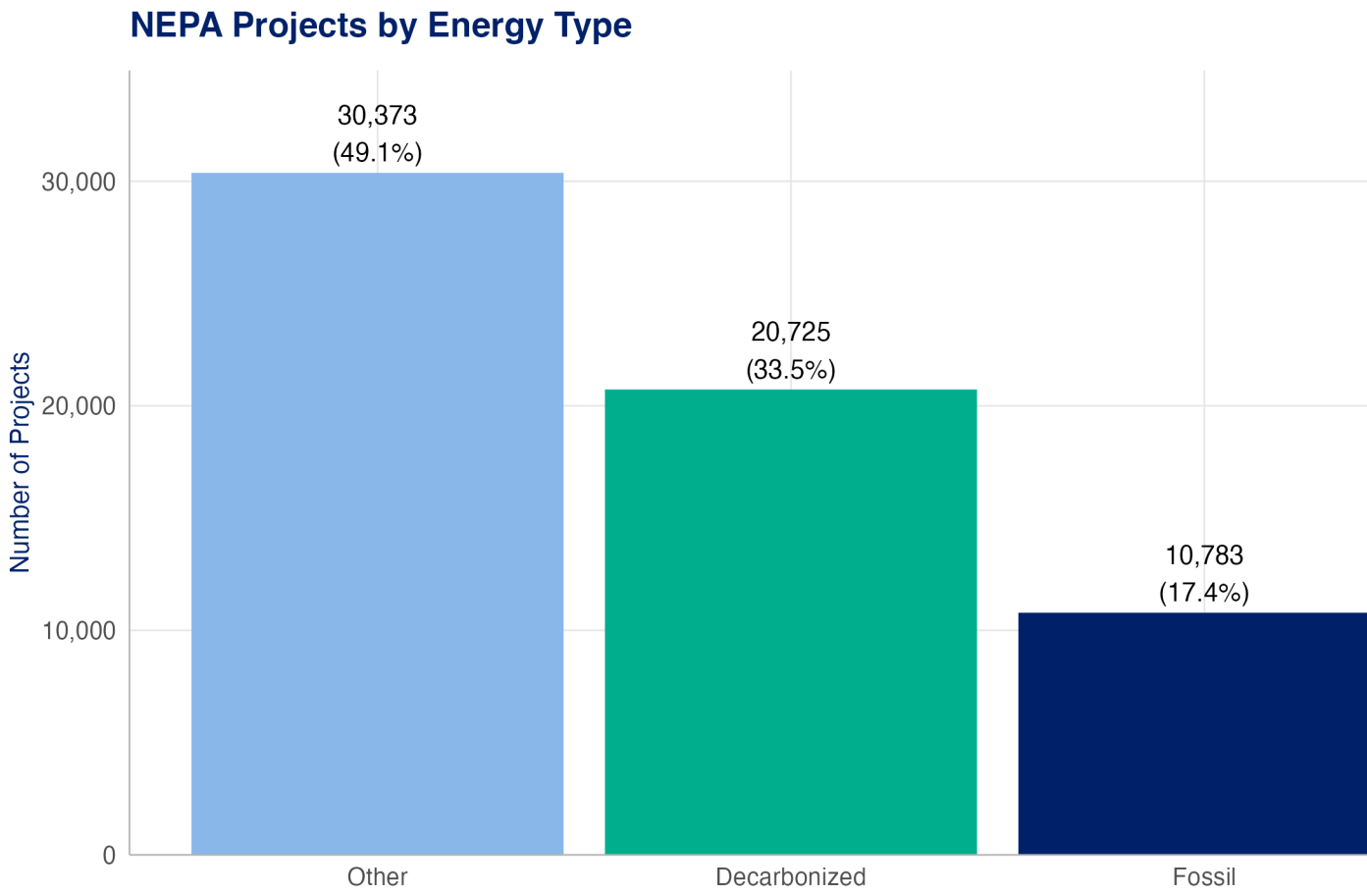

Figure 3: Distribution of all NEPA projects by energy type classification.

- **Nuclear + Waste Management:** We started with approximately **4,000 Nuclear Waste projects** (those tagged with "Waste Management" AND ("Nuclear Technology" OR "Conventional Energy Production - Nuclear")), excluded all projects sponsored by DOE's National Nuclear Security Administration (NNSA), Office of Environmental Management (EM), and Office of Legacy Management (LM), along with their associated field offices. This reduced the dataset to approximately **1,588 projects**, and then CAFT staff identified only **34** projects that were relevant to this analysis. You can see a list of the 34 nuclear-tagged projects that remain in this analysis here.