

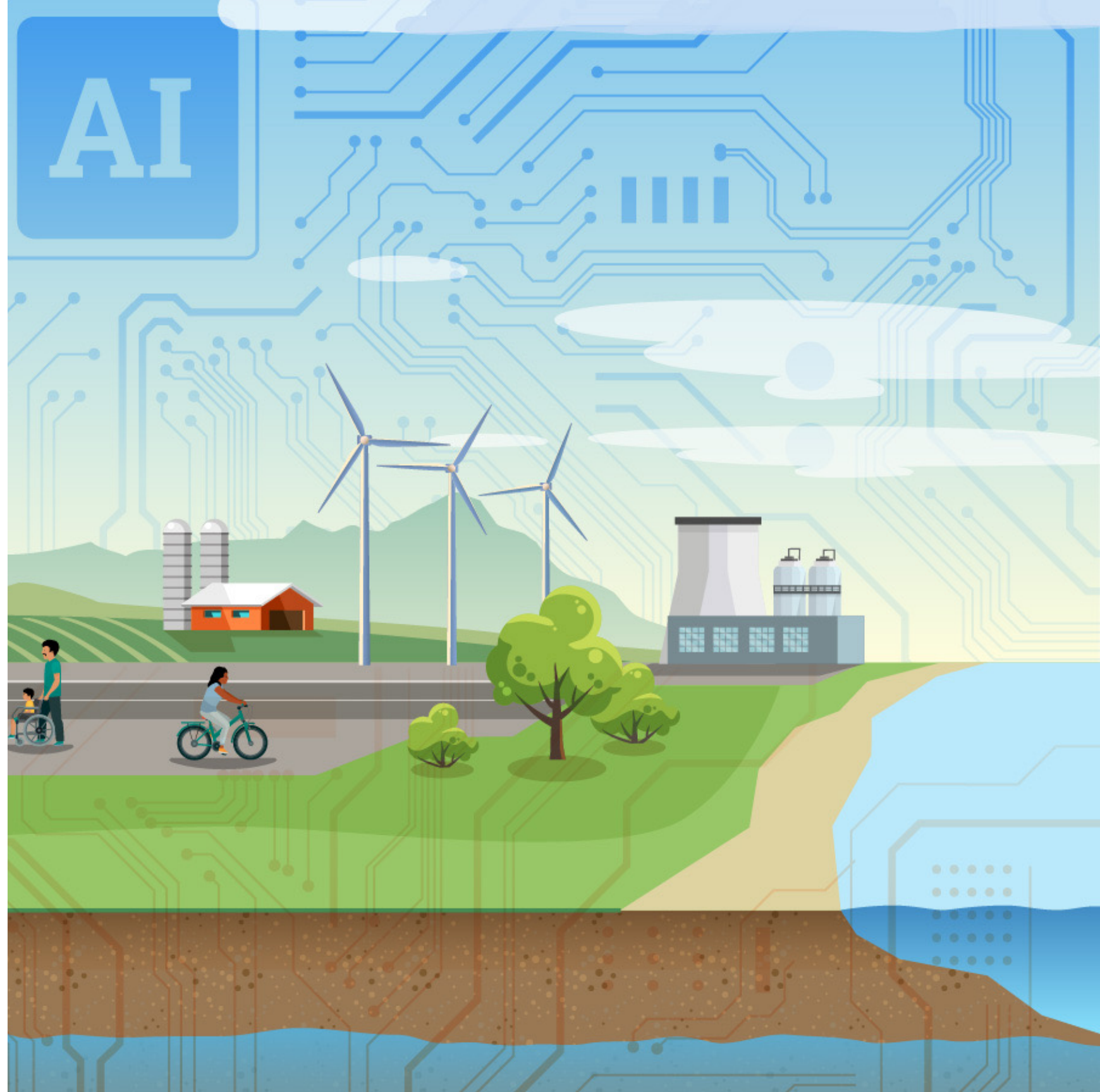
NEPATEC1.0: First Large-Scale Text Corpus of National Environmental Policy Act PDF Documents

Dan Nally

NEPA Project Manager

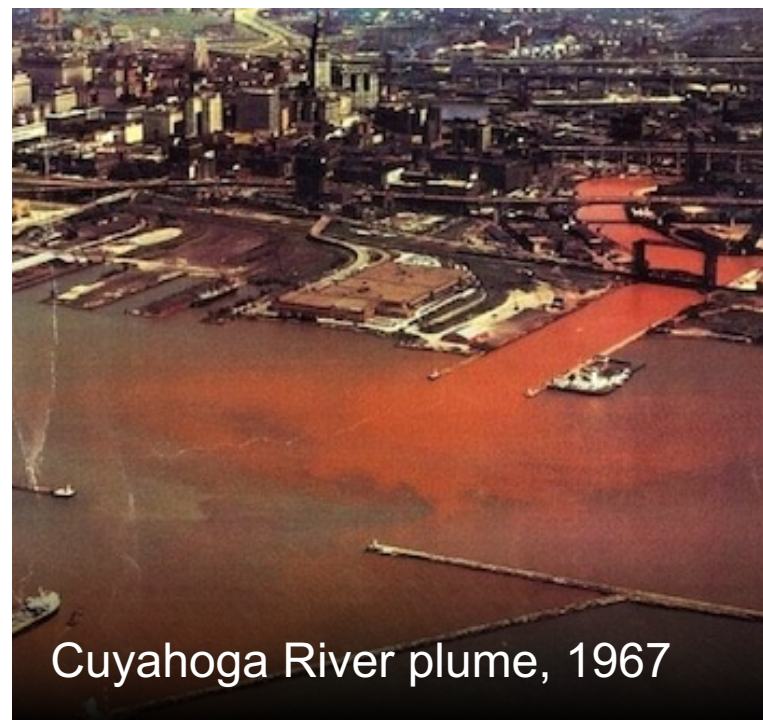
Shivam Sharma

Data Scientist

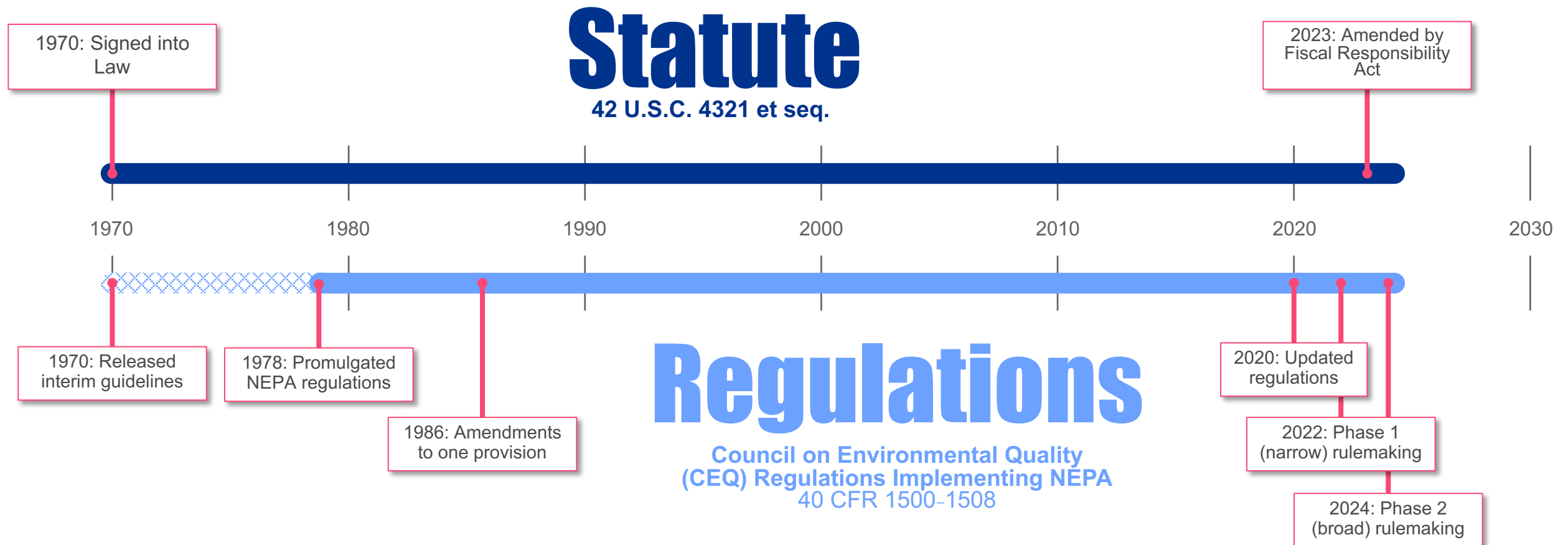


Origin and Purpose of NEPA

- Enacted in 1969; signed into law in 1970
- Landmark and often replicated environmental law
- Requires Federal agencies to consider significant environmental consequences of their proposed actions and inform the public
- Plays critical role in promoting sound decision making and reducing damage to the environment

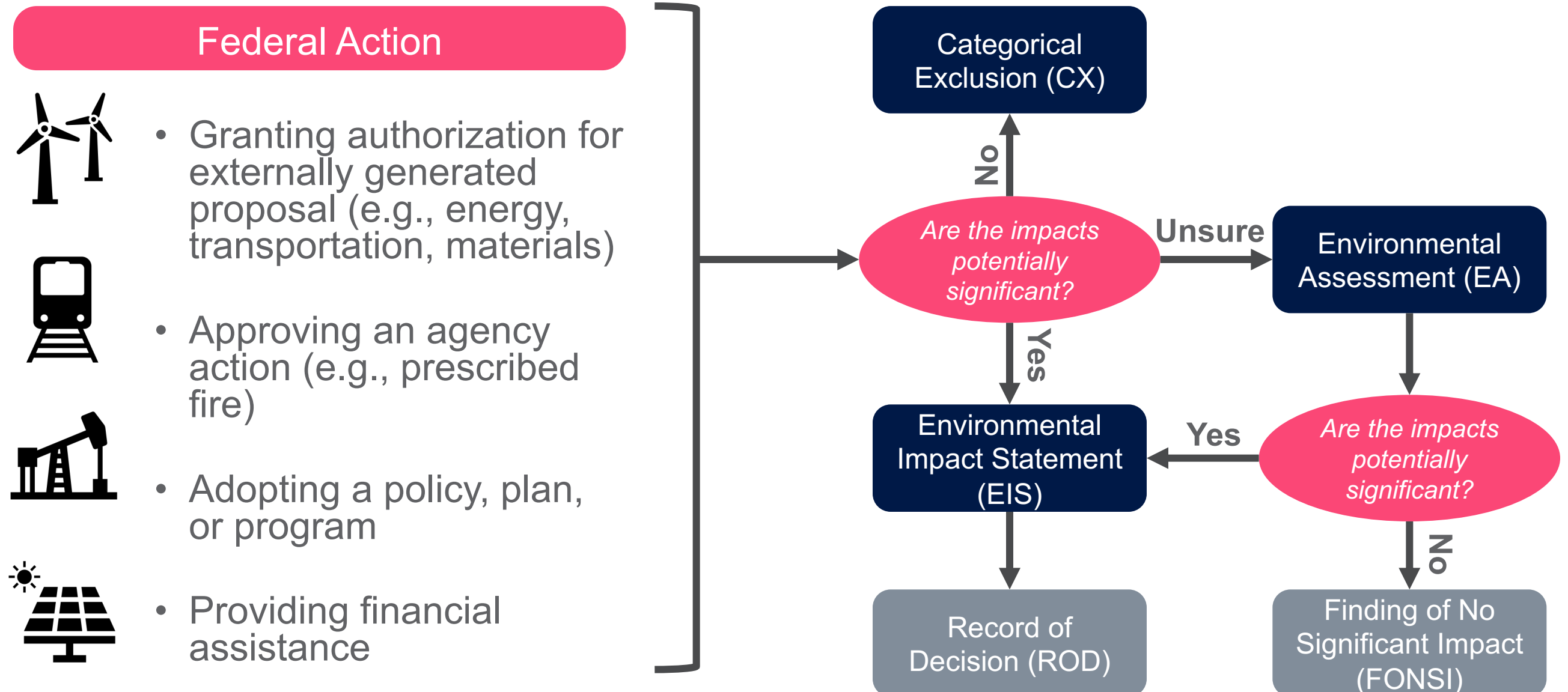


Substantial Changes to NEPA



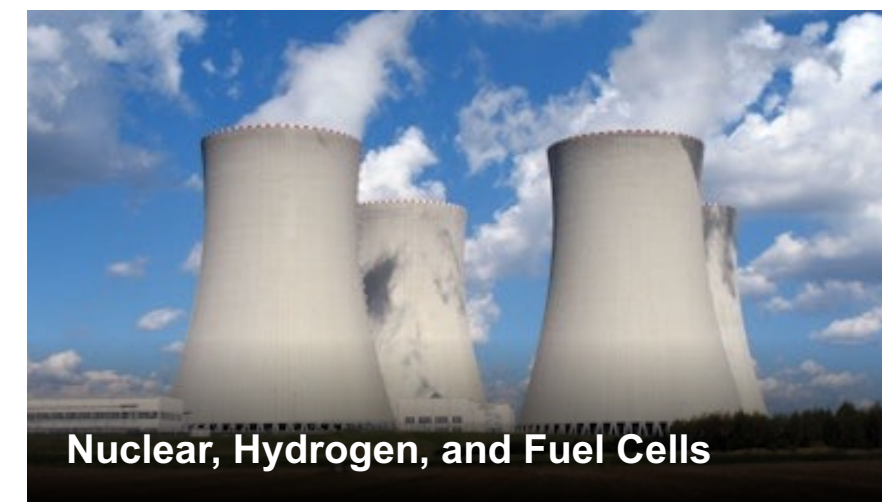
Note: Minor technical and typographical changes not shown.

The NEPA Process



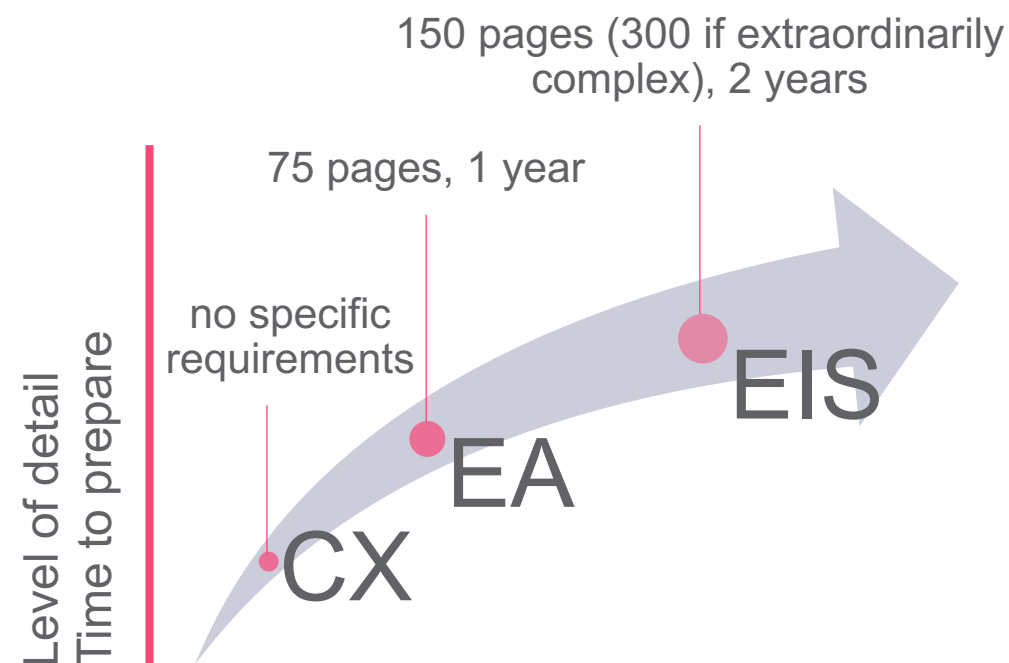
NEPA for Clean Energy Actions

- Commercial-scale projects often involve Federal lands or financial assistance, triggering NEPA requirements



Timelines by Document Type

Regulatory Requirements



Source of regulatory requirements: [40 CFR 1501.10](#)

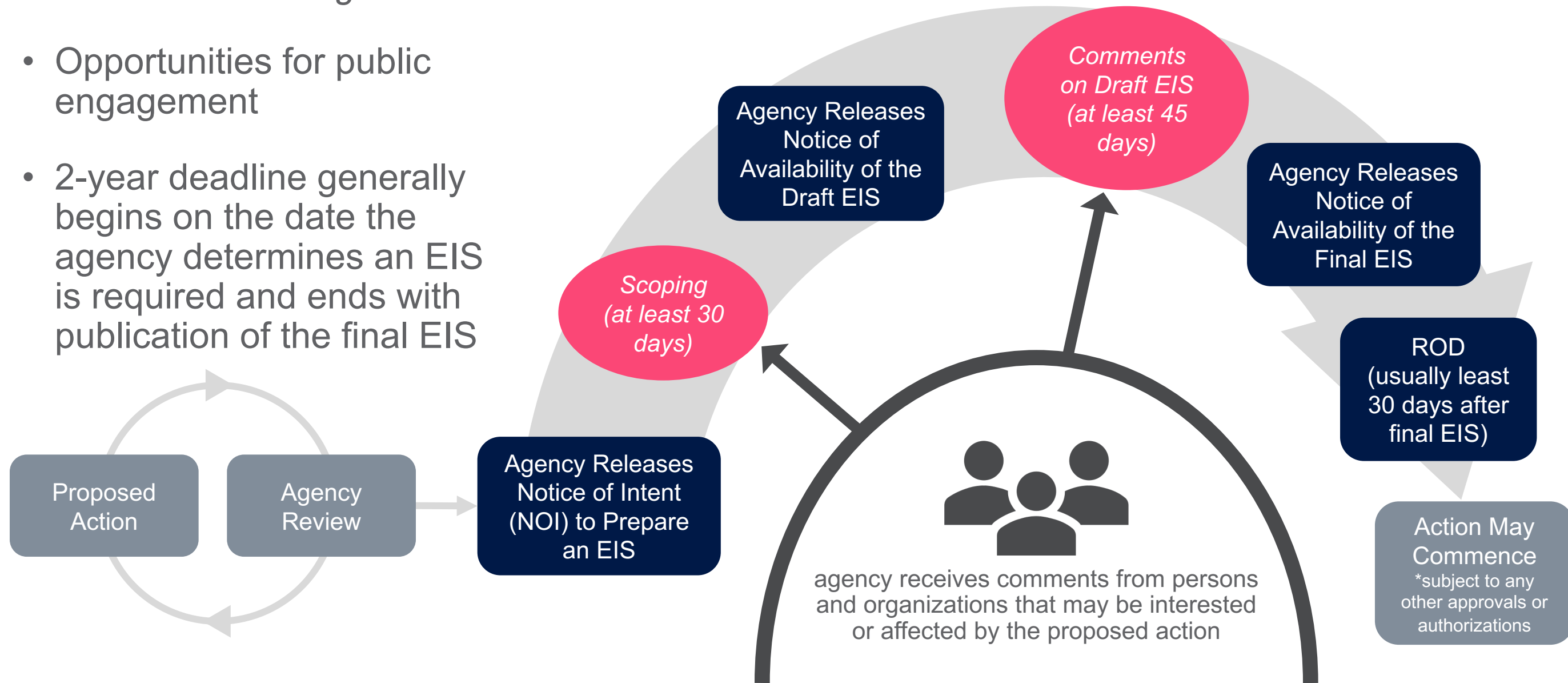
Note: Page limits exclude citations, appendices, and information displayed graphically.

Typical Timelines

- Actual time needed to complete EAs and EISs varies considerably by agency and specific action, but has historically been substantially greater than required under current regulations
- Timelines are influenced by a multitude of factors, including incomplete applications and inadequate funding and staffing of reviewing agencies

EIS Preparation Process

- Major milestones published in the *Federal Register*
- Opportunities for public engagement
- 2-year deadline generally begins on the date the agency determines an EIS is required and ends with publication of the final EIS



Typical Roles

Management and Production

- comment response manager
- document architect (i.e., formatting specialist)
- GIS specialist
- graphic designer
- legal counsel
- NEPA advisor
- project manager
- public engagement specialist
- reference manager
- technical editor

Subject Matter Experts

- air quality and greenhouse gases
- alternatives to the proposed action
- aquatic ecology
- climate change
- cumulative impacts
- environmental justice
- geology, seismology, and soils
- historic and cultural resources
- human health and safety
- hydrology
- land use and visual resources
- meteorology
- noise
- paleontological resources
- socioeconomics
- terrestrial ecology
- transportation and access
- tribal engagement
- waste management

EIS Format

- No universally standardized format, but common elements
- May contain one or more volumes, divided by chapters and appendices
- Predominantly text, but typically contain tables and figures



Example EIS

Document Name	File Size	Release Date	Document Date	Document Type
Dear Interested Party_Juniper-EIS_signed.pdf	286.77 KB	5/24/2024	5/24/2024	PDF
_Juniper_EIS_Vol_1.pdf	45.17 MB	5/24/2024	5/24/2024	PDF
App_A_References.pdf	327.85 KB	5/24/2024	5/24/2024	PDF
App_B_Detailed Description.pdf	52.27 MB	5/24/2024	5/24/2024	PDF
App_C_ACEPMs.pdf	376.74 KB	5/24/2024	5/24/2024	PDF
App_D_GRS Consistency.pdf	346.91 KB	5/24/2024	5/24/2024	PDF
App_E_BMM Eagle Conservation Plan.pdf	15.1 MB	5/24/2024	5/24/2024	PDF
App_F_RFFAs.pdf	388.94 KB	5/24/2024	5/24/2024	PDF
App_G_Supplemental.pdf	2.76 MB	5/24/2024	5/24/2024	PDF
App_H_Glossary.pdf	163.59 KB	5/24/2024	5/24/2024	PDF
App_I_Index.pdf	76.43 KB	5/24/2024	5/24/2024	PDF
App_J_DEIS Comments.pdf	716.13 KB	5/24/2024	5/24/2024	PDF
App_K_FRA Conformance.pdf	33.99 KB	5/24/2024	5/24/2024	PDF

Often divided into multiple files



Table of Contents	
TABLE OF CONTENTS	
CHAPTER 1. INTRODUCTION.....	1-1
1.1. Identifying Information.....	1-1
1.2. Background.....	1-5
1.3. Summary of the Proposed Action.....	1-5
1.3.1. Proposed NOA Plan Amendment (Juniper Project).....	1-5
1.3.2. Eagle Take Permit.....	1-6
1.4. Purpose and Need.....	1-7
1.4.1. BLM.....	1-7
1.4.2. USFWS.....	1-7
1.4.3. RG-BM Objective.....	1-7
1.5. Decisions to Be Made.....	1-7
1.5.1. BLM.....	1-7
1.5.2. USFWS.....	1-8
1.6. Relationship to BLM and Non-BLM Policies, Plans, and Programs.....	1-9
1.6.1. Resource Management Plan Conformance.....	1-9
1.6.2. Federal Regulations, Statutes, and Policies.....	1-10
1.6.3. Bald and Golden Eagle Protection Act.....	1-11
1.6.4. State and Local Land Regulations and Policies.....	1-12
1.6.5. Permits and Approvals.....	1-13
CHAPTER 2. ALTERNATIVES.....	2-1
2.1. BLM Alternatives.....	2-1
2.1.1. No-Action Alternative.....	2-1
2.1.2. Proposed Action.....	2-6
2.1.3. Alternative A (BLM Preferred Alternative).....	2-35
2.2. USFWS Eagle Permit Decision Alternatives.....	2-46
2.2.1. USFWS No-Action Alternative.....	2-47
2.2.2. USFWS Action Alternatives.....	2-47
2.3. Alternatives Considered but Eliminated From Detailed Analysis.....	2-55
2.3.1. Complete Backfill of Underground Workings.....	2-56
2.3.2. Complete Backfill of Open Pits.....	2-56
2.3.3. No or Minimal Surface Disturbance Increase in Ruby Valley Hydrographic Basin.....	2-57
2.3.4. No New Water Production Wells in Ruby Valley Hydrographic Basin.....	2-57
2.3.5. No New Surface Disturbance in Ruby Lake National Wildlife Refuge Viewshed.....	2-58
CHAPTER 3. AFFECTED ENVIRONMENT AND ENVIRONMENTAL CONSEQUENCES.....	3-1
3.1. Introduction.....	3-1
3.2. Geology and Mineral Resources.....	3-4
3.2.1. Affected Environment.....	3-4
3.2.2. Environmental Consequences.....	3-13
3.3. Water Quality and Quantity.....	3-24
3.3.1. Affected Environment.....	3-24

Organized into chapters and appendices

Information contained in text, figures, tables, and cited references

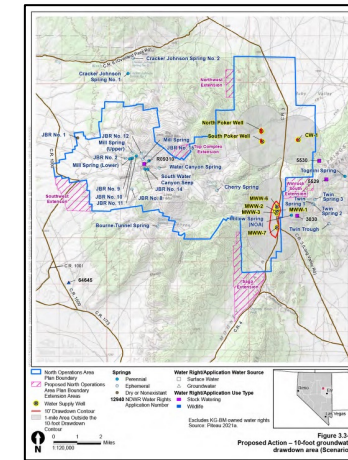


Table 3-8. Comparison of proposed pit floor and groundwater elevations

Pit	Floor Elevation (feet amsl)		Groundwater Elevation (feet amsl) ¹			Estimated Minimum Depth below Pit Floor to Groundwater (feet) ^{2,3}	
	Authorized	Proposed	Minimum	Maximum	Average ⁴	Authorized	Proposed
Redbird	6,620	6,100	5,995	6,054	6,050	566	46
Rat	7,225	6,975	<6,800	<6,800	-	>425	>175
Top expansion	-	7,540	6,702	7,425	6,903	-	115
Royale	-	6,125	6,060	6,095	6,059 ⁵	-	70 ⁶
South Duke	-	6,125	6,049	6,069	6,059	-	56
Bida	7,030	6,525	<6,515	6,061	6,061	969	464
Saga	6,225	6,200	5,922	6,143	6,098	82	57 ⁷
Winrock South	6,600	6,400	6,193	6,556 ⁸	6,369	216	16 ⁹

As shown in Table 3-62, the seven tribal communities in the study area had a vacancy rate of 19% in 2020. The South Fork Reservation and Off-Reservation Trust Land had the highest vacancy rate of 44%, and Elko Colony the lowest vacancy rate at 5%. The median home value was 55% lower in the tribal communities than in the study area at \$79,400 in 2018. The median rent was 48% lower than in the study area, as well (U.S. Census Bureau 2010b, 2022b, 2022c).

Link to Example EIS: [BLM 2024](#)

Variable Format of EISs

CHAPTER 3. AFFECTED ENVIRONMENT AND ENVIRONMENTAL CONSEQUENCES	3-1
3.1. Introduction	3-1
3.2. Geology and Mineral Resources	3-4
3.2.1. Affected Environment	3-4
3.2.2. Environmental Consequences	3-13
3.3. Water Quality and Quantity	3-24
3.3.1. Affected Environment	3-24
3.3.2. Environmental Consequences	3-46
3.4. Soils and Reclamation	3-84
3.4.1. Affected Environment	3-86
3.4.2. Environmental Consequences	3-92
3.5. Vegetation, Special Status Plants, and Wetlands	3-98
3.5.1. Affected Environment	3-98
3.5.2. Environmental Consequences	3-104
3.6. Noxious Weeds and Nonnative Invasive Plant Species	3-115
3.6.1. Affected Environment	3-117
3.6.2. Environmental Consequences	3-117
3.7. Wildlife and Fisheries Resources	3-120
3.7.1. Affected Environment	3-120
3.7.2. Environmental Consequences	3-139
3.8. Special Status Wildlife Species	3-164
3.8.1. Affected Environment	3-166
3.8.2. Environmental Consequences	3-168
3.9. USFWS Golden Eagle Permit Decision	3-209
3.9.1. Affected Environment	3-209
3.9.2. Environmental Consequences	3-217
3.10. Livestock Grazing	3-227
3.10.1. Affected Environment	3-229
3.10.2. Environmental Consequences	3-233
3.11. Wild Horses	3-239
3.11.1. Affected Environment	3-239
3.11.2. Environmental Consequences	3-241
3.12. Paleontological Resources	3-243
3.12.1. Affected Environment	3-244
3.12.2. Environmental Consequences	3-248
3.13. Cultural Resources	3-251
3.13.1. Affected Environment	3-253
3.13.2. Environmental Consequences	3-257
3.14. Native American Traditional Values	3-260
3.14.1. Affected Environment	3-260
3.14.2. Environmental Consequences	3-261

CHAPTER 3 Affected Environment	3-1
3.1 Introduction	3-1
3.1.1 Scoping Issues and Concerns	3-1
3.1.2 Study Area	3-5
3.2 Species of Special Concern	3-7
3.2.1 Gray Wolf	3-7
3.2.2 Other Federally Listed Species	3-12
3.2.3 State-Listed Species	3-16
3.3 Other Wildlife Species	3-17
3.3.1 Elk and Deer	3-17
3.3.2 Other Ungulates	3-18
3.4 Tribal Resources	3-19
3.4.1 Archaeological and Historical Sites	3-19
3.4.2 Natural Resources of Cultural Importance	3-19
3.4.3 Tribal Treaty Rights and Reservations	3-20
3.4.4 Government-to-Government Consultation	3-21
3.5 Socioeconomic Resources	3-22
3.5.1 Human Activity in Colorado	3-22
3.5.2 Industry Sectors in Colorado	3-25
3.6 Environmental Justice	3-29
3.6.1 Methodology	3-29
3.6.2 Existing Conditions	3-30
CHAPTER 4 Environmental Consequences	4-1
4.1 Introduction	4-1
4.2 General Methodology for Assessing Impacts	4-1
4.3 General Analysis Methodology and Assumptions	4-1
4.3.1 Assessing Impacts Using Council on Environmental Quality Criteria	4-1
4.3.2 Assumptions	4-2
4.3.3 Jurisdiction and Compliance	4-2
4.4 Species of Special Concern	4-3
4.4.1 Gray Wolf	4-3
4.4.2 Other Species of Special Concern	4-7
4.5 Other Wildlife Species	4-8
4.5.1 No-Action Alternative	4-9
4.5.2 Alternative 1	4-9
4.5.3 Alternative 2	4-10

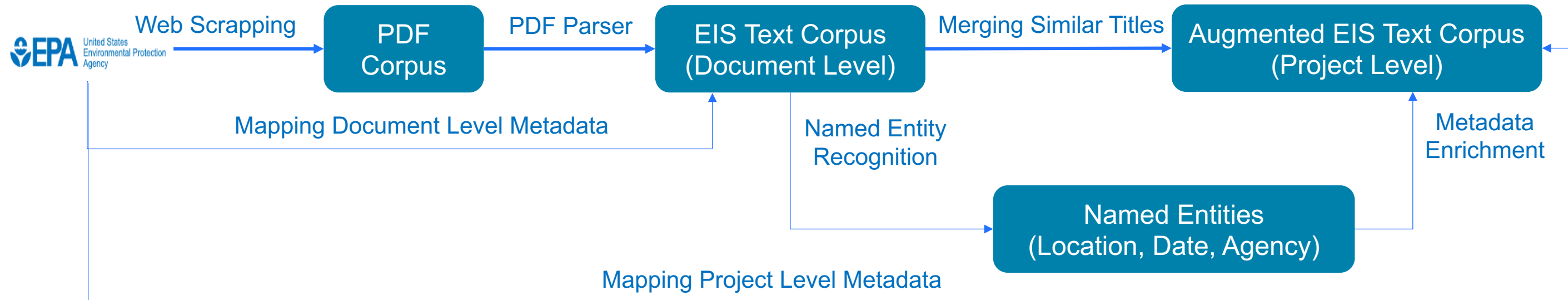
Chapter 3 Affected Environment and Environmental Consequences	3-1
3.1 Impact-Producing Factors	3.1-1
3.2 Mitigation Identified for Analysis in the Environmental Impact Statement	3.2-1
3.3 Definition of Impact Levels	3.3-1
3.3.1 Activities Terminology	3.3-1
3.3.2 Impact Terminology	3.3-2
3.4 Physical Resources	3.4-1
3.4.1 Air Quality	3.4.1-1
3.4.2 Water Quality	3.4.2-1
3.5 Biological Resources	3.5-1
3.5.1 Bats	3.5.1-1
3.5.2 Benthic Resources	3.5.2-1
3.5.3 Birds	3.5.3-1
3.5.4 Coastal Habitat and Fauna	3.5.4-1
3.5.5 Finfish, Invertebrates, and Essential Fish Habitat	3.5.5-1
3.5.6 Marine Mammals	3.5.6-1
3.5.7 Sea Turtles	3.5.7-1
3.5.8 Wetlands	3.5.8-1
3.6 Socioeconomic Conditions and Cultural Resources	3.6-1
3.6.1 Commercial Fisheries and For-Hire Recreational Fishing	3.6.1-1
3.6.2 Cultural Resources	3.6.2-1
3.6.3 Demographics, Employment, and Economics	3.6.3-1
3.6.4 Environmental Justice	3.6.4-1
3.6.5 Land Use and Coastal Infrastructure	3.6.5-1
3.6.6 Navigation and Vessel Traffic	3.6.6-1
3.6.7 Other Uses (Marine Minerals, Military Use, Aviation, and Scientific Research and Surveys)	3.6.7-1
3.6.8 Recreation and Tourism	3.6.8-1
3.6.9 Scenic and Visual Resources	3.6.9-1

Table of contents of three EISs from different agencies illustrating different scope of resource analyzed and different formats for organizing the discussions of the affected environment and environmental consequences.

EISs pictured (left to right): [BLM 2024](#), [USFWS 2023](#), [BOEM 2024](#)

NEPATEC1.0 Construction

NEPA Text Corpus



	Title	Document	EPA Comment Letter Date	Federal Register Date	Agency	State
10	Virginia Reliability Project and Commonwealth ...	Final	-1	09/22/2023	Federal Energy Regulatory Commission	VA
80	Virginia Reliability Project and Commonwealth ...	Draft	06/05/2023	04/21/2023	Federal Energy Regulatory Commission	VA

Exact Duplicates

	Title	Document	EPA Comment Letter Date	Federal Register Date	Agency	State
6802	~ VOIDED ~ Council Bluffs Interstate System Im...	Draft	-1	09/30/2005	Federal Highway Administration	IA
6829	Council Bluffs Interstate System Improvements...	Final	10/07/2005	09/09/2005	Federal Highway Administration	IA
7211	Council Bluffs Interstate System Improvements ...	Draft	01/14/2005	12/23/2004	Federal Highway Administration	IA

Partial Duplicates

NEPATEC1.0 Statistics

- We scrapped a total of **35,427** PDFs from the EPA website from **12,376** EIS Project Links
- We had a total of **16,310** EIS Project Metadata
- We did a 2-step EIS Project Title merging:
 - Duplicate Title Merging
 - Fuzzy Title Merging
- After mapping the PDFs to corresponding metadata, we had a total of **28,212** PDFs from **2,917** Unique EIS Projects
 - Total number of pages: **4.5 Million**
 - Total number of tokens (GPT2 tokenizer): **3.6 Billion**

3K Projects

28K PDFs

4.5M Pages

3.6B Tokens

NEPATEC1.0 Structure

- Dataset is organized by the project, where a project can have multiple documents
- Each project has the following data:
 - Title
 - Project metadata
 - ✓ Agency
 - ✓ States
 - ✓ Dates
 - Page-wise text
 - ✓ NER for each page
 - ✓ Each NER consist of the following features:
 - Text: text for the named entity
 - Label: Label for the named entity
 - Score: Confidence score for the text to belong to the given label

NEPATEC1.0 Named Entity Recognition

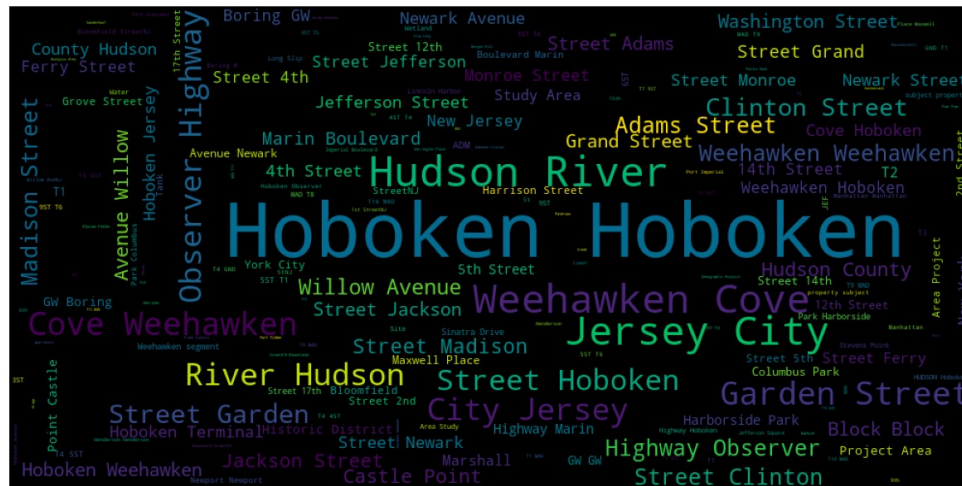
- We extracted a list of five entities from the text:
 - Name: Any name, ranging from name of person to project name
 - Date: Any reference to a specific data or just the year
 - Agency: Any organization
 - Location: Any location, ranging from site location to street, county, state, or country
 - Title: Aimed to extract title of the document and any relevant titles of mentioned documents
- Low threshold to retain major entities
 - Subsample based on score as needed

NORTH CITY Location PROJECT Name EIR EIS RESPONSE TO COMMENTS CAPCOA 2017 Date is the default wind speed for San Diego County Location which is taken from data from the Gillespie Field Location meteorological station Title and includes data from 1996 Date through 2006 Date WRCC 2017 Date This dataset includes hourly wind data as recorded by that station for that time period which includes high wind events Therefore the fugitive dust emissions calculated within CalEEMod Agency account for high wind events within its results From historical records Santa Ana Location winds can easily exceed 50 miles per hour and during a high wind event earth disturbing work would not occur This would be a standard approach by the contractor Name to comply with SDAPCD Agency Rules 55 Fugitive Dust 50 Visible Emissions and 51 Nuisance As stated within the Draft EIR EIS Title the Project Name will comply with all SDAPCD Agency applicable rules Specifically the Project Name would be prevented from allowing emissions during a high wind event by SDAPCD Agency

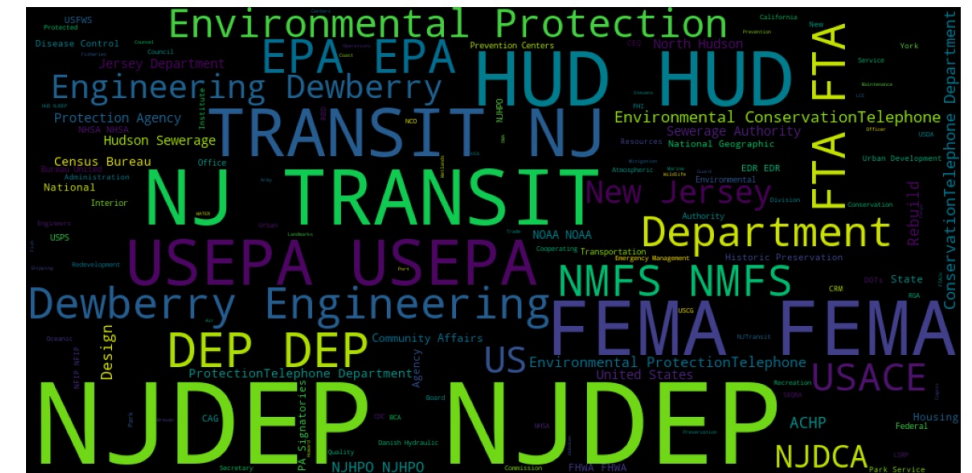
NEPATEC1.0 Named Entity Recognition

Example Word Clouds

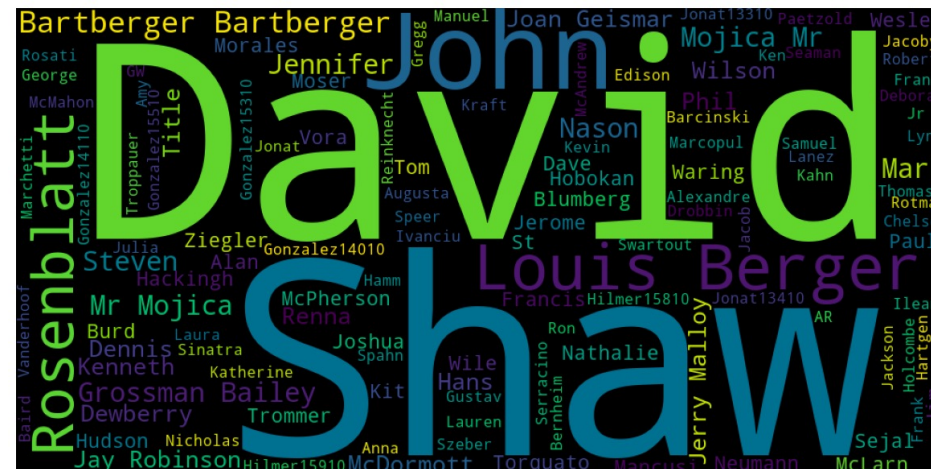
ADOPTION - Rebuild by Design - Hudson River (RBD-HR)



Location Word Cloud



Agency Word Cloud



Name Word Cloud

NEPATEC1.0 Named Entity Recognition

Location NER Tags

ADOPTION - Rebuild by Design - Hudson River (RBD-HR)

```
['Bloomfield StreetNJ',  
'12th Street',  
'Park Avenue',  
'Hoboken Avenue',  
'Paterson Avenue',  
'Adams StNJ',  
'Jefferson Street',  
'GW 5',  
'Willow Avenue',  
'Vezzetti Way',  
'Jersey Avenue',  
'Willow AveNJ',  
'6th Street',  
'Shippen Street',  
'Park Avenue',  
'Clinton Street',  
'8th Street',  
'Frank Sinatra Drive',  
'16th Street',  
'Newark Street',  
'Jefferson Street',  
'Newark Avenue',
```

Street Locations

```
['Hudson county', 'Hudson County']
```

County Locations

```
['NJ', 'Massachusetts', 'NJ11', 'New Jersey', 'NJ']
```

State Locations

```
['Weehawken Cove',  
'Newark',  
'Stevens Point',  
'Princeton',  
'Castle Point',  
'Warrington Plaza',  
'Hoboken',  
'Winfield',  
'Hudson Falls',  
'Englewood Cliffs',  
'Hudson',  
'Newport',  
'Hoboken',  
'Hoboken',  
'New York City',  
'Fort Lee',  
'Manhattan',  
'Madison',  
'Henderson',  
'Philadelphia',  
'Weehawken',  
'Vanderhoof',  
'Hudson PlaceNJ',  
'NYC',  
'San Francisco']
```

City Locations

NEPATEC1.0 Named Entity Recognition

Location NER Tags

ADOPTION - Rebuild by Design - Hudson River (RBD-HR)

References LR 3 Title HBLR Agency MOS 1 Long Slip Location Jersey City Location Hoboken Location Hudson County New Jersey Location Field Testing and Data Recovery Report On file Trenton Location NJHPO Agency 2004a Date HBLR MOS 2 2nd Street Station Location

Hoboken Hudson County New Jersey Memo Report on Archaeological Field Testing On file Trenton NJHPO 2004bHans Date Jim Name 100 Hoboken Firsts Hoboken Hoboken Historical Museum 2005Hartman Date David Name and Barry Lewis Name "A Walk through Hoboken" Accessed February 9 2016 Date http wwwthirteenorg hoboken Location historycfm Historical Perspectives Inc and The Louis Berger Group Inc Archaeological Documentary Study No 7 Line Extension Hudson Yards Location Rezoning New York Location New York On File New York Landmarks Preservation Commission Agency

2004Hoboken Date 411 Photograph "Photos of Hoboken Flooding " May 2015 Date Accessed on December 2016 Date http hoboken411com archives 112803Hoboken Green Infrastructure Strategic Plan Together North Jersey October 2013 Date Retrieved from https onedrivecom view.aspx resid 2F0F556D773BC90F 2343 app WordPdfHoboken Historical Museum The Stevens Family Accessed online at https wwwhobokenmuseumorg history the stevens family"Hoboken's Posthistory the stevens family"Hoboken's Post Sandy Resilience" Center for Science and Democracy Agency Fact sheet February 2014 Date Retrieved from http wwwucsusaorg sites default files legacy assets documents center for science and democracy Agency hoboken Location post sandy resiliencepdfHoboken Planning Board City of Hoboken Master Plan City of Hoboken Hoboken Planning Board 2004Hoboken Date Resiliency and Readiness Plan August 2013 Date City of Hoboken Hudson County Location NJHolochuck N 2000 Date Hudson River Location Submerged Aquatic Vegetation New York Location Sea Grant Extension Program Kingston NY Location Hudson County 2008 Date Reexamination of the Master Plan http wwwhudsoncountynjorg wp content uploads 2013 06 Hudson County Master Plan Reexamination Report 2008pdfHudson River Foundation Available online at wwwhudsonriverorg "Hudson River Hoboken Weehawken Location Jersey City Location New Jersey Location Draft scoping Document Environmental Impact Statement Draft Scope of Work Rebuild by Design" Free Electronic Library Page 3 Available at http wwwdislibxinfo dd other 360804 3 hudson river Location hoboken weehawken Location jersey city Location new jersey Location draft scopphpJersey City Online "Jersey City America's Golden Door""Jersey City America's Golden Door" Accessed online May 30 2016 Date at http wwwjerseycityonlinecom history jc historyhtmKalm James Name Aka Loren Munk Name Photograph October 2012 Date Accessed November 29 2016 Date http joannematterablogspotcom 2012 11 marketing mondays hell and high waterhtmlKarnoutsos Carmela Name Jersey City Past and Present Title Jersey City New Jersey City University 2007 Date Accessed online at https wwwnjcu edu programs jchistory pages D Pages Dutch West India Location CohtmLevandowsky M and D Vaccari Name

2004 Date Analysis of Phytoplankton Data from Two Lower Manhattan Location Sites Final Report of a Grant from the Hudson River FoundationLichvar Agency RW Name DL Banks Name WN Kirchner Name and NC Melvin Name 2016 The National Wetland Plant List 2016 wetland ratings Phytoneuron 2016 30 1 17 Published 28 April 2016 Date ISSN 2153 733XLurie Maxine N Name and Marc Mappen Name Encyclopedia of New Jersey New Brunswick Location Rutgers University Press 2004Marcopul Date Katherine Name 2017 Date New SHPO Agency Opinion RB Davis Company Hudson County Location Hoboken City Location Weehawken Township Location Jersey City Rebuild by Design Title Resist Delay Store DischargeDesign Resist Delay Store Discharge Project Community Development Block Grant Disaster Recovery Agency Department of Housing and Urban DevelopmentMargolin Agency Josh Name New York Post Location " NJ Transit Agency Investigating its Preparations for Hurricane Sandy after Numerous Trains Damaged in Storm" November 2012 Date Accessed December 2016 http nypostcom 2012 11 17 Date nj transit Agency investigating its preparations for hurricane sandy after numerous trains damaged in storm Mary Delaney Name Krugman Associates Inc and William Sandy Name Preliminary Cultural Resource Study Title Existing Conditions for the Frank Sinatra Drive Location Visioning and Conceptual Design Plan On File Trenton Location New Jersey Historic Preservation Office Agency 2014Matheson Date Mark P Name Department of Psychiatry Medical Sciences Building Queen Mary University of London Location London UKMeyers Russell W Name Deputy State Historic Preservation Officer Agency Correspondence and attached comments regarding Weehawken to Edgewater Reach Location aka Hoboken to North Bergen Reach Location New York Harbor Location Collection and Removal of Dirt to Colonel FH Griffis Name District Engineer Name US Army Corps of Engineers Agency September 12 1984Mid Date Atlantic Fisheries Council12 1984Mid Atlantic Fisheries Council MAMFC Agency Available online at http wwwmamfcorg Moss Linda Name April 3 2014 Date Thomson Reuters Agency relocating 450 workers to Hoboken Location http wwwnorthjerseycom news business thomson reuters Agency relocating 450 workers to hoboken Location 1841182 Date Accessed November 2 2016Munoz Date Eduardo Name NPR Agency Photograph "Latest on Sandy Death Toll Rises Wait for 'Normal' Life Continues" October 2012 Date Accessed December 2016 Date http wwwnprorg sections thetwo way 2012 10 31 Date 164014421 Date latest on sandy death toll rises wait for normal life continuesNational Historic Preservation Act of 1966 as amended 2004 Date NHPA Location 36 CFR 8001 Title 16dNational Historic Preservation Act 16 USC 470 Title as amended re codified under Title 54 Title " National Park Service and Related Programs Agency " 54 USC 300101 Title et seq 2015Advisory Council on Historic Preservation Protection of Historic Properties Regulations implementing Section 106 Title of the NHPA 36 CFR 800National Marine Fisheries Service NMFS Agency Summary of Essential Fish Habitat EFH Designation Available online at http wwwnoaa gov indexhtmlNational Park Service National RegisterindexhtmlNational Park Service National Register Bulletin 15 Title How to Apply the National Register Criteria for Evaluation Washington DC Location US Department of Interior Agency National Park Service Cultural Resources Agency Interagency Resources Division Agency 1995LIST Date OF references Cont d

Reference Section/Document

Challenges and Opportunities

- Existing online repositories containing EISs and other NEPA documents are incomplete and have limited search capabilities
- EISs are long documents that do not have a universally consistent or standardized structure
- AI can assist people involved in the preparation and review of NEPA documents and enhance efficiency by:
 - searching for, interpreting, and synthesizing content from existing NEPA documents
 - assessing trends in NEPA documents and processes over time
 - generating draft content or reviewing text of new NEPA documents
 - assisting with analysis and summarization of public comments

Technical Opportunities, Users, and Solutions

Below we present a set of problems, ways solutions would benefit potential users (high-level stakeholders, researchers/authors of EISs, and members of the public), and examples of potential solutions.

- **NEPA documents often contain technical jargon and are challenging for the public to understand.**

- Need: Apply language modeling to suggest revisions to NEPA documents to improve clarity through use of plain text (e.g., clear, understandable prose)
- Solutions:
 - ✓ Develop AI-powered tools for enhancing document clarity and readability.
 - ✓ Implement a plain language checker and recommendation for NEPA documents.

- **High-Level Stakeholders:** Ensures documents meet policy and regulatory standards.
- **Researchers:** Helps in producing clear and concise research outputs.
- **General Public:** Improves document readability and comprehension.



Technical Complexity Level

- **Reviewing and synthesizing information from multiple NEPA documents is time consuming.**

- Need: Enable rapid comparative analysis and summarization of multiple NEPA documents.
- Solutions:
 - ✓ Develop natural language processing algorithms to summarize documents.
 - ✓ Create a comparative analysis tool for multiple NEPA documents.

- **High-Level Stakeholders:** Provides quick insights for decision-making.
- **Researchers:** Assists in comparative studies across multiple documents.
- **General Public:** Simplifies complex information for better understanding.



Technical Complexity Level

Technical Opportunities, Users, and Solutions (Continued)

- **NEPA documents are information rich but have limited to no metadata.**

- Need: Extract multi-level metadata to facilitate information retrieval and analysis by LLMs.
- Solutions:
 - ✓ Develop software to reliably harvest metadata from a variety of file types (PDFs, Word documents, etc.)
 - ✓ Implement a metadata standardization system to ensure consistency across data sets.

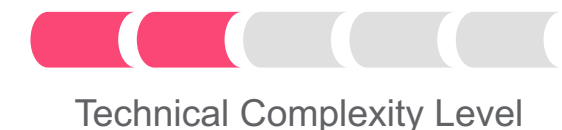
- **High-Level Stakeholders:** Helps in policy-making and oversight.
- **Researchers:** Aids in academic and practical research.
- **General Public:** Assists in understanding project context and relevance.



- **How do NEPA projects and processes change with place and time?**

- Need: Find documents or assess spatio-temporal trends based on project location and points in time.
- Solutions:
 - ✓ Create visualization tools for spatio-temporal data analysis.
 - ✓ Develop machine learning models to predict trends based on historical data.

- **High-Level Stakeholders:** Enables strategic planning based on spatial and temporal trends.
- **Researchers:** Aids in identifying patterns and anomalies over time.
- **General Public:** Enhances public awareness of environmental changes.



Technical Opportunities, Users, and Solutions

(Continued)

- **NEPA documents are mostly text and not designed to convey information quickly or visually.**

- Need: Display specific subsets of information graphically to enhance understanding (e.g., word cloud, complex numerical taxonomic clustering).
- Solutions:
 - ✓ Develop dynamic visual tools like word clouds and taxonomic clustering.
 - ✓ Create dashboards for visual representation of complex data sets.

- **High-Level Stakeholders:** Assists in data-driven decision-making.
- **Researchers:** Helps to visually interpret complex data patterns.
- **General Public:** Makes technical information accessible through visuals.



Technical Complexity Level

- **Few quantitative metrics are available to measure the efficiency of the NEPA process.**

- Need: Assess trends in document length, process length, project types, and other characteristics over time.
- Solutions:
 - ✓ Develop tools to track document length, process duration, and other efficiency metrics.
 - ✓ Implement dashboards for real-time monitoring of NEPA efficiency.

- **High-Level Stakeholders:** Informs policy reviews and legislative changes.
- **Researchers:** Supports empirical studies and trend analysis.
- **General Public:** Offers transparency in governmental processes.



Technical Complexity Level

Technical Opportunities, Users, and Solutions (Continued)

- **Geospatial information is not directly embedded in NEPA documents.**

- Need: Extract detailed project location data to enable geographical search and localization capabilities.
- Solutions:
 - ✓ Create a geographic information system (GIS) integration module.
 - ✓ Develop APIs to extract and standardize location data from NEPA documents.

- **High-Level Stakeholders:** Enhances geographical planning and resource allocation.
- **Researchers:** Facilitates spatial analysis and correlation studies.
- **General Public:** Provides an easy way to identify projects in their vicinity.



Technical Complexity Level

- **Are NEPA documents using best available science?**

- Need: Identify and assess trends in scientific concepts and studies cited in NEPA documents over time and space
- Solutions:
 - ✓ Develop text mining and analysis tools to identify key scientific concepts.
 - ✓ Implement trend analysis modules for scientific studies cited in NEPA documents.

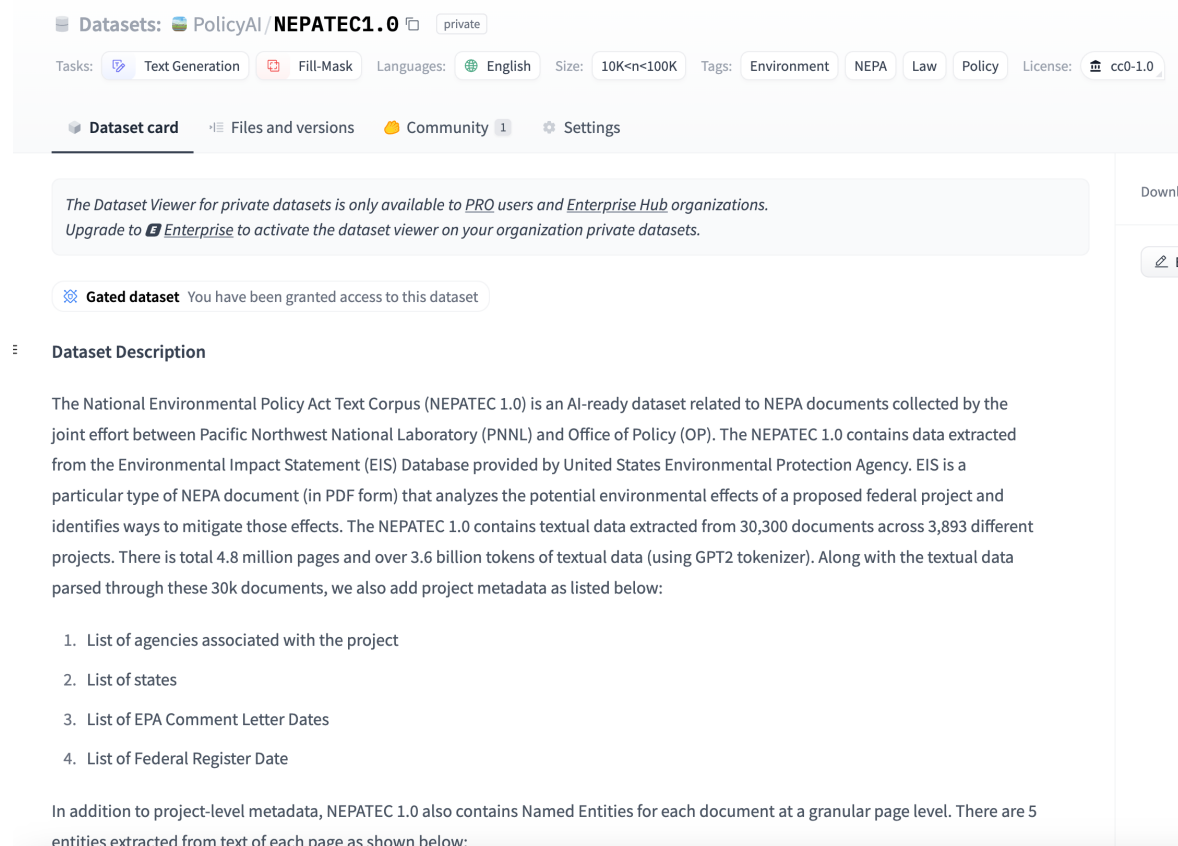
- **High-Level Stakeholders:** Informs evidence-based policy-making.
- **Researchers:** Enables trend analysis in scientific research.
- **General Public:** Promotes awareness of the scientific foundations of projects.



Technical Complexity Level

NEPATEC1.0 Dataset

- NEPATEC1.0 Dataset is publicly available in [HuggingFace](#) under [Creative Commons “0” license \(public domain dedication\)](#)



The screenshot shows the HuggingFace dataset page for NEPATEC1.0. The dataset is categorized under 'PolicyAI' and is marked as 'private'. It includes filters for tasks (Text Generation, Fill-Mask), languages (English), size (10K<n<100K), tags (Environment, NEPA, Law, Policy), and license (cc0-1.0). The page features tabs for 'Dataset card', 'Files and versions', 'Community', and 'Settings'. A message states that the dataset viewer is only available to PRO users and Enterprise Hub organizations. A 'Gated dataset' notification indicates that access has been granted. The 'Dataset Description' section explains that the dataset is an AI-ready corpus of NEPA documents, containing 30,300 documents across 3,893 projects, with 4.8 million pages and 3.6 billion tokens of text. It also lists project-level metadata: agencies, states, EPA comment dates, and federal register dates. Finally, it mentions that the dataset includes named entities at the page level, with 5 entities extracted per page.

NEPATEC1.0: First Large-Scale Text Corpus of National Environmental Policy Act PDF Documents

Abstract

An environmental impact statement (EIS) is a written document that contains detailed analysis of the potential environmental effects of a proposed major federal action. The preparation of an EIS and other procedural requirements of the National Policy Act (NEPA) are mainstays of federal decision-making and natural resource management. NEPA serves as a critical environment safeguard and opportunity for public engagement, while also facing scrutiny from efforts to streamline and expedite environmental permitting processes enabling the deployment of critical energy and infrastructure projects. Directed retrieval and interpretation of information contained in completed EISs, individually and in aggregate, could help improve the efficiency and outcomes of future NEPA reviews. To encourage developers to build AI tools with this objective, we release a text corpus of NEPA PDF documents, **National Environmental Policy Act Text Corpus (NEPATEC1.0)**. NEPATEC1.0 consists of textual data extracted from more than 32,000 EIS documents associated with 3,893 projects reviewed under NEPA. This textual data consists of page-wise content from each of the documents and a set of named entities flagged from the page-wise text. In addition, we organize the documents by the level of projects and enrich with metadata (e.g., project title, agency, and location).

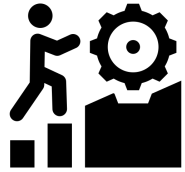
1 Introduction

The National Environmental Policy Act of 1969, as amended (NEPA), is a bedrock and enduring environmental law in the United States with the express intent of fostering a productive harmony between humans and the environment for present and future generations. The NEPA statute (42 U.S. Code 4321 et seq.) and implementing regulations of the Council on Environmental Quality (40 Code of Federal Regulations parts 1500 through 1508) establish procedures requiring all federal agencies to consider environmental effects in their planning and decisions and to inform the public. As a first step, federal agencies must determine whether NEPA applies to a proposed action and then determine the appropriate level of environmental review. A categorical exclusion is the most basic level of NEPA review and addresses those categories of actions that do not individually or cumulatively have a significant effect on the environment. An environmental impact statement (EIS) is the most detailed level of NEPA review and is required for major federal actions with significant environmental effects. If it is unknown whether a proposed action has the potential to have a significant effect on the environment, and agency must first prepare a more concise document called an environmental assessment (EA) to support its determination (Figure 1).

Each type of NEPA review requires preparation of a written document disclosing relevant information that supports the agency’s decision-making process. Recent changes to NEPA now limit EAs to 75 pages and EISs to 150 pages, excluding citations, appendices, and information displayed graphically. Historically, most EISs have been substantially longer. Average document length for EISs sampled by the Council on Environmental Quality from 2013 to 2018 was 575 pages for draft documents and 661 pages for final documents (excluding appendices, which accounted for, on average, another 584 pages and 1,042 pages, respectively) [1]. An agency typically begins the NEPA process after determining

Preprint. Under review.

Community Outreach



Generative AI Challenge for Environmental Review

Kaggle competition: “LLM for Environmental Review”

Dataset: NEPAQuAD1.0

Task: Develop LLM for Question Answering

Start Date: **May 15, 2024.**

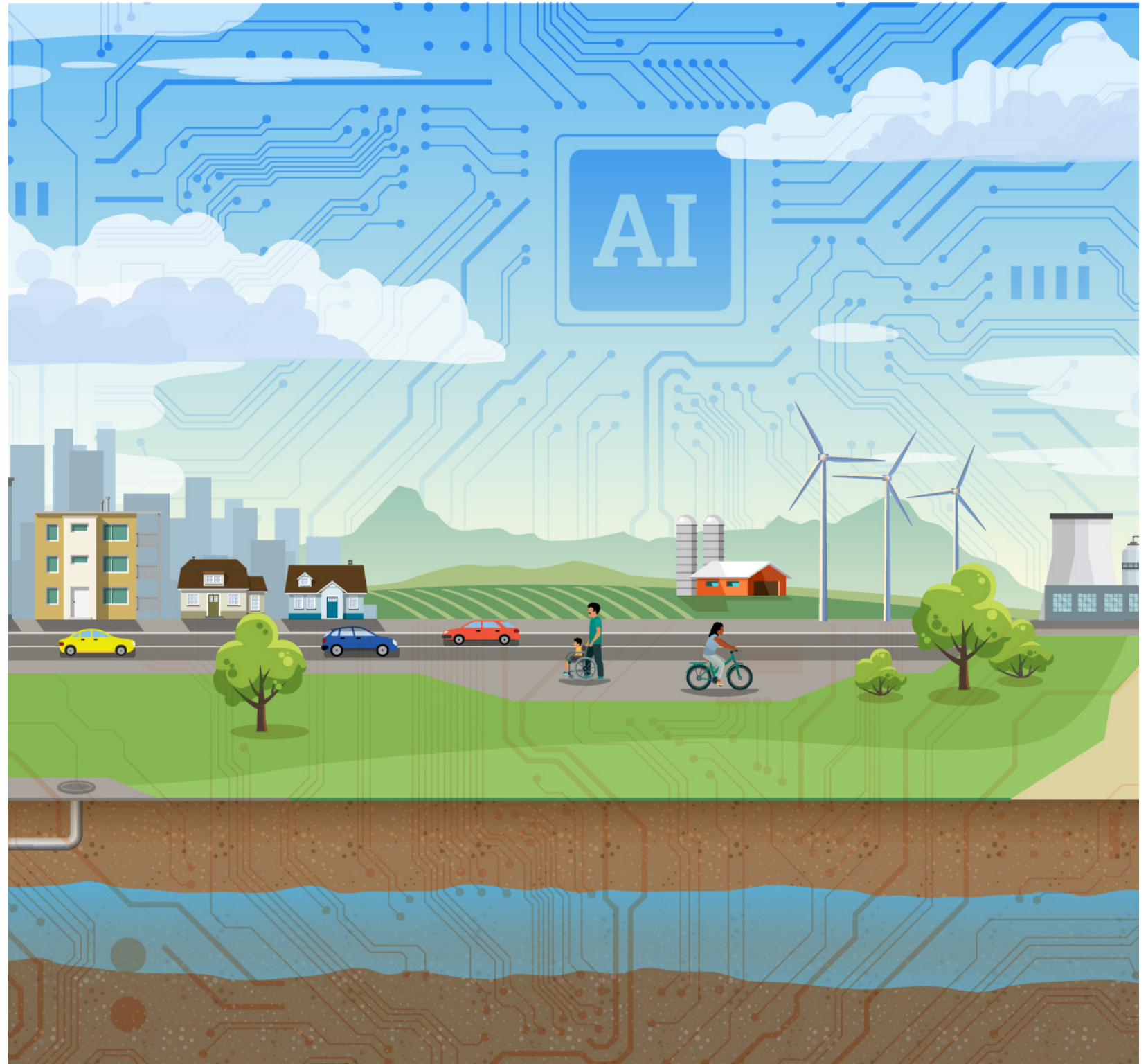
End Date: June 30, 2024



PolicyAI

Acknowledgement

This work was supported by the Office of Policy, U.S. Department of Energy, and Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RLO1830.



Thank you

