SAND2022-9525C

A Natural Language Processing Tool to Support Stakeholder Engagements During a Consent-Based Siting Process

Thushara Gunda

Matthew Sweitzer

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

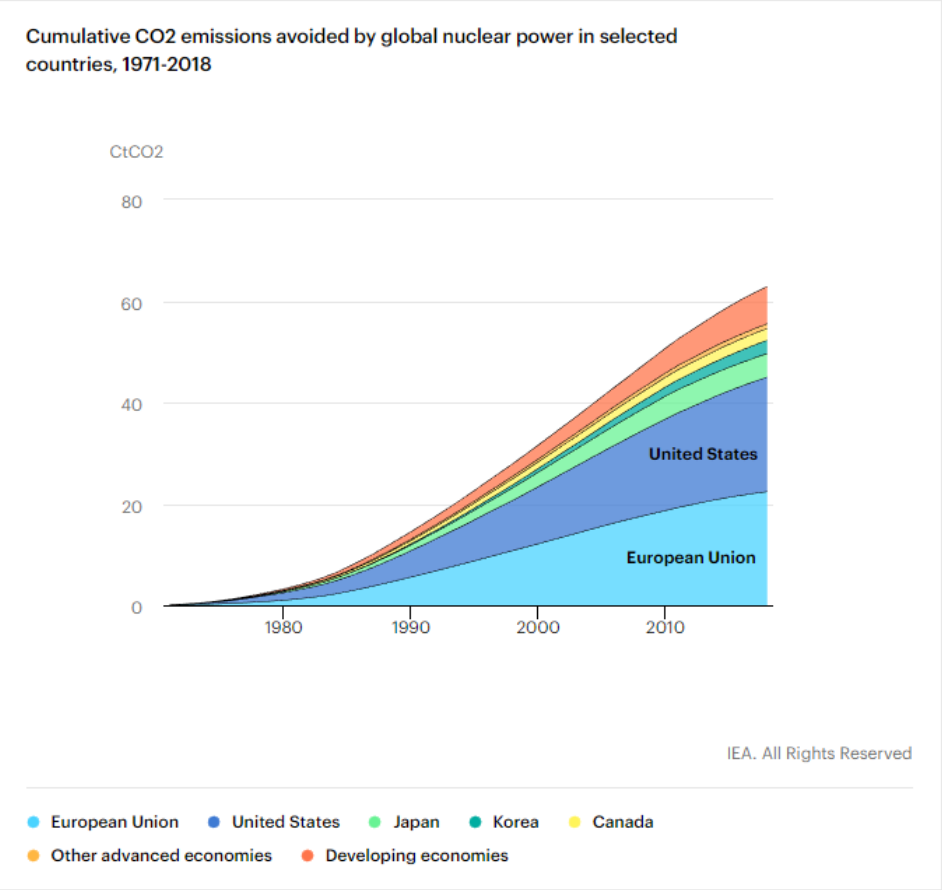
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

- **Background**
  - Motivation
  - Study Objective
- **Methods**
  - Datasets
  - NLP Techniques
- **Findings**
- **Discussion**

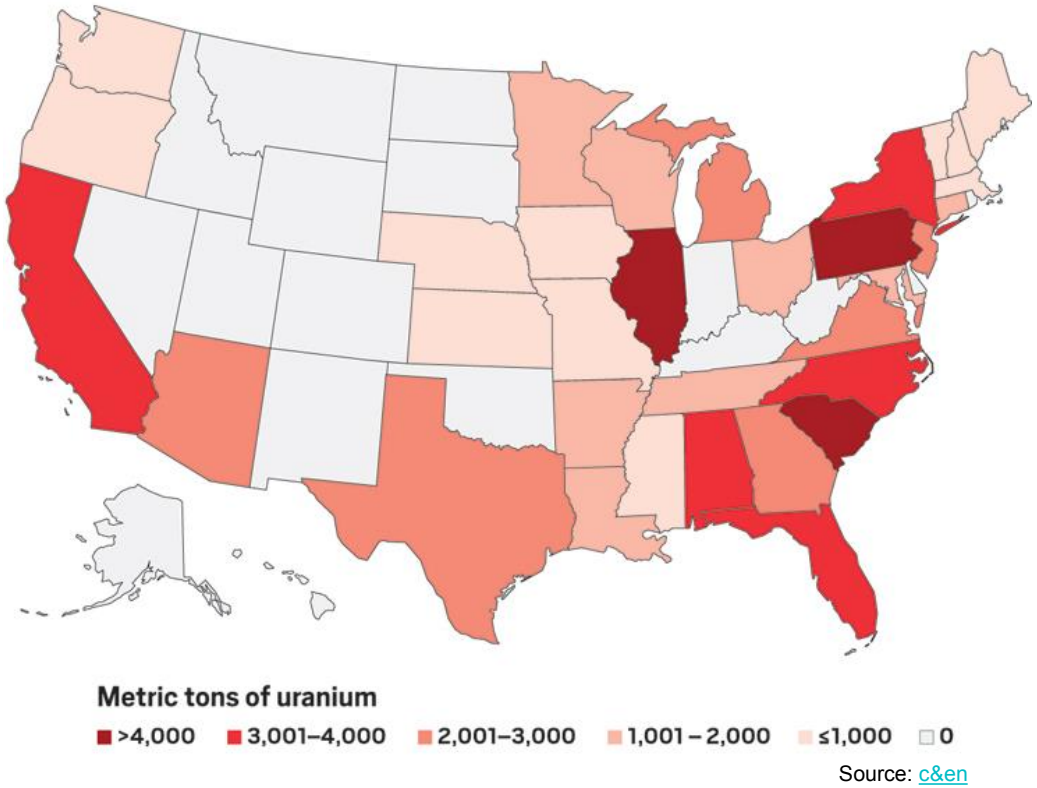
Nuclear Energy is a Leading Low-Carbon Electricity Generation Source

- **Nuclear energy generation accounts for 10% of the global electricity supply**
  - US: 20% of the nation's electricity supply
- **Use of nuclear power has reduced CO2 emissions by 60+ gigatonnes**



Management and Disposal of Nuclear Waste

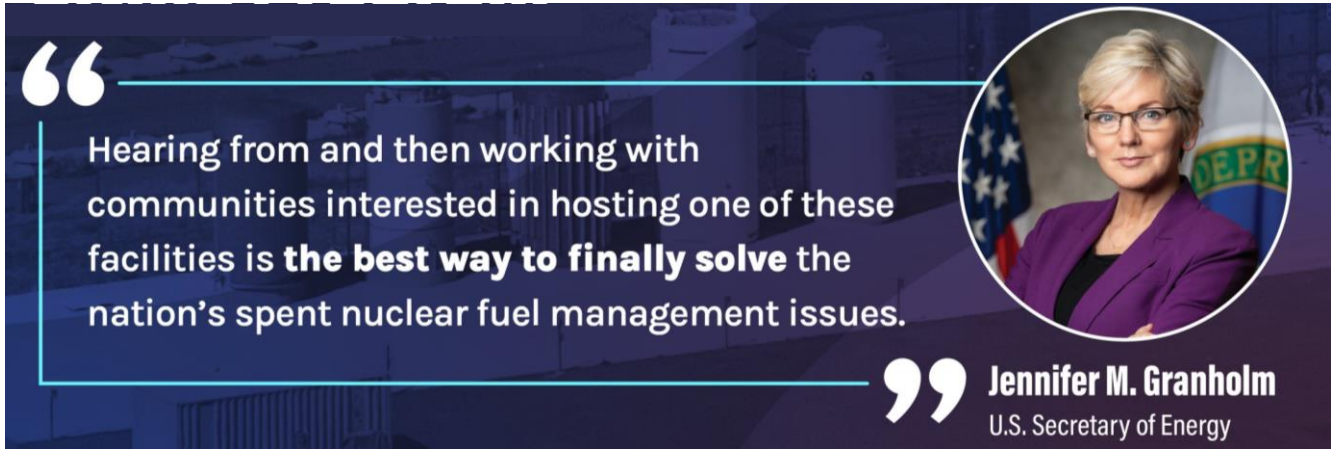
- **The Department of Energy has the management and implementation authority for nuclear waste in the United States**
- **Currently, stored in temporary containers on-site of nuclear power plants**
  - 85K+ metrics tons of from commercial nuclear power plants in the US
  - Growing at 2K metrics tons/year
- **Deep geological disposal as a long-term solution**



Source: c&en

Consent-Based Siting Process

- Recognizes community-driven perspective is required to ensure equitable and just management of nuclear waste
- Collaboration between government officials as well as the public and interested groups
- Responses to RFIs have provided insights into general awareness of communities and a number of relevant organizations to engage



Understanding Local Narratives

Communities are Complex Systems

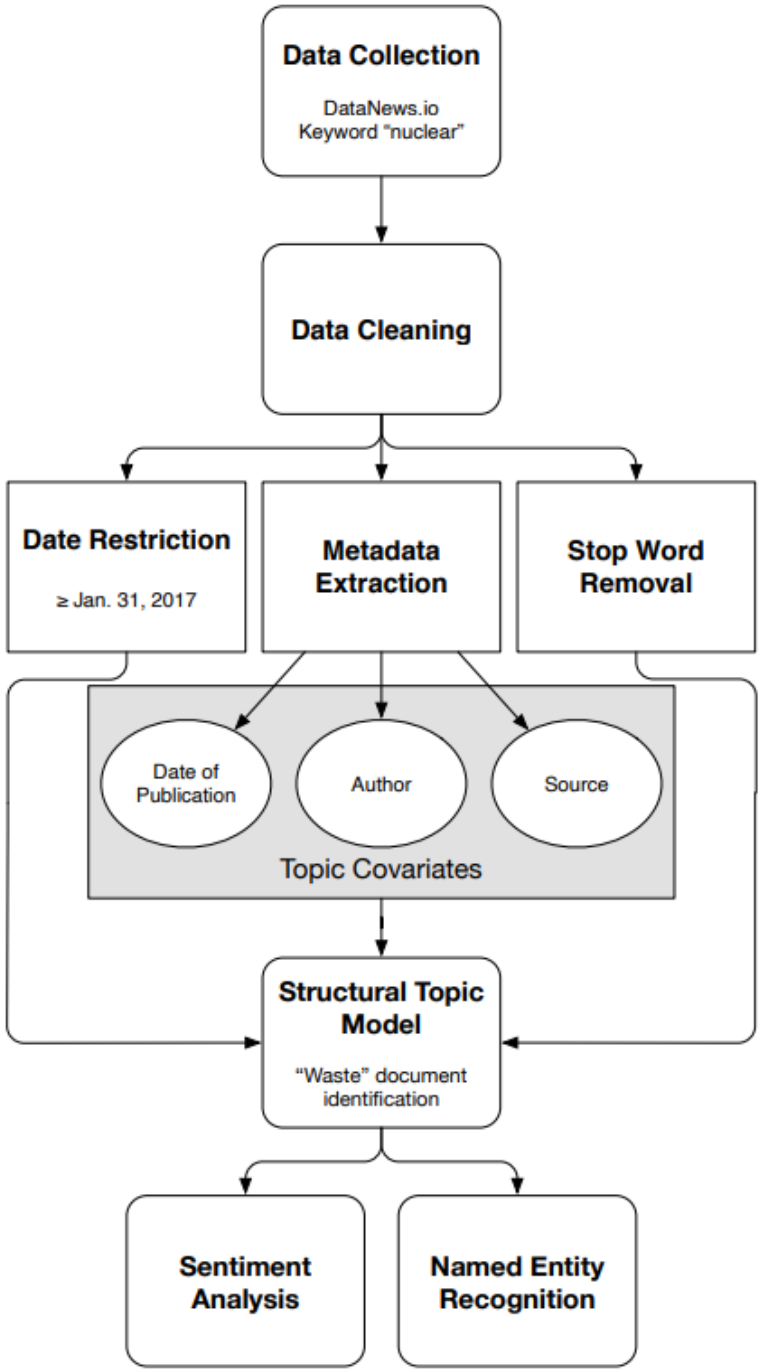
Source: First5

- **Common "sensors" used to gauge community narratives include:**
  - Public opinion surveys
  - Observational studies
  - Journalism analysis
- **Data with meaningful size and depth of context require the use of computational analyses in order to be interpreted**

Study Objectives

Source: Pixabay

- Leverage natural language processing (NLP) techniques to evaluate nuclear waste narratives within local newspapers
- Identify opportunities to incorporate community priorities and concerns into ongoing dialogues



General Methodology

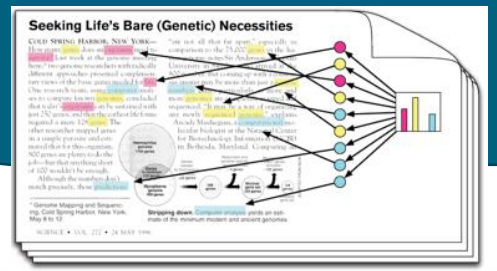
- **Downloaded articles for 5 years (Feb 2017-Jan 2012)**
  - Case-insensitive
  - United States source restriction
- **All data collection and analysis conducted in open-source software R**
- **Human-Machine Teaming: Combines strengths of reliability of computers with interpretability by humans**
- **Final Corpus: 148K articles with 4.8K unique sources**

NLP Approaches

Structural Topic Models

stm

- Unsupervised approach for finding co-occurring words
- Includes covariates and metadata within text analysis
- Particularly well-suited for open-ended textual data
- Used to support filtering of newspaper articles to those focused on nuclear waste

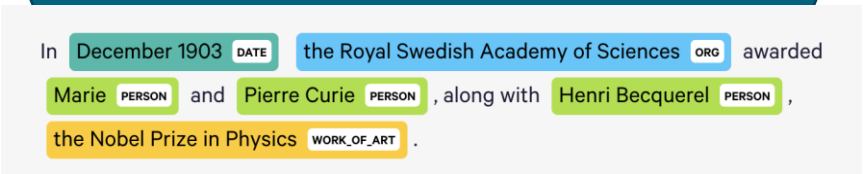


Source: Blei2002

Named Entity Recognition

spacyr

- Locates and classifies entities into pre-defined categories
- Used to support searching and summarization of content
- Used spaCy implementation: based on a convolutional neural network



Source: Medium

Sentiment Analysis

sentimentr

- Extracts and quantifies the polarity of affective language
- Used to understand customer content and generate recommendations
- Leverage existing lexicons with pre-defined categorized terms and assoc. polarities
- Accounts for (de)amplifiers, negators, & adv conjunctions

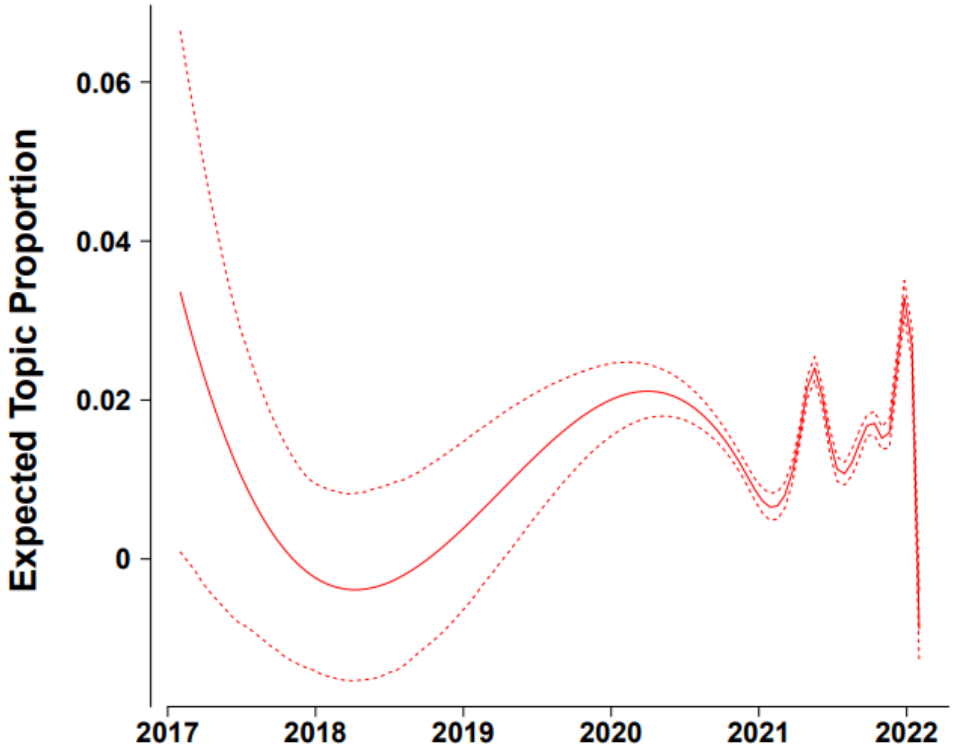
Source: Apptentive

Nuclear Waste Articles using STM

- **Topic #70 grouped a number of terms most relevant to nuclear waste siting issues**

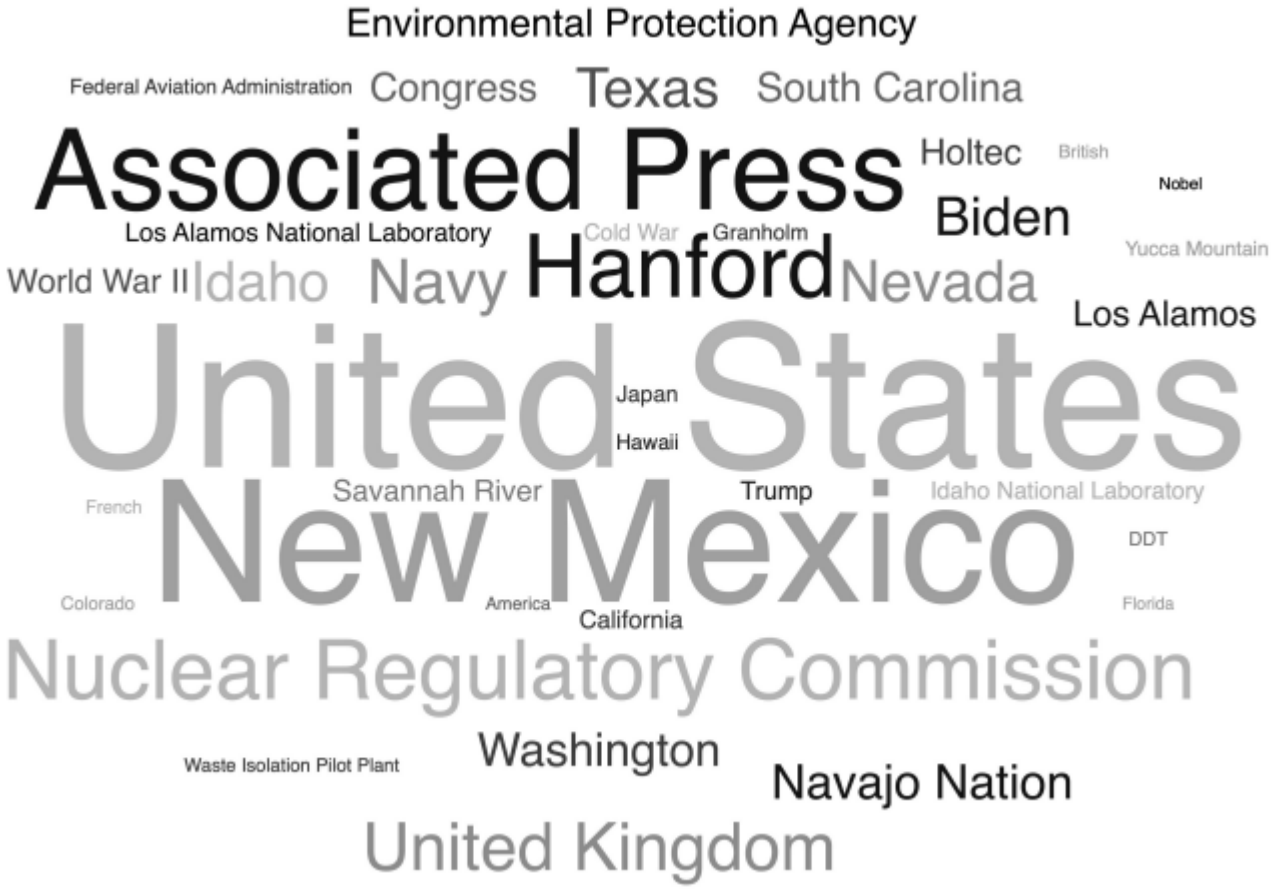
  “nuclear” “waste” “site” “energy” “aquifer” “wipp” “microreactor” “repository” “gallons”

- **Coverage of nuclear waste issues ebbed and flowed over the last 5 years, but reached a peak in late 2021**
- **Waste subset: 2.5K articles from 370 sources**



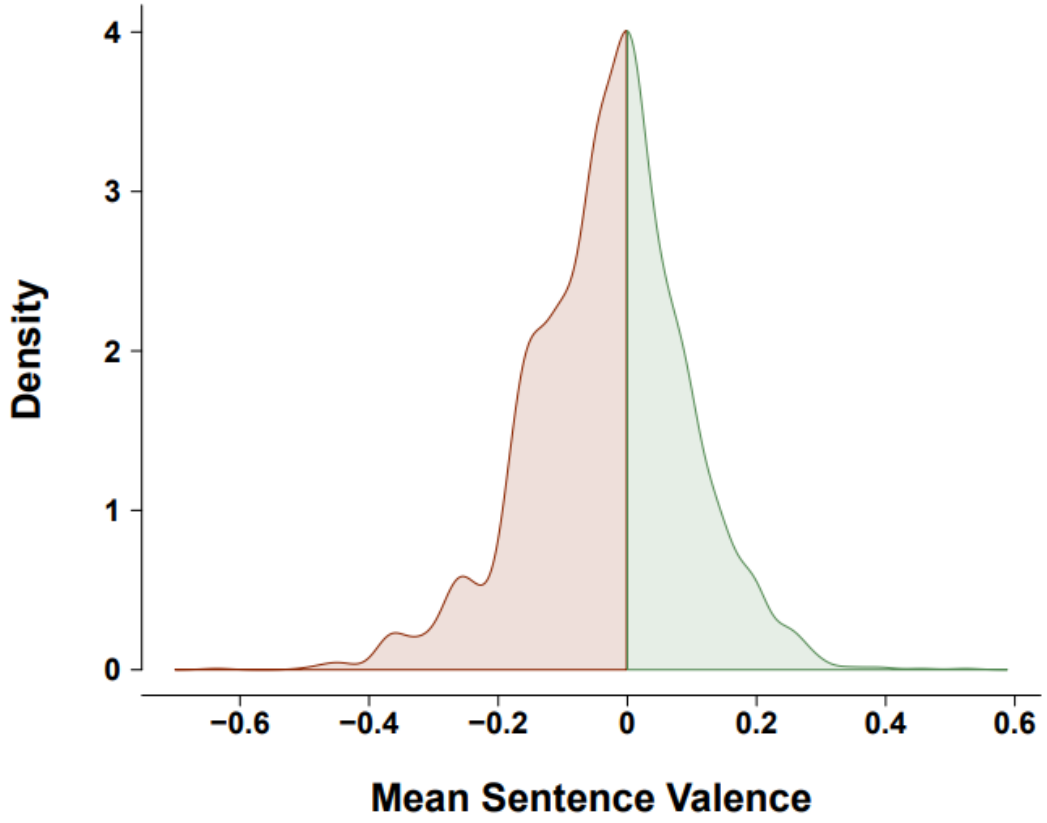
Frequently Occurring Named Entities

- **Commonly occurring entities**
  - Specific geographies
  - Organizations
  - Persons
  - Historical events



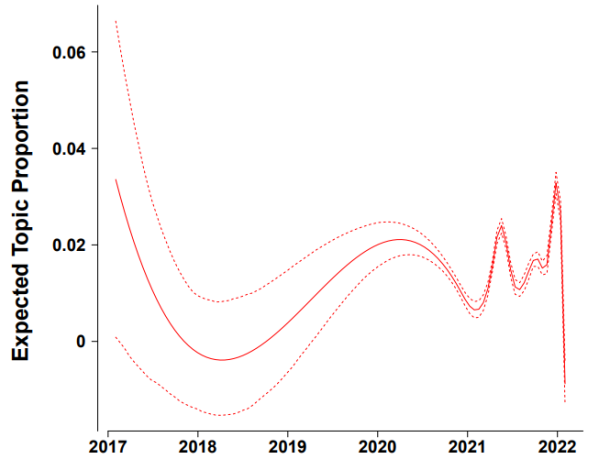
Sentiment Analysis

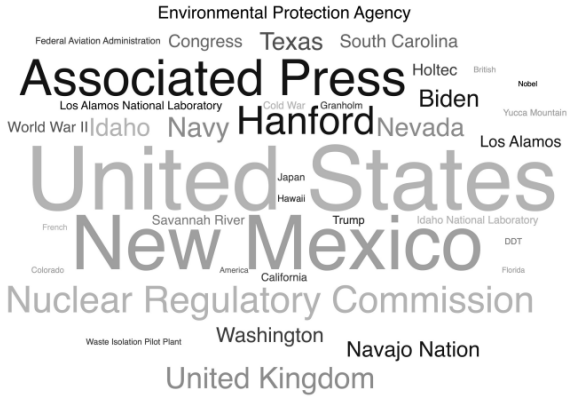
- **Vast majority of nuclear waste articles generally use neutral language**
- **Tendency to use of negative valenced words compared to positive ones**
  - M = −0.030, SD = 0.123
  - Statistically significant difference

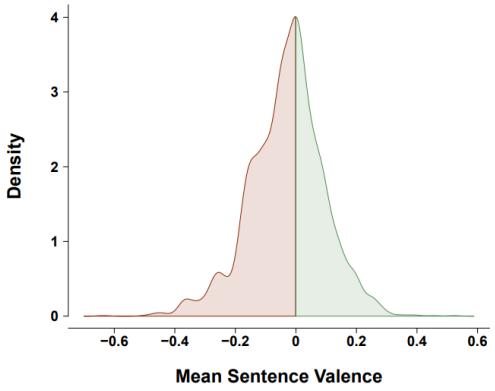


Conclusion

- **Demonstrated significant potential in application of NLP methods to gain insights into narratives around nuclear waste**
  - Cyclical patterns
  - Common entities
  - Affective language
- **Ongoing work**
  - Extend analysis
  - Expand datasets



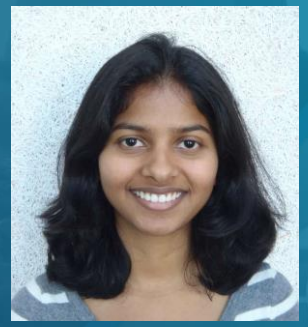




Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly-owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

U.S. DEPARTMENT OF ENERGY | Office of NUCLEAR ENERGY

Thank you for your Time!



Thushara Gunda
tgunda@sandia.gov

Matt Sweitzer
msweitz@sandia.gov