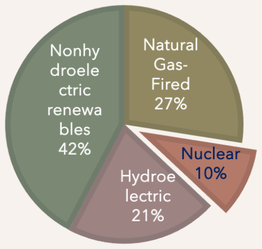
Natural Language Processing of Safety Reports in Nuclear Power Plants

Adrianne Nguyen[1], Rahul Nair[1], Shourya Kothari[1], Tejaswi Chaudhari[1], Xiang Li[1], Najmedin Meshkati[1*], Yolanda Gil[1*], Ulf Hermjakob[1*]

[1]University of Southern California, *Faculty Adviser

Introduction

Nuclear Generated Electricity in CA



3M+

Est. $1.1B

Governor Gavin Newsom signed SB 846, called for **extension until 2030**

State Legislators and Newsom approved a **$1.4B loan** to maintain operation at the plant until 2030

CA.Gov; DCISC; EIA.Gov; INPO

Nuclear Power Plant Safety Matters

Chernobyl 1986

Fukushima 2011

DCISC[1] Annual Reports

INPO[2] Traits of a Healthy Nuclear Safety Culture

Goals

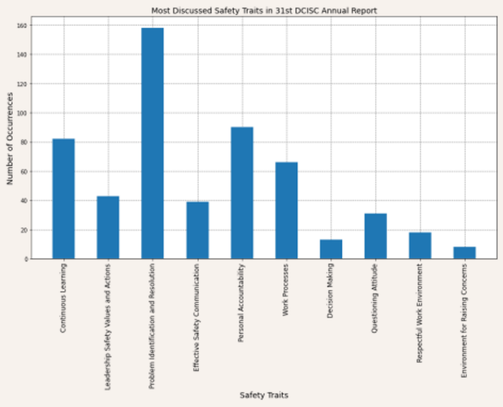
Analyze the nuclear safety culture at DCPP by employing the INPO 12-012 Traits of a Healthy Nuclear Safety Culture to identify whether instances of the traits and their attributes can be found in the DCISC 32nd Annual Report

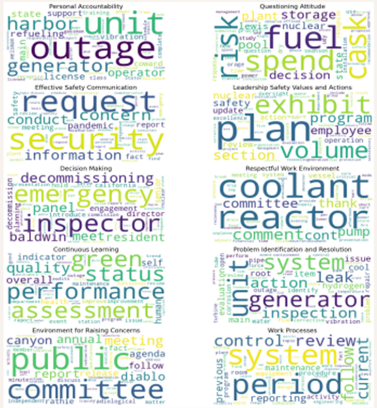
Utilize NLP and clustering techniques to develop a model that can identify traits and can extract valuable information from any DCISC annual report and apply the model to any nuclear report

Create a dashboard to summarize safety-related issues from a report with their associated safety traits, aiming to help identify key improvement areas in terms of safety traits

Results from the 31st DCISC Annual Report

	Accuracy	Precision	Recall	F1-score
Prototype Model	36%	46%	54%	48%





Alkatheeri, A., Neshatvar, Y., Saxena, S., Srinivasan, S., Jain, R. G., Phalak, A., Vatsaraj A., Gordo, I., Wu, S. (2023). Technical Report - NLP for Nuclear Safety.

Milestones

1

Data Cleaning &Preparations

- Manually read through 900+ 32nd DCSIC Annual Report to extract key insights including:
  - Report structure
  - Issue statements
  - Identify the content type (scanned pdf, or text pdf)
  - Key sections for sentiment analysis
- Manually classify issue statements based on INPO 12-012 Traits of a Healthy Nuclear Safety Culture for 'Golden Standards'
- OCR (Optical Characteristic Recognition) for scanned pdf pages

2

Baseline Model Selections

- KNN-clustering
- Random Forest
- Binary Relevance

3

LLM & Advanced Models Selections

- APIs: ChatGPT 3.5 or Llama 2
- Local fine-tuning: Select from Hugging Face
- Neural Network & Deep Learning Models

4

Dashboard Design

- Inputs: DCISC Annual Report or issue statements
- Processing: Utilize finalized models to classify issue statements
- Outputs: Network and statistic visualization of classified traits associated with each issue statement