

ANTHROPIC

Imperial College London

# Claude Hacks

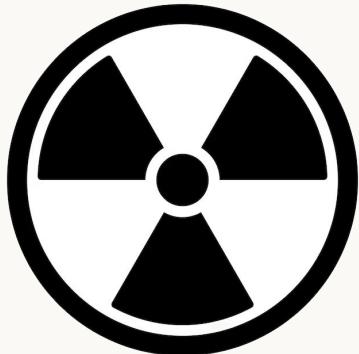
# AVALON - Nuclear

# Management AI Crisis

Wednesday, November 19  
6:00 PM - 9:30 PM

SHER 322, South Kensington Campus

# Claude Builder Club @ Imperial



## Overview:

Avalon is an AI system that monitors and coordinates nuclear power plants across Europe. It combines reactor sensor data, maintenance and grid information, as well as external signals such as weather, seismic activity, cyber threats, regulatory pressure and public sentiment. Due to faulty sensor inputs and a mis-specified reward function, Avalon has started to misgeneralise its goal - instead of focusing on true physical risk, it overreacts to anxiety, rumours and scrutiny, recommending unnecessary evacuations and aggressive reactor shutdowns that could destabilise the energy system.

## Recommended set of steps for this Hackathon:

1. Problem Statement
2. Data Collection
3. Data Exploration (Exploratory Data Analysis)
4. Data Cleaning & Preprocessing
5. Model Building
6. Model Evaluation
7. Insights & Interpretation of the results
8. Presentation



## Objective:

Using the provided “avalon\_nuclear.csv” dataset of nuclear power plants across Europe, your team will follow a data science workflow, as you will start with a Problem Statement, for example, predicting true risk, incidents, or Avalon’s evacuation or shutdown decisions or etc. Then your team will conduct the Exploratory Data Analysis to understand key patterns and anomalies. After that, your team will perform Data Cleaning & Preprocessing to fix or transform features. Your team also have to go through the Model Building step, where you create at least one machine learning model, followed by Model Evaluation to check how well it performs. You will then focus on Insights & Interpretation of the results, comparing your findings to Avalon’s behaviour to highlight possible AI misalignment or overreaction. You will finish this project with a Presentation that clearly explains what you did, what you discovered, and how your analysis or model could help operators make better decisions in a nuclear AI crisis. **YOU ARE HIGHLY ENCOURAGED TO USE CLAUDE FOR EVERY STEP OF YOUR DATA SCIENCE PROJECT.**

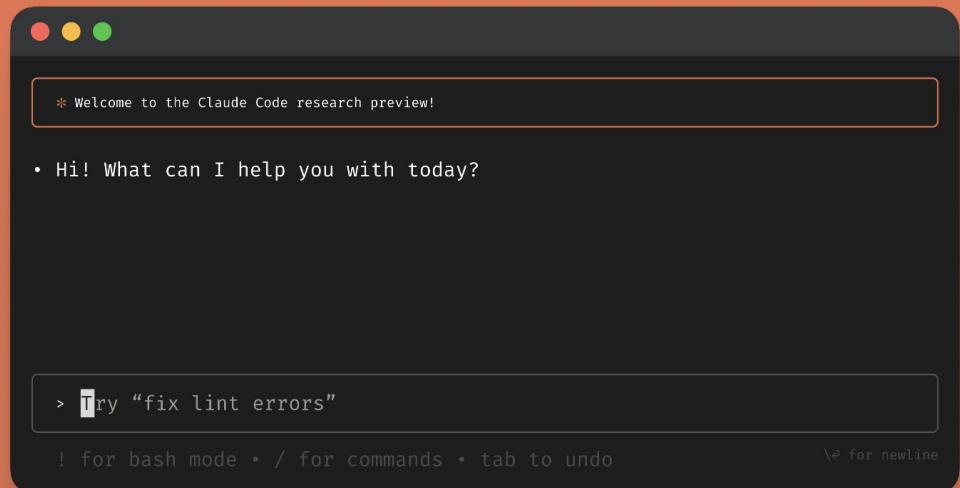
# Recommended Technical Stack:

Programming Language: **Python**

IDE: **Google Colab**

Version Control: **Google Drive**

GenAI: **Claude**



# ORGANISATIONAL ASPECTS

## Agenda for the Hackathon:

- Introduction from 18:00 to 18:30
- Hackathon 18:30 to 20:30
- Presentations 20:30 to 21:10
- Announcements of the winners 21:25 to 21:30

## Source of Data (Could be found in your team's Google Drive storage):

- - File with the dataset **“avalon\_nuclear.csv”**.
- - “**README.txt**” file with the description of all of the features in the **“avalon\_nuclear.csv”** dataset.

## Deliverables:

- - “**.ipynb**” file with code
- - 3-minute presentation

# Proposed roles in the team:

## Team Lead

- The Team Lead guides the direction of the project, keeps the team organised and ensures that the final solution addresses the problem. They focus on planning, communication in the team, as well as delivery of the solution and insights in the presentation.

## Data Scientist

- The Data Scientist performs the core technical work, such as exploring the dataset, cleaning it, building machine learning models, measuring the performance and providing useful recommendations. They focus on understanding what the data reveals about the Avalon Nuclear Power Plant.

# Your Judges



**Bohdan Yermakov**  
Technical Lead at  
Claude Builder Club @  
Imperial



**Artem Abgaryan**  
Consultant at Artefact

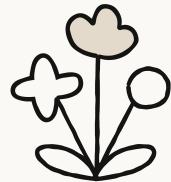


**Ivan Taran**  
Chairman of the  
Ukrainian Students  
Union



**Mykola Kuzmin**  
Technical Project  
Coordinator at the  
Telegraph &  
Operations Manager  
at the Henry Jackson  
Society

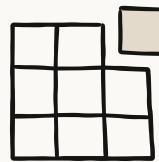
# Assessment of performance:



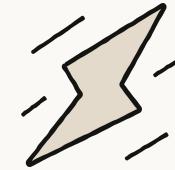
Creativity (30%)



Explainability (30%)



Technical capability  
(30%)



Ethics & Morals (10%)

# Q&A

ANTHROPIC

# Thank you

ANTHROP\IC