



# BHARATIYA ANTARIKSH HACKATHON

2024

---

presented by

---

Team Suryakiran

# PROBLEM STATEMENT

Voice-Enabled User Interface for Geospatial Map-Based Web Applications

## Core Requirements

- **Intuitive Interaction** : Create a voice-enabled interface for natural geospatial map interaction.
- **Scalability** : Develop a lightweight, scalable solution using on-device GPUs/NPUs.
- **Integration** : Ensure seamless compatibility with Leaflet and OpenLayers.
- **Accessibility** : Design for user-friendliness and accessibility.

# PROPOSED SOLUTION

---

- **Speech-to-Text Conversion** : Converted user command to text using Whisper ASR model.
- **Sentence Parsing** : Used a fine-tuned Named Entity Recognition (BERT) model trained on our custom dataset to extract location, action, and layer from the transcription of the user command.
- **Layer Description Matching** : Match extracted layer names with descriptions from a file using an LM sentence similarity model.
- **Command Execution** : Execute the most accurate command on the geospatial map for visualization and interaction.

## Whisper ASR model

- Lightweight and runs on modern device hardware.
- Suitable for offline or low-connectivity scenarios.
- High accuracy and adaptable to various accents.



## BERT NER model

- Context-aware and robust in handling diverse command structures.
- Enables precise identification of user intent from natural language commands.



## Sentence Similarity model [MinILM]

- Match extracted layer names with predefined layer descriptions.
- Calculates accuracy score to determine the best match using cosine similarity.
- Ensures the most relevant layer(s) is selected for display.

# INNOVATION HIGHLIGHTS

---

- Majorly client-side processing.
- Fine-tuned model with 96% accuracy on single layer multi-action multi-location commands.
- On-device execution leveragings CPUs.
- Trained on custom dataset with 1000+ map commands.

# FUTURE WORK

---

- Act on confidence pre-calculated by the BERT model.
- Commands having any word with accuracy less than threshold value are sent to LLM (Llama-3) for token classification. So complex commands are easily executed.
- Increase sentence similarity calculations (layer matching) speed.
- Work on responsive (TTS system).



THANK YOU