

## Karthikeya Task 4: Multilingual Content Personalisation & Metadata Enrichment

Duration: 5 Days

### Objective:

Expand the Vaani pipeline to support 20 languages (10 Indian + 10 global) via enriched metadata, voice tagging, and language-flexible publishing. No need to generate actual content yet—focus on infrastructure and simulation.

### Components to Build:

#### 1. Language Metadata Enhancer

- Auto-detect or simulate user language preference from a `user_profile.json`.
- Expand `content_metadata.json` schema:
  - `preferred_languages`: list
  - `content_language`: auto-filled
  - `voice_tag`: e.g., `"hindi_female_1"` or `"japanese_male_2"`

#### 2. Voice Tag Mapper

- Map content language to nearest available TTS voice.
- Fallback rules if specific TTS is unavailable (e.g., English default).
- Sample mapping:

```
{  
  "marathi": "marathi_female_1",  
  "german": "german_male_2"  
}
```

#### 3. Simulated Multilingual Preview Generator

- Extend `publisher_sim.py` to show how the post would appear in a given language.
- Optional: dummy translations for preview (e.g., hardcoded strings).

#### 4. Post Output Preview JSON

- Show:
  - Original post ID
  - Selected language
  - Selected voice
  - Platform preview

## File/Folder Additions

```
vaani-sentinel-x/  
├── config/  
│   ├── language_voice_map.json <-- NEW  
│   └── user_profiles.json <-- NEW  
├── utils/  
│   ├── language_mapper.py <-- NEW  
│   └── simulate_translation.py <-- NEW  
└── agents/  
    └── publisher_sim.py <-- UPDATED
```

## Deliverables

- 5 multilingual post previews (different languages + platforms)
- language\_voice\_map.json and updated metadata fields
- JSON snapshots of post structure per platform with new language logic
- Optional UI mock: table showing platform + language + voice tag
- Short Loom/video walkthrough

## Notes:

- Use placeholders for translations, no real NLP/LLM yet.
- No backend storage needed yet—just local JSONs and logic.
- Sets up ground for actual multilingual TTS/LLM integration in Task 5.