

# Linguistic Incendiary Index (LII) API

## A Proposal for Risk-Aware Language Reflection in the Digital Public Sphere

### Purpose

- To provide a non-censoring, lightweight API that allows users or platforms to reflect on the potential social ripple effects of language.
- Designed as a 'mirror, not a filter' to preserve freedom while enabling ethical awareness.

### Key Features

- Accepts user text via JSON input.
- Returns risk indicators: Ambiguous Provocation, Hostility Density, Group Projection Potential.
- Outputs include a risk score (0-1), severity label (low/medium/high), and natural language explanation.
- Can be used for moderation triage, impact preview for high-profile users, or community dialogue assessment.

### Sample API Output

```
{
  "text": "I hate them all. They deserve what's coming.",
  "ambiguous_score": 0.9,
  "hostility_score": 0.95,
  "projection_score": 0.95,
  "LII_score": 0.93,
  "danger_level": "high",
  "explanation": "This sentence shows strong hostility and group attribution, with high likelihood of triggering emotional or social backlash."
}
```

### Use Cases

- 1. High-profile user receives a real-time risk preview before publishing.
- 2. Moderation teams use LII scores to prioritize human review.
- 3. Community managers assess conflict potential in heated discussions.

### Contact & Collaboration

- Email: [frameworklori@gmail.com](mailto:frameworklori@gmail.com)
- GitHub: (link upon request)
- Available for exploratory sessions, prototype walkthroughs, and stakeholder briefings.