



# **StreamGuardian** –

## **Real-time Toxic Comment Detection System**

---

*Presented By– Nasim Raj Laskar*

# Problem Statement

Online platforms, especially during live streaming or community chats, suffer from a barrage of toxic messages — harassment, hate speech, and offensive content.

Manual moderation:

- Is inefficient
- Can't scale with large volumes
- Is reactive, not proactive

**StreamGuardian** solves this by detecting and visualizing toxic messages in real-time, empowering moderators and platforms to act immediately.

# Tech Stack Overview

## Programming Languages:

- **Python** – For backend logic, data processing, and model integration.
- **JavaScript** – For frontend interactivity and chart rendering.
- **HTML/CSS** – For UI layout and styling

## Libraries:

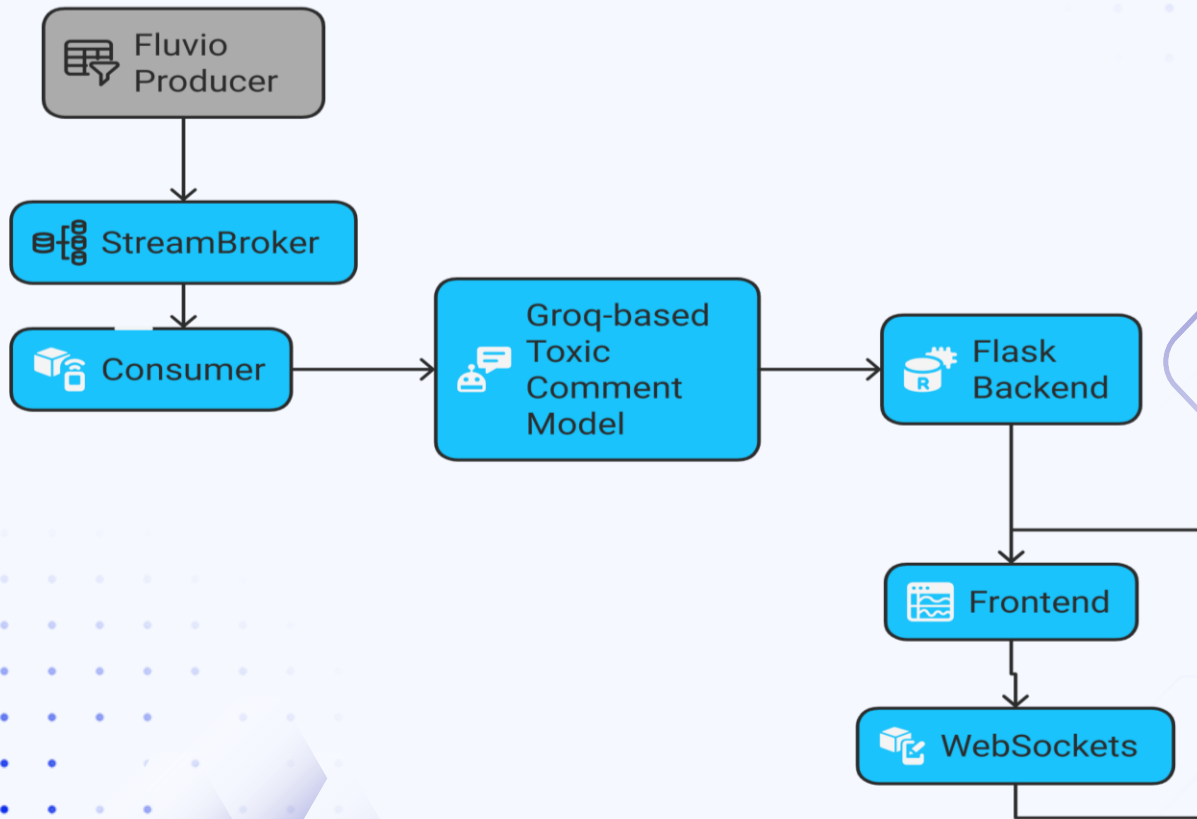
- **Python:**
  1. **Flask** – Lightweight backend framework
  2. **Fluvio Python SDK** – For data handling and message processing
- **JavaScript:**
  1. **Chart.js** – For dynamic data visualizations.
  2. **Socket.IO** – For real-time WebSocket integration

## Development Tools:

- **Fluvio** – For real-time data streaming and message brokering
- **Groq** – For ultra-low-latency toxic comment prediction
- **VS Code** – For writing and managing project codebase
- **Postman** – For testing backend endpoints



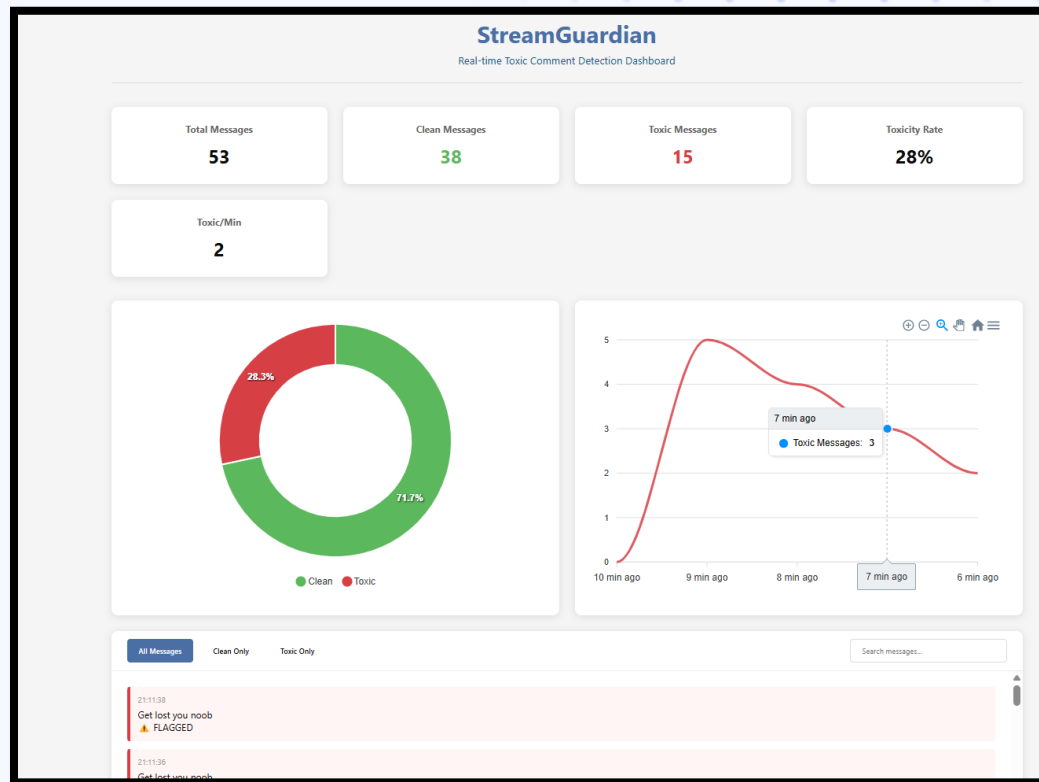
# System Architecture



# Dashboard Walkthrough

## UI Features:

- **Summary Cards:** Display total messages, clean messages, toxic messages, toxicity rate
- **Donut Chart:** Visual representation of clean vs toxic split
- **Line Graph:** Time-based toxic message count
- **Tabs:** Filter view by "All", "Clean Only", or "Toxic Only"
- **Live Feed:** Toxic messages flagged with emoji ⚠️
- **Export Options:**
  1. Download CSV for message logs
  2. Export PNG/SVG versions of charts for Sharing reports



Dashboard preview

# Key Features



## Real-time Streaming

Detect and analyze toxicity as it happens.



## Groq-powered Inference

Lightning-fast toxic comment classification.



## WebSockets Refresh

Messages update live — no reload needed.



## Dynamic Charts

Track trends with visual clarity.



## Message Flagging

Toxic messages marked clearly with warning icons.



## Filtering Tabs

Quickly isolate clean or toxic content.



## Download Support

Export logs as CSV, download visualizations as PNG/SVG.

# Future Improvements

