

Concept Note: – ShieldAI

Technology-Facilitated Gender-Based Violence (TFGBV) Detection Browser Extension

1. Introduction

Technology-Facilitated Gender-Based Violence (TFGBV) is rapidly increasing across digital platforms. Many victims cannot identify early signs of coercion, harassment, and abuse. ShieldAI is a privacy-first Chrome extension that analyzes text for TFGBV in English, Amharic, and Afaan Oromo. It provides classification, severity scores, and downloadable evidence reports.

2. Objective

- Help users recognize online abuse.
- Provide safe documentation of incidents.
- Support digital safety awareness.
- Enable NGOs and support centers to use structured reports.
- Ensure accessibility using offline detection.

3. Target Users

Primary:

- Women and girls experiencing online harassment.
- Teens/students facing cyberbullying.
- Social media users who receive abusive messages.

Secondary:

- Parents and guardians.
- Schools, universities, counselors.
- NGOs and GBV support organizations.
- Law enforcement partners.

4. Key Features

- Multilingual TFGBV detection (English, Amharic, Afaan Oromo)
- Offline local classifier (no internet required)

- Optional AI enhancement via OpenRouter/OpenAI
- Real-time classification (Safe, Uncertain, Abusive)
- Severity scoring and analysis reasons
- JSON evidence report generator
- Privacy-first (no monitoring, no screenshots)

5. Implementation Pathway

Phase 1 – Prototype:

- Chrome Extension MV3
- Multilingual classifier
- JSON report generator

Phase 2 – AI Integration:

- Optional cloud-AI mode
- Auto-language detection

Phase 3 – UX Enhancements:

- Color-coded severity
- History tracking (local)

Phase 4 – Deployment:

- NGO partnerships
- School pilot programs

6. Potential Scalability

- Add more languages (Tigrinya, Somali, Arabic, Swahili)
- Expand to Firefox/Edge and mobile keyboard app
- Integrate with Telegram/WhatsApp bots
- NGO training modules
- Local ML fine-tuning

Conclusion

ShieldAI is a scalable, multilingual, privacy-first solution that empowers users to identify and document TFGBV safely. It reduces barriers to reporting and enhances digital safety for vulnerable communities.