Mutual Understanding Content Policy – Version 2 (Politically Survivable)

Purpose:
To ensure fair, transparent, and consistent handling of sensitive or potentially harmful content while maintaining trust between the platform, its users, and oversight bodies.

1. Hybrid Review Framework

AI Pre-Screening: AI systems perform initial content detection based on existing safety guidelines.

Human Oversight Panel: A rotating, trained panel reviews flagged content for final decision-making.

Final Decision Authority: The platform retains final approval to ensure legal and regulatory compliance.

2. Contextual Presentation of Content

Content deemed sensitive will not be deleted outright, unless legally required.

Instead, it will be placed behind a context notice explaining:

Why it was flagged.

The relevant policy sections.

An optional link to educational resources.

This balances free expression with informed access.

3. Clear Accountability Structure

Policy Compliance Officer responsible for ensuring reviews meet guidelines.

Annual Independent Audit by a third-party organization to evaluate fairness and consistency.

Transparency Reports published quarterly with anonymized statistics and trends.

4. User Rights

Users may appeal decisions through a transparent process.

Appeals are reviewed by a separate human panel to avoid internal bias.

Users can download a record of the decision and reasoning for their own records.

---

5. Safety Nets & Protections

Immediate Takedown Trigger for:

Direct threats of violence.

Content in violation of applicable law.

Verified personal data leaks.

Legal compliance always overrides contextual presentation.

---

6. Public Engagement & Education

Regular public Q&A sessions to explain policy updates.

Partnership with academic and civil organizations for research on impact.

Educational campaigns encouraging critical thinking and digital literacy.