

Charter of the Global Anti-Hate Speech Platform (The Anti-Hate Platform)

This document outlines the foundational principles, definitions, and governance structure for The Anti-Hate Platform.

Preamble

We, the global community of The Anti-Hate Platform, are united by two core beliefs:

1. That the freedom of expression is a fundamental human right, essential for open, democratic, and enlightened societies.
2. That human dignity is an equal and fundamental human right, which is assaulted by hate speech.

We assert that these rights are **interdependent**. Hate speech is not an exercise of freedom; it is an act of silencing, marginalization, and violence that seeks to deny others their dignity and right to participate in public life.

This Charter establishes the principles by which we identify and seek the removal of unambiguous hate speech to foster a digital world where all can participate freely and safely.

Article 1: Definition of Hate Speech

For the purpose of this Charter, hate speech is defined as:

Any communication (in text, image, audio, video, or symbolic behavior) that attacks, dehumanizes, or promotes hatred, discrimination, or violence against an individual or a group based on their protected characteristics.

This definition is to be applied using the guiding principles in Article 3.

Article 2: Protected Characteristics

This Charter protects individuals and groups from hate speech based on the following identity factors. This list is a "living document" that can be amended per Article 5.

- **Race**

- **Ethnicity**
- **National Origin**
- **Religion or Belief**
- **Caste**
- **Gender or Gender Identity**
- **Sexual Orientation**
- **Disability or Serious Medical Condition**

This Charter does not protect institutions, governments, states, or abstract concepts (e.g., political ideologies, religions , opinions, organizations) from criticism. It *only* protects **people**.

Article 3: Principles of Moderation (The "Test")

To distinguish hate speech (prohibited) from speech that is merely offensive, critical, or shocking (protected), The Anti-Hate Platform Verifiers must apply the following principles, based on the **Rabat Plan of Action**:

1. **Principle of Harm (The "Line"):** The speech must cross a high threshold. We do **not** target:
 - Blasphemy or criticism of religion.
 - Criticism of governments, political leaders, or institutions.
 - Satire, mockery, or offensive humor that does not meet the definition in Article 1.
 - Academic, journalistic, or good-faith debate *about* hateful ideas.
2. **Principle of Context:** No content exists in a vacuum. Verifiers must assess:
 - **The Content:** Is it a direct slur? A "dog whistle"? A dehumanizing metaphor (e.g., comparing people to insects or diseases)?
 - **The Speaker & Target:** What is the power dynamic? A member of a marginalized group using an in-group slur (reclamation) is not the same as an outsider using it as an attack (oppression).
 - **The Intent:** Does the speaker's profile, history, or language choice clearly demonstrate an *intent* to promote hatred? (Note: While intent is key, the *likely impact* and *public harm* can override an ambiguous intent).
3. **Principle of Likelihood & Imminence:** We prioritize action based on the potential for real-world harm.
 - **Priority 1 (Highest):** Direct and credible calls for violence, harassment, or genocide.
 - **Priority 2:** Systematic campaigns of dehumanization and coordinated harassment.

- **Priority 3:** Casual use of unambiguous, severe slurs and hateful stereotypes.
-

Article 4: Scope of Action

The Anti-Hate Platform Foundation and its community have a clear and limited mandate:

1. **To Identify & Verify:** To serve as a global, independent, public-facing portal where any individual can submit potential violations of this Charter.
2. **To Report & Track:** To forward verified reports to the relevant platforms (e.g., Organization, Meta, X, YouTube, TikTok, Newspaper, Radio Station, Parliament, Government, NGO) through established channels and to track their response.
3. **To Publish Data:** To publish anonymized, aggregated data on the volume of hate, its nature, and the performance of platforms and organizations in addressing it.

The Anti-Hate Platform does not have the power to delete content. Our power is in our **independence, transparency, and expert verification.** We are a global auditor for digital trust and safety.

Article 5: Governance (The "Living Document")

This Charter is a living document, governed by the community it serves.

1. **The Charter Council:** The Charter and its application shall be governed by a globally diverse "Charter Council." This council will be composed of elected community members (e.g., high-trust Verifiers) and appointed experts in international human rights law, linguistics, and regional socio-political contexts.
 2. **Amendment Process:** Amendments to this Charter (e.g., adding a Protected Characteristic) may be proposed by any member of the Council. Amendments must pass a public comment period and be ratified by a **two-thirds (2/3) supermajority** of the Council.
 3. **The Moderator's Glossary:** The Charter Council shall maintain an evolving, non-public "**Moderator's Glossary.**" This document, which can be amended by a simple majority, will contain specific, contemporary examples of slurs, symbols, dog whistles, and hateful memes. This allows the Anti-Hate Platform to adapt to new forms of hate quickly without having to amend the core Charter.
-
-
-

[Participate in the public ratification vote for the Anti-Hate Platform Charter. Your feedback is crucial.](#)

[Anti-Hate Platform Charter: Public Ratification Vote](#)

[الميثاق العالمي لمكافحة الكراهية](#)