

**\*\*[Draft Bill]**

## **A BILL**

To mandate transparency in online content moderation practices of the largest online platforms, ensuring the publication of detailed data regarding all specific enforcement actions and government communications which will enable third-parties such as academics, media and other companies to measure and publish reports on the performance of the platforms in terms of online safety, viewpoint neutrality and government influence.

### **SECTION 1. SHORT TITLE.**

This Act may be cited as the "Transparent Online Content Moderation Act" or the "TOCOMO Act".

### **SECTION 2. DEFINITIONS**

- (1) LARGEST ONLINE PLATFORMS means any online platform company with more than 1,000,000 unique visitors or users in a calendar month.
- (2) ENFORCEMENT ACTIONS means any actions taken against users for posting or sharing content that violates the platform's published acceptable use rules. Such actions may include: labeling the content, warning to the user, blocking the content, restricting the dissemination or sharing of the content, temporarily suspending the user, banning the user, reducing any monetization or advertising revenue to the user, de-amplifying the algorithmic sharing of the content or any other action.
- (3) GOVERNMENT COMMUNICATIONS means any electronic, verbal, written or other communications from any federal or state government entity, or government-funded entity, with the exception of communications related to actual specific law enforcement or national security issues.

### **SECTION 3. FINDINGS.**

Congress finds the following:

- (1) Opaque enforcement actions with unclear explanations contribute to the public's lack of trust in the content moderation efforts of online search and social media platforms.
- (2) People in the United States increasingly rely on interactive computer services to communicate, gather information, and conduct transactions that are central to our economic, political, social, and cultural life.
- (3) There is a compelling government interest in having providers of interactive computer services provide information to the public about their content moderation policies and practices because of the impact those policies may have on the free speech interests and online safety of their consumers.
- (4) Transparency in content moderation practices will motivate online platforms to improve online safety, ensure viewpoint neutrality and fairness, and strengthen user trust in the online platforms.
- (5) Many other industries transparently publish their performance including examples such as: automobile safety, airline safety and on-time performance,

pharmaceutical safety, food product ingredient safety, financial safety of banks, and employee safety in factories.

(6) The largest online platform companies already publish generalized information and statistics about their content moderation actions, but these reports currently lack the specificity of data needed for measurement and peer-to-peer comparison of the content moderation of online platforms in terms of online safety, viewpoint neutrality and government influence.

#### **SECTION 4. MANDATORY TRANSPARENCY AND PUBLICATION REQUIREMENTS FOR THE ONLINE PLATFORMS.**

The largest online platforms shall each publish a report on publicly accessible website on a quarterly basis using a standard spreadsheet format detailing the following data, statistics and information:

- (a) **ENFORCEMENT ACTIONS** All content moderation enforcement actions, including:
  - (1) The specific content rule(s) violation that triggered the action,
  - (2) The specific category of content affected,
  - (3) the specific enforcement actions taken,
  - (4) The role and identity of any third-party fact-checker in the decision-making process for the action,
  - (5) Any government requests that led to enforcement actions, except in cases of actual and specific national security or law enforcement confidentiality issues,
  - (6) The status and outcome of disputes and appeals related to enforcement actions,
  - (7) Name of user or content provider affected by the enforcement action, provided such user or content provider has opted-in to be named in such reports and agrees to void their relevant privacy protections.
- (b) **GOVERNMENT COMMUNICATIONS.** Platforms must disclose all communications with government agencies or government-funded entities related to content moderation, except those involving specific national security or law enforcement matters that require confidentiality. This would include date of communication, specific content moderation requests, specific name of government entity or government-funded entity, and name of person making such request.
- (c) **FACT-CHECKER BACKGROUND.** For any enforcement action relying on third-party fact-checkers, platforms must publicly disclose the specific enforcement action triggered by a fact-check along with the specific name, affiliations, funding sources, and prior history of the fact-checker(s).
- (d) **CRITERIA AND RANKING OF WEBSITES AS AUTHORITATIVE SOURCES FOR SEARCH RESULTS.**
  - (1) For online search services, the specific criteria for ranking of websites and online sources used as sources for search results will be published.

(2) Any content moderation policy that restricts or affects search results will be published, and users affected by such restrictions on search results will be notified in real-time at the top of the first online page of search results.

(3) Any government communications, as defined by Section 4(b) above or any requests from government or government-funded entities that affect search results will be published.

**(e) USER NOTIFICATION REQUIREMENTS.**

(1) For each enforcement action taken against a user, platforms must provide:

(1a) A clear explanation of the specific rule(s) violated;

(1b) A clear indication of the specific content which violated the specific rule(s), including the specific portion of any lengthy content or video which triggered the enforcement action;

(1c) A description and updated status of the review and adjudication process;

(1d) Steps available for appeal.

(2). For each search result impacted by government communications or other search result modification, the search platforms must include an explanation for the restricted results for the search terms requested by the user.

**SECTION 5. ENFORCEMENT AND PENALTIES.**

**(a) FINANCIAL AND CIVIL PENALTIES.**

(1) Failure to comply with the transparency requirements set forth in this Act may result in civil penalties, the amount of which shall be determined based on the severity and frequency of violations.

(2) Federal enforcement actions may be initiated by the Department of Justice (DOJ), the Federal Communications Commission (FCC), or the Federal Trade Commission (FTC).

(3) Financial penalties shall start at \$10,000,000, double to \$20,000,000 if compliance is not resolved within 90 days of the initial penalty, and further escalate aggressively based on intent, repeated offenses, and the harm caused by non-compliance.

**(b) WHISTLEBLOWER PROTECTIONS.**

(1) Employees or contractors of online platforms who disclose non-compliance with this Act shall be protected from retaliation.

(2) Whistleblowers who provide verifiable information about efforts to evade public transparency requirements shall be eligible for financial rewards funded by the financial penalties resulting from the verifiable information.

**SECTION 6. IMPLEMENTATION AND REVIEW.**

(a) The FTC shall determine the companies to be included in the list of the largest online platforms, reach out to those companies within 30 days, and establish regulations to implement the requirements of this Act within 180 days of its enactment.

- (b) The list of companies will be expanded or updated on a quarterly basis (every 90 days) based on changes in which companies meet the criteria to be included in the list of largest online platforms.
- (c) A public database shall be created as part of the FTC website to host the transparency reports from the largest online platforms, enabling third-party and public analysis and reports based on the information and statistics provided in the transparency reports.

#### **SECTION 7. EXEMPTIONS.**

- (a) This Act does not require the disclosure of proprietary algorithms, software techniques or trade secrets related to content moderation methodologies of the largest online platforms or any other online companies.

#### **SECTION 8. EFFECTIVE DATE.**

This Act shall take effect 180 days after the date of enactment.

#### **SECTION 9. SEVERABILITY.**

If any provision of this Act is found to be invalid or unconstitutional, the remaining provisions shall remain in full force and effect.