Stopping Abuse and Fraud Enabled by Platforms (SAFE Platforms) Act

BY THE CONSUMER FEDERATION OF
AMERICA AND CONSUMER REPORTS

PUBLISHED SEPTEMBER 2026

Consumer Federation of America

Consumer Reports

SUMMARY

Social media and AI-enabled scams, especially those targeting seniors, are on the rise. The FBI's 2025 Internet Crime Report found that Americans reported $20.877 billion in losses from internet scams in 2025 — a 26% increase from the prior year. Older Americans are the most vulnerable age-group to this type of scam, with $7.749 billion in reported losses to the IC3 from individuals age 60 or older with an average loss of $38,500 per victim. The FTC found that social media was the costliest contact method for fraud in 2025 with reported losses over $2 billion. These figures likely understate the true scale of the problem: a Consumer Federation of America report estimated that the total true losses from scams may be as high as $148.2 billion for Americans nationwide.

Criminals are exploiting platform features such as direct messaging, fake profiles, recommendation algorithms, and ad-targeting tools to gain victims' trust and spread deceptive content, exposing seniors and other vulnerable groups to financial loss and emotional distress. Even worse, platforms are turning a blind eye to this conduct, or in some instances, even encouraging and profiting off of it.

The bill creates three pillars of protection: (1) imposing liability on platforms that facilitate fraud, (2) requiring platforms to implement safer features to prevent scammers from reaching users, and (3) providing recourse for victims of platform-enabled scams.

Recent federal rollbacks to consumer protection agencies such as funding cuts, weakened enforcement, or withdrawn rulemakings have made it even harder to hold platforms accountable for scam activity. With weakened federal oversight and fewer consequences, platforms face even less pressure to protect vulnerable communities including seniors from fraud.

Lawmakers have the opportunity to help prevent this proliferation of harm from AI and tech-enabled online scams with the SAFE Platforms Act. This first-of-its-kind legislation was introduced in New York in 2026 (S.9417/A.10561) and is expected to be introduced in more states in 2027.

Scammers continue to exploit platform features to target users — especially seniors, veterans, and other vulnerable communities — with few consequences for the platforms that enable and profit from that abuse. By setting clear standards, the SAFE Platforms Act ensures platforms take meaningful steps to stop fraud before it spreads.

MODEL LEGISLATION

A BILL – To establish liability for online platforms that materially contribute to fraudulent content

SECTION 1. SHORT TITLE.

This Act may be cited as the "Stopping Abuse and Fraud Enabled by Platforms (SAFE Platforms) Act."

SECTION 2. DEFINITIONS.

For purposes of this Act:

(a) "ADVERTISER" means any person or entity that pays a platform to display, promote, or otherwise disseminate content to users of the platform.

(b) "FRAUD" means any representation, omission, or practice in or affecting commerce that:

(1) Is likely to mislead a consumer acting reasonably under the circumstances; and

(2) Is material to consumer decision-making.

(c) "FRAUDULENT CONTENT" means content that constitutes fraud as defined in this Act.

(d) "HAS REASON TO KNOW" means circumstances that would make the fraudulent nature of content apparent to a reasonable platform operator, including:

(1) Receipt of reports substantially complying with the reporting requirements established under Section 3(a) indicating fraudulent activity;

(2) Information from law enforcement, regulatory agencies, or other authoritative sources identifying specific fraudulent content;

(3) Detection systems implemented by the platform, whether automated or manual, flagging content as potentially fraudulent; or

(4) Other specific information brought to the platform's attention indicating that particular content is fraudulent.

(e) “INTERACTIVE COMPUTER SERVICE” means any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.

(f) “INFORMATION CONTENT PROVIDER” means any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.

(g) “MATERIALLY CONTRIBUTE” means conduct that goes beyond providing neutral tools for developing or disseminating fraudulent content, including

(1) Encouraging or inducing a third party to develop or disseminate fraudulent content;

(2) Targeting, recommending, promoting, curating, or prioritizing fraudulent content directed at residents of this State that the platform knows or has reason to know that the content is fraudulent;

(3) Other conduct that the Attorney General determines to be a material contribution through regulations promulgated pursuant to the rulemaking authority conferred under Section 7.

(h) “MUTUAL CONNECTION” means a relationship where two users on a platform are each directly connected to the same third-party user through a formal connection mechanism such as a friend or follow. Shared participation between two users in the same group, forum, or public page does not constitute a mutual connection.

(i) “PLATFORM” means a provider of an interactive computer service that:

(1) Publishes information provided by an information content provider;

(2) Enables users to view or interact with content; and

(3) Has at least 100,000 monthly active users in this State or generates annual gross revenue in excess of $25,000,000, adjusted annually to reflect adjustments in the Consumer Price Index.

(j) “PRECISE GEOLOCATION INFORMATION” means information derived from technology, including but not limited to global positioning system coordinates, that identifies the real-time or recent location of an individual with sufficient precision to identify a street name and the name of a city or town.

(k) “REASONABLE STEPS” means proactive and reactive measures that can reasonably be expected from a platform considering its size, resources, user base, and the nature of the fraudulent content at issue, including but not limited to:

(1) Implementing and enforcing clear policies against fraudulent content;

(2) Maintaining an accessible and responsive mechanism for users to report suspected fraudulent content;

(3) Reviewing reported content in a timely manner;

(4) Removing or limiting the distribution of content determined to be fraudulent;

(5) Warning potentially affected users when fraudulent content has been identified; and

(6) Implementing reasonable measures to prevent repeated or similar fraudulent content.

(l) “UNCONNECTED USER” means a user who

(1) is not verified by the platform, and

(2) does not share a mutual connection with the user whose content or account is at issue.

SECTION 3. PLATFORM LIABILITY

(a) IN GENERAL— A platform that conducts business in this State may not

(1) Materially contribute to the development of fraudulent content directed at residents of this State.

(2) Target, recommend, promote, curate, or prioritize fraudulent content directed at residents of this State if it knows or has reason to know that the content is fraudulent.

(3) Share consumer data or behavioral profiles with third parties it knows or has reason to know are likely to engage in fraud.

(4) Fail to enforce material published terms, policies, or community standards concerning fraudulent content where

(A) Users reasonably relied on specific representations regarding fraud prevention, detection, or removal, or

(B) Such failure demonstrates systematic disregard for the platform's anti-fraud commitments.

SECTION 4. NOTICE AND TAKEDOWN SAFE HARBOR.

(a) REPORTING REQUIREMENTS — The Attorney General shall promulgate regulations pursuant to the rulemaking authority conferred under Section 7 requiring platforms to provide easily accessible mechanisms for users to report suspected fraudulent content and content creators to submit counter-notices contesting the removal of content.

(1) Reporting mechanisms shall include

(A) A clearly visible mechanism for reporting suspected scams on all content that could contain advertisements, commercial, or promotional material;

(B) Simplified reporting processes that

(i) Do not require users to navigate multiple pages or provide extensive information

(ii) Are accessible to users with disabilities, compatible with common assistive technologies, and available in the user's selected interface language; and

(iii) Allow users to provide

(I) Their identity and contact information;

(II) The specific factual basis for believing that the reported content is fraudulent;

(C) Acknowledgment of receipt of reports within [x] days; and

(D) Provision of notice to the content creator.

(2) Counter-notice mechanisms shall include

(A) The content creator's identity and contact information; and

(B) Specific facts supporting the statement that the reported content is not fraudulent.

(b) SAFE HARBOR.—

(1) A platform shall not be liable under Section 3(a)(4) if it:

(A) Has published and implemented reasonable policies and procedures to address fraudulent content;

(B) Reviews and updates such policies and procedures at least yearly;

(C) Reviews reported fraudulent content within 48 hours of receiving notice; and

(D) Takes reasonable steps to detect and initiate appropriate steps with respect to the fraudulent content within _ days of receiving notice, unless:

(i) The platform reasonably determines after investigation that the content is not fraudulent; or

(ii) Additional time is necessary due to the complexity of the investigation, volume of reports, or other reasonable factors, provided that

(I) The platform provides written notice to the reporting party within the initial 7-day period explaining the need for additional time and the expected timeline for resolution;

(II) Any extension beyond the initial 7-day period shall not exceed 30 days unless the platform obtains written approval from the Attorney General for additional 30-day extensions.

(2) A platform shall not be liable for removing or restricting access to content based on good faith reliance on reports that substantially comply with the requirements in subsection (a), regardless of whether such content is ultimately determined to be fraudulent or lawful, if it:

(A) Promptly provides a copy of the counter-notice to the original reporter;

(B) Restores the removed content not less than [], nor more than [], business days following receipt of the counter-notice, unless the platform, after reviewing the counter-notice and any additional evidence, reasonably determines that the content is fraudulent.

(C) Notifies both parties of the platform's decision and timeline for any restoration.

(3) A platform may refuse to process reports from users who have submitted three or more reports determined to be false or made in bad faith within a 12-month period, provided the platform gives notice and opportunity to contest such determination.

SECTION 5. PLATFORM DESIGN ELEMENTS [Consumer Reports has concerns about the effects of provisions B, C, D, and G in Sec. 5 (a)(1) on consumer use and free expression. However, these provisions are only defaults that can be changed by a user of a service, and we include this language because of the disproportionate damage that functions such as unsolicited Direct Message are leading to for victims.]

(a) FRAUD-RESISTANT DEFAULT PRIVACY SETTINGS.—

(1) A platform subject to this Act shall configure all default privacy settings to reduce fraud vulnerabilities, including through the following settings:

(A) Not displaying a user's phone number, email address, birthdate, or birth year to any Unconnected User unless the user has expressly and unambiguously chosen to make that information visible, either generally or for specific users;

(B) Not displaying media created or posted by a user on a platform to any Unconnected User unless the user has expressly and unambiguously chosen to make their media publicly available, either generally or for specific users;

(C) Not permitting any Unconnected User to like, comment on, or otherwise provide feedback on a user's media unless the user has expressly and unambiguously allowed such feedback from Unconnected Users, either generally or for specific users;

(D) Not permitting direct messaging on a platform between a user and any Unconnected User unless the user has expressly and unambiguously decided to allow direct messaging, either generally or for specific users;

(E) Not displaying a user's Precise Geolocation Information to any Unconnected User, unless the user expressly and unambiguously shares their location, either generally or with specific users;

(F) Not displaying the users connected to a user to any Unconnected User unless the user expressly and unambiguously chooses to share the information, either generally or for specific users; and

(G) Disabling search engine indexing of a user's account profile unless the user expressly and unambiguously opts into indexing.

(2) A platform shall not

(A) Provide a user with a single setting that makes all of the default privacy settings less protective at once;

(B) Request or prompt a user to make their privacy settings less protective unless the change is strictly necessary for a user to access a service or feature they have expressly and unambiguously requested; or

(C) degrade, limit, or reduce the functionality, performance, or quality of services as a result of a user maintaining higher privacy settings unless such limitation is strictly necessary due to the technical requirements of providing the requested service and the platform cannot reasonably provide the service through alternative means that would preserve the user's privacy settings.

(b) EVASION PREVENTION REQUIREMENTS.— Platforms shall implement reasonable measures to detect and prevent the creation of multiple accounts by the same individual for the purpose of evading enforcement actions related to fraudulent content, including:

(1) Accounts created shortly after suspension or removal of an account for fraud violations;

(2) Accounts that exhibit substantially similar patterns of fraudulent behavior to previously suspended accounts;

(3) Coordinated networks of accounts promoting fraudulent schemes.

(c) RULEMAKING.— The Attorney General shall promulgate regulations pursuant to this section under the rulemaking authority conferred under Section 7 that:

(1) Prohibit additional platform design features that facilitate fraud;

(2) Mandate the adoption of reasonable measures to prevent the creation of fraudulent profiles and pages; and

(3) Mandate the implementation of systems to prevent fraudulent reviews, likes, and other engagement metrics.

SECTION 6. ACCESSIBILITY REQUIREMENTS.

(a) DESIGN REQUIREMENTS.—The Attorney General may promulgate regulations pursuant to the rulemaking authority conferred under Section [x] establishing accessibility requirements for advertising content on platforms, including:

(1) Minimum font size requirements for material terms, conditions, and pricing information;

(2) Requirements for clear and conspicuous labeling;

(3) Contrast standards between text and background;

(4) Standards for making advertising content accessible to users with disabilities; and

(5) Additional design requirements to prevent deceptive or misleading advertising practices.

SECTION 7. KNOW YOUR CUSTOMER REQUIREMENTS.

(a) VERIFICATION REQUIREMENTS.—

(1) Platforms subject to this Act shall establish and maintain reasonable procedures to verify the identity of advertisers before allowing them to purchase advertisements.

(2) Such verification procedures shall include, at minimum:

(A) Collection and verification of the advertiser's:

(i) Legal name and physical address;

(ii) Valid contact information, including email address and phone number;

(iii) Tax identification number or other government-issued identifier;

(iv) For businesses, proof of registration or incorporation in the relevant jurisdiction; and

(v) A declaration of industry type and whether the advertiser engages in the sale or promotion of high-risk products, services, or opportunities.

(B) Verification of payment methods and accounts used by the advertiser; and

(C) Reasonable steps to confirm the authenticity of the information provided.

(b) HIGH-RISK ADVERTISERS.—

(1) For advertisers in high-risk categories, or that engage in the sale or promotion of restricted or high-risk products, services, or opportunities, platforms shall implement enhanced verification, including:

(A) Additional documentation to validate the advertiser's identity and business operations;

(B) Review of the advertiser's online presence and business history; and

(C) Periodic re-verification.

(2) High-risk categories include, but are not limited to:

(A) Financial services and investment opportunities;

(B) Health products and services;

(C) Educational credentials and certification;

(D) Housing and real estate;

(E) Gambling;

(F) Unregulated financial products, including cryptocurrency and non-fungible tokens;

(G) Tobacco, alcohol, cannabis, and other consumables restricted by age; and

(H) Categories identified by the Attorney General through rulemaking.

(c) RECORD KEEPING.—

(1) Platforms shall maintain records of advertiser verification information for a period of at least three years from the date of the last advertisement or commercial activity.

(2) Upon receipt of a valid subpoena, court order, or written request from a law enforcement agency in connection with an investigation of fraudulent activity, platforms shall promptly provide relevant advertiser information.

(d) TRANSPARENCY.—

(1) Platforms shall provide users with access to basic information about advertisers whose content they view, including:

(A) The legal name of the advertiser;

(B) The country or jurisdiction where the advertiser is based; and

(C) How long the advertiser has maintained an account on the platform.

(2) This information shall be accessible through direct display on the advertisement, a clearly visible link or button associated with the advertisement or commercial content, or through other reasonably accessible means.

(e) REPEAT OFFENDERS.—

(1) Platforms shall establish and maintain systems to:

(A) Track advertisers found to have violated platform policies related to fraudulent content;

(B) Prevent such advertisers from creating new accounts under different identities; and

(C) Apply enhanced scrutiny to new accounts that share characteristics with previously suspended accounts.

(2) Platforms may not provide advertising services to advertisers who have repeatedly engaged in fraudulent or deceptive practices.

SECTION 8. ENFORCEMENT.

(a) ATTORNEY GENERAL ENFORCEMENT.—

(1) A violation of this Act shall constitute an unfair or deceptive act or practice under [insert state UDAP law] and shall be subject to all the investigatory and enforcement powers conferred upon the Attorney General under that statute.

(2) In pursuing enforcement action, the Attorney General may:

(A) Seek an injunction to restrain any violation of this Act;

(B) Seek civil penalties not to exceed $___ for each violation, or, in the case of a continuing violation, $___ for each day during which the violation continues;

(C) Seek restitution for residents of this State who suffered monetary harm; and

(D) Recover reasonable attorney's fees and costs.

(b) PRIVATE RIGHT OF ACTION.—

(1) Any person who suffers a monetary loss as a result of a platform's violation of this Act may bring a civil action against such platform.

(2) In such action, the plaintiff may:

(A) Recover actual damages;

(B) Obtain injunctive relief to prevent further violations;

(C) In the case of willful violations, recover treble damages; and

(D) Recover reasonable attorney's fees and costs if the plaintiff is the prevailing party.

(c) NOTICE REQUIREMENT FOR PRIVATE ACTION.—

(1) At least 30 days before filing an action under subsection (b), a person must provide written notice to the platform and to the Attorney General identifying:

(A) The specific alleged violation;

(B) The fraudulent content at issue; and

(C) The monetary loss suffered.

(2) If, within 30 days of receiving notice, the platform removes the fraudulent content and offers the person reasonable compensation for the monetary loss, no action may be brought by that person against the platform under subsection (b) for that specific violation.

(d) COORDINATION WITH EXISTING STATE LAWS.—

(1) Nothing in this Act shall be construed to limit or preclude enforcement of the State's consumer protection, unfair trade practices, or false advertising laws.

(2) A violation of this Act shall also constitute an unfair or deceptive practice under [State's Unfair and Deceptive Acts and Practices statute].

(e) RULEMAKING AUTHORITY

(1) The Attorney General shall have the authority to promulgate rules and regulations to implement, interpret, and enforce all provisions of this Act, including but not limited to the specific rulemaking authorities referenced elsewhere in this Act.

SECTION 9. LIMITATIONS AND EXCEPTIONS.

(a) NO REQUIREMENT TO MONITOR.—Nothing in this Act shall be construed to require a platform to monitor all content before it is posted or transmitted.

(b) EXCEPTIONS.—This Act does not apply to:

(1) Interactive computer services that function primarily as internet service providers, email providers, or data storage providers;

(2) Content that is transmitted through a platform but not stored on the platform;

(3) Actions taken by a platform in good faith to restrict access to or availability of material that it considers to be fraudulent, deceptive, or otherwise objectionable.

SECTION 10. SEVERABILITY

The provisions of this Act are severable. If any provision of this Act or any part of any section is declared invalid or unconstitutional, the declaration shall not affect the validity or constitutionality of the remaining portions.