Trusted Data and Technology Institute
TECHNOLOGY WORTHY OF TRUST

RESEARCH PAPER 2026-09-01 · TRUSTED DATA & CONSUMER PROTECTION

Complaints About The Machine

What consumers told the CFPB about financial firms' chatbots, phone systems and automated decisions, 2019–2026: a validated analysis of 3.49 million published complaint narratives

Author
Joseph Atalla, PhD

Published
September 2026

ABSTRACT

Banks and card issuers increasingly route customers through chatbots, automated phone lines and automated account decisions. We screened all 3,485,222 complaint narratives the CFPB published for complaints received from January 2019 to August 13, 2026 with a pre-registered lexicon, and coded a stratified sample of 528 against a written codebook (κ = 0.92 on 176 double-coded). In checking, savings and card complaints, about 1.0% of narratives describe a firm's automated system as part of the problem on the screen alone (95% CI 0.9–1.1%), and an estimated 4.9% (1.8–8.9%) including cases the screen misses. Phone systems that block a person and automated decisions each account for over 40% of detected cases, chatbots about 10%, and the rate was flat from 2019 to 2025. Half of coded cases describe being unable to reach a human; a quarter describe losing access to funds. Matched on product, issue, year and firm, these complaints obtained relief at the same rate as others. No structured field can stand in for the narrative, and narrative publication fell from 45% of complaints in 2019 to under 3% in 2026 before the CFPB ended it on August 14, 2026. This study cannot be repeated with the public data that remain.

1% –5%	52%	1.5%	45% →0
of bank and card complaint narratives describe a firm's chatbot, phone system or automated decision (detected floor, and estimate including undetected cases)	of coded cases say the consumer could not reach a human; 82% of those about phone systems	precision of the best structured-field proxy for these complaints at 50% recall; the narrative is the only signal	share of complaints published with a narrative, January 2019 to August 2026; 1.5 million complaints since August 1 have none

About TDTI. The Trusted Data and Technology Institute is a California nonprofit public benefit corporation that advances data and technology quality, integrity, security and responsible use through research, education and practical collaboration, with a focus on trusted data, responsible AI evaluation, cybersecurity assurance and public-sector technology. Contact: info@trusteddti.com · trusteddti.com · California

Atalla, J. (2026). *Complaints About the Machine: What consumers told the CFPB about financial firms' chatbots, phone systems and automated decisions, 2019–2026.* TDTI Research Paper 2026-09-01. Trusted Data and Technology Institute.

Executive summary

Financial firms have moved much of their customer contact to automated channels: chatbots, phone menus, and systems that freeze accounts, decline transactions or deny disputes without a person in the loop. Consumers who are harmed by those systems have one public place to say so, the CFPB's complaint database, and until August 2026 their own words were published there. This paper measures how often consumers described a firm's automation as part of their problem, what went wrong, and whether those complaints were resolved differently.

KEY FINDINGS

1 A small but steady share. In checking, savings and credit card complaints from 2019 to 2025, the screen and coding confirm about 100 complaints per 10,000 narratives (1.0%) that describe a firm's automated system as part of the problem. A random check of complaints the screen missed suggests the true share is several times higher, about 4.9% (95% CI 1.8–8.9%), because most consumers describe automation in their own words.

2 Phone systems and automated decisions, more than chatbots. Of detected cases, 46% concern phone systems or no path to a human, 44% automated decisions such as fraud freezes, closures and dispute denials, and 10% chatbots. The detected rate did not rise between 2019 and 2025 (100 and 101 per 10,000).

3 The harm is access. In coded cases, 52% of consumers say they could not reach a human, and the most common concrete harm is losing access to their own funds (25%). Nearly all (95%) describe the problem as unresolved when they complained.

4 No outcome penalty, and no outcome bonus. Within the same product, issue, year and firm, these complaints were closed with relief at the same rate as others (bank products: odds ratio 0.99, 95% CI 0.94–1.05).

5 Firms differ about fourfold. Among the ten firms with the most bank and card complaints, screened candidates range from 61 to 254 per 10,000 narratives.

6 The evidence base is gone. No structured field identifies these complaints (best proxy: 1.5% precision). The share of complaints published with a narrative fell from 45% (2019) to under 3% (2026), then to zero when the CFPB stopped publication on August 14, 2026.

What should change. The CFPB should add structured complaint fields for automated channels and decisions (R1) and publish counts from them (R2); restore access to narratives for research, at least in redacted bulk form (R3); and firms should report and fix "no path to a human" failures (R4–R5). Congress and state regulators should treat a working route to a person as a baseline consumer protection (R6).

1 Why complaints about automation matter

The CFPB's 2023 issue spotlight on chatbots warned that automated customer service can leave consumers stuck in loops, give wrong answers, and fail to recognize when a consumer is disputing a charge or asserting a legal right.[1] It drew on complaints, but reported no counts, trends or firm comparisons. Survey evidence points the same way: in Consumer Reports surveys of about 2,000 adults each in February and October 2024, chat users who believed they were talking to a bot were roughly three times as likely to say they did not get the help they needed (58% vs 21%).[2] The regulatory backdrop has since shifted. In May 2025 the CFPB withdrew 67 guidance documents, including its circulars on adverse action notices for algorithmic credit decisions, though not the chatbot spotlight.[3] On August 14, 2026 it stopped publishing complaint narratives, moving those already published to its FOIA reading room.[4,5]

Research questions

RQ1. How often do consumers describe a firm's chatbot, phone system or automated decision as part of their complaint, and has that changed since 2019? **RQ2.** What harms do they describe, and do these complaints end differently? **RQ3.** Can the data the CFPB still publishes reveal these complaints without narratives?

Prior work and contribution

Machine learning studies of CFPB narratives predict outcomes or classify issues, but do not isolate complaints about firms' automation.[6,7] Shin, Shin and Kim study the opposite direction, complaints written *with* ChatGPT, and find they are more likely to obtain relief.[8] This paper is, to our knowledge, the first validated count of complaints *about* firms' automated systems across the full public narrative record, and the first to measure what the end of narrative publication removes.

2 Data and methods

Source	Coverage	Records	Use
CFPB Consumer Complaint Database, full CSV[9]	Received Dec 1, 2011 – Oct 14, 2026; 15 structured fields, no narratives	18,399,780	Firm volumes; post-August 2026 volumes
CFPB FOIA reading room, narratives archive, exports 2–21[5]	Received May 2018 – Aug 2026; all fields including narrative	16,502,055	Screening, outcomes, narrative availability
Analysis window (narratives received Jan 1, 2019 – Aug 13, 2026)	All products; bank subset = checking/savings and credit/prepaid card	3,485,222	All estimates (bank subset 387,270)
Validation sample (stratified, seeded)	328 screened hits, 200 non-hits; 176 double-coded	528	Precision, missed cases, harms

Screening. A lexicon fixed in a dated pre-analysis plan flags narratives in three primary categories: (A) chatbots and virtual assistants; (B) automated decisions attributed to a system, algorithm or "AI"; and (C) automated phone lines and reports of no path to a human. Two further categories are recorded but excluded: outbound robocalls (D) and consumers who say they used AI to write the complaint (X). One correction was made after screening and is reported as a deviation: matches on the name of Wells Fargo's assistant, "Fargo," were dropped because the pattern failed to exclude lower-case "wells Fargo" (2,267 false matches). Bank of America's "Erica" produced no matches, probably because the CFPB redacts personal names (an inference).

Validation. We drew 80 screened hits per category in bank products and 30 elsewhere, plus 100 random non-hits in bank products and 100 across all products. Each was coded blind to stratum and matched terms against a written codebook: the consumer must attribute the problem to an automated system operated by or for the firm complained about. Coding was done by AI coding agents (Claude) instructed with the codebook, not by people; a second, independent agent coded one third of the sample. Agreement on relevance was 96.6% (Cohen's $\kappa = 0.92$) and on category 95.5% ($\kappa = 0.91$). Agreement between AI coders measures consistency, not correctness; the full coding sheet is released for human verification. Counts are estimated by applying each stratum's coded category mix to all hits in that stratum (a stratified ratio estimate, with bootstrap 95% intervals). The *detected floor* counts screened hits only; the *full estimate* adds missed cases from the random non-hit strata.

Outcomes. Relief means closure with monetary or non-monetary relief, among closed complaints. We compare flagged and unflagged narratives raw, within exact strata of product × issue × year × firm, and by logistic regression with product, issue, year and firm fixed effects.

3 Findings

3.1 How often, and how the screen performed

The screen flagged 29,099 of 3,485,222 narratives (0.83%). Hand coding confirmed 55% of sampled hits as complaints about a firm's automated system (181 of 328), with the rest mostly incidental mentions, form letters described as "computer generated," or automation run by a third party such as a merchant. Precision was higher in bank products than elsewhere for automated decisions (65% vs 30% of sampled hits), largely because credit reporting complaints use "automated" for template dispute letters. The screen misses much more than it finds: 4 of 100 random bank-product non-hits (95% CI 1.6–

9.8%) also described a firm's automation, in phrases the lexicon did not anticipate, such as "automatic service" for a phone line.

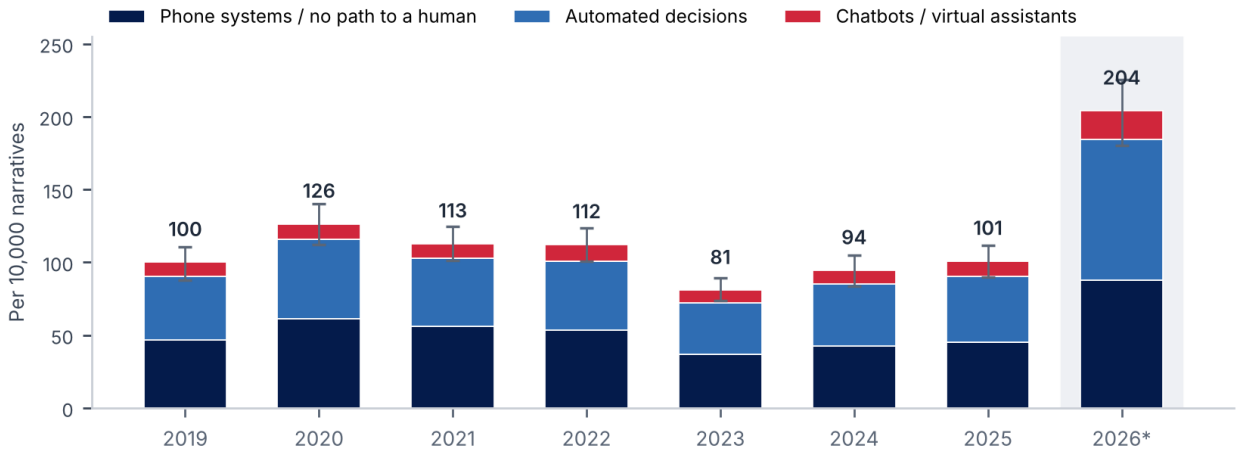
Bank products, 2019–2025 (344,038 narratives)	Per 10,000	95% CI	Share of detected
Phone systems / no path to a human (C)	46	37–57	46%
Automated decisions (B)	44	34–54	44%
Chatbots / virtual assistants (A)	10	6–14	10%
Any, detected floor	100	90–111	3,444 complaints
Any, including estimated missed cases	494	179–885	~17,000 complaints
All products, detected floor (3,363,828 narratives)	35	28–42	11,815 complaints

Source: CFPB narratives archive; TDTI screening and coded sample. Category rows are detected-floor estimates and sum to the "any" row. The wide interval on the full estimate reflects a 100-complaint random check; a larger check would narrow it.

3.2 No rise through 2025

Despite rapid adoption of chatbots and generative AI by large banks in this period, the detected rate in bank products was 100 per 10,000 in 2019 and 101 in 2025, with a peak of 126 in 2020 and a low of 81 in 2023 (Figure 1). Chatbot complaints held near 10 per 10,000 every year. The 2026 rate (204) is not comparable: only 2.9% of 2026 complaints were published with a narrative, and the median 2026 narrative is about half again as long as in 2025 (1,014 vs 668 characters), so the published 2026 subset differs in kind. The screen also cannot tell whether consumers increasingly use AI to write complaints; only 24 narratives say so explicitly.

Figure 1. Complaints describing a firm's automated system, bank products, per 10,000 narratives



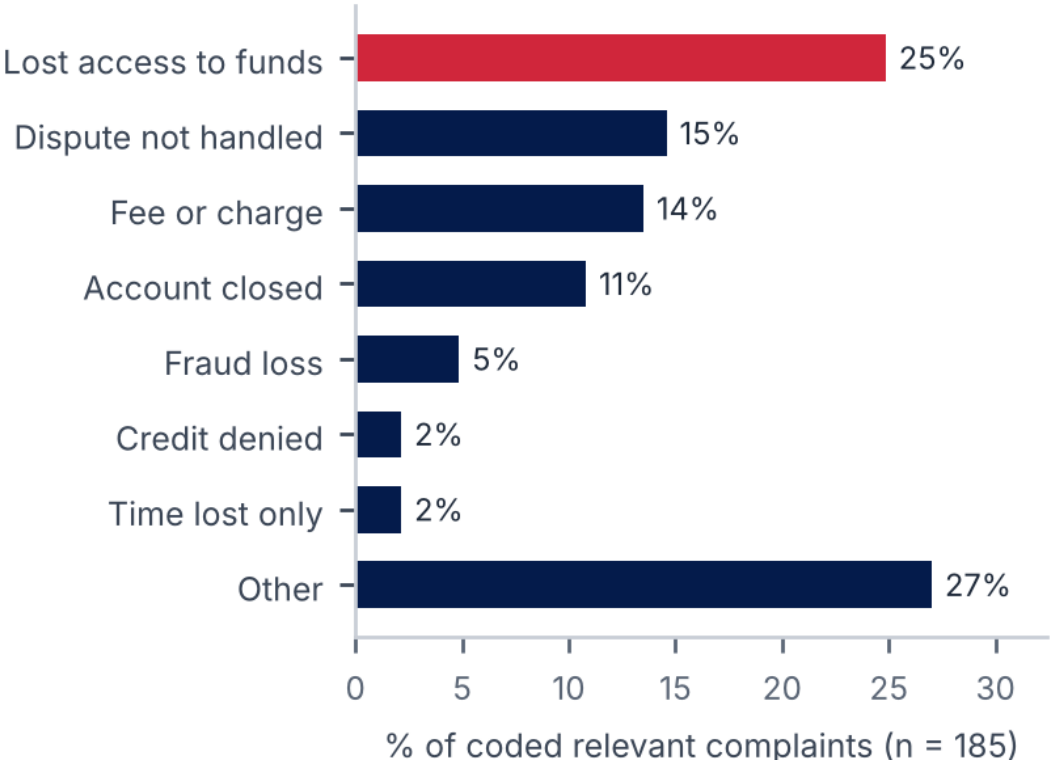
Source: CFPB narratives archive; TDTI screening and coded sample. Detected floor: screened hits weighted by coded precision; error bars are bootstrap 95% intervals for the total. Bank products: checking/savings and credit/prepaid card. *2026 covers complaints received January 1 to August 13, of which only 2.9% were published with a narrative; it is not comparable with earlier years.

3.3 What goes wrong: access, and no way to reach a person

Of the 185 sampled complaints coded as relevant, 52% (95% CI 45–60%) say the consumer could not reach a human or was kept from one. That rises to 82% for phone-system complaints and 50% for chatbot complaints, but only 7% for automated decisions, where the consumer usually did reach a person who then deferred to the system ("the automated system said").

The most common concrete harm is losing access to one's own money: frozen or restricted accounts, held deposits and blocked cards (25%, Figure 2). Disputes closed without review (15%), fees or charges (14%) and account closures (11%) follow. In 176 of 185 cases (95%) the consumer describes the problem as unresolved when they turned to the CFPB, which is expected of a complaint, but confirms that automated channels were not resolving these problems first.

Figure 2. Harm described in coded complaints



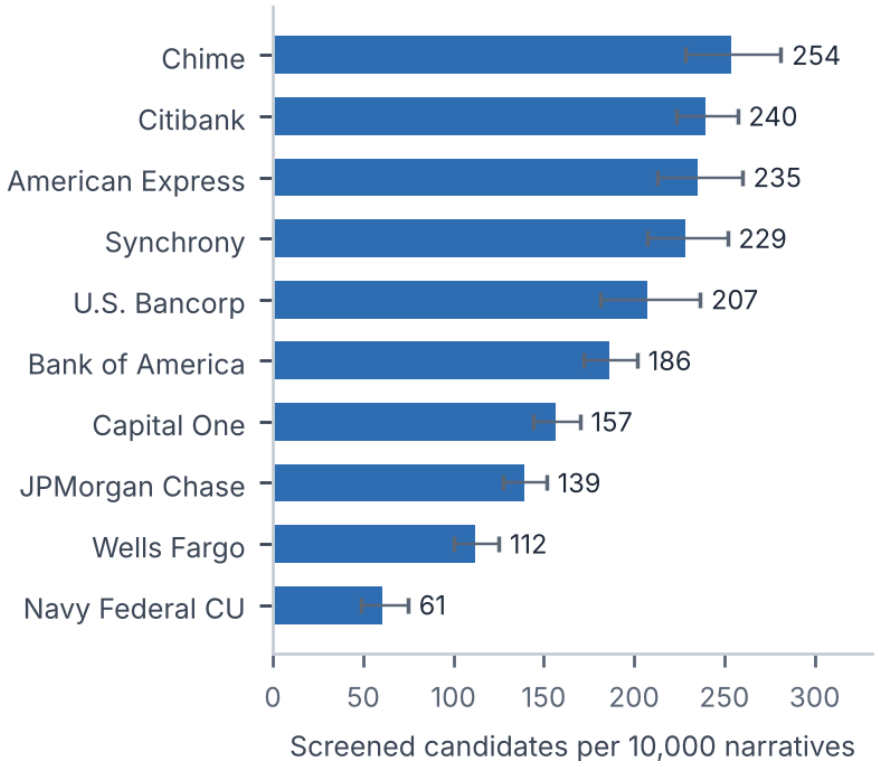
Source: TDTI coded sample, complaints coded relevant (n = 185), all strata. One main harm per complaint. Shares describe the sample, not the population.

3.4 Do these complaints end differently?

Raw comparisons mislead. Across all products, flagged complaints were closed with relief less often (27.6% vs 34.9%), but only because credit reporting complaints, which rarely mention automation, are relieved more often. Within the same product, issue, year and firm, the difference is zero (+0.02 percentage points across 6,797 matched strata). In bank products, relief rates were 25.2% for flagged and 25.9% for other complaints; the matched difference is +0.5 points and the fixed-effects odds ratio 0.99 (95% CI 0.94–1.05; n = 386,920). The all-product model on a 1-million-complaint random sample gives 1.12 (1.06–1.18). Timely response was 98.0% vs 99.0% raw and indistinguishable after matching (−0.1 points). Flagged chatbot complaints had the lowest raw relief rate in bank products (24.2%). Because flags are about 55% precise, these comparisons are diluted toward zero and describe flagged complaints rather than confirmed ones.

3.5 Firms: a uniform rule

Figure 3. Screened candidates per 10,000 narratives, ten firms with the most bank and card complaints



Source: CFPB data; TDTI screening (lexicon v1.1). Firms ranked by total checking/savings and card complaints, Jan 2019–Jul 2026. Bank products only; Wilson 95% intervals. Screened candidates, not confirmed cases.

To limit discretion, and because of the author's employment, firms are named only if they rank in the top ten by total checking, savings and card complaint volume, a rule fixed before analysis. Among those ten, the rate of screened candidates ranges about fourfold, from 61 per 10,000 bank and card narratives at Navy Federal Credit Union to 254 at Chime (Figure 3). Chime has the highest phone-system rate (104 per 10,000) and Citibank, American Express and Synchrony the highest automated-decision rates (152–154).

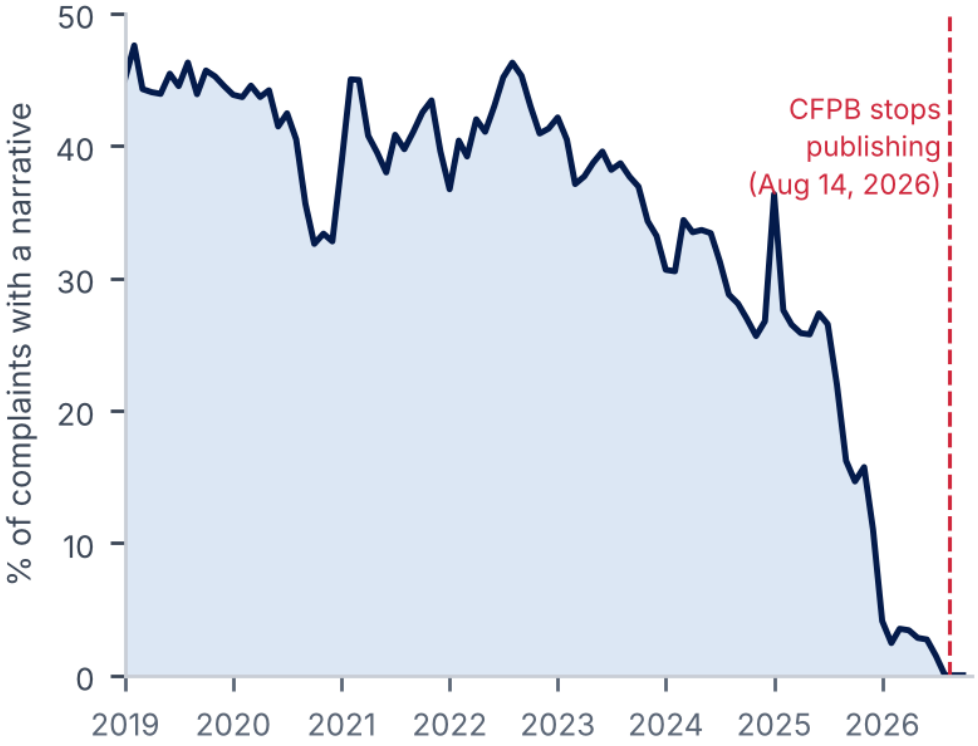
These are screened rates. The validation sample is too small to estimate precision firm by firm, so differences partly reflect how each firm's customers write, and how much each firm's complaints are about card disputes, where automated decisions are common. They indicate where a regulator or the firm itself should look first, not a ranking of conduct.

3.6 The data problem: only the narrative carries the signal, and it is gone

The CFPB's structured fields have no category for automated channels or decisions. Using issue and sub-issue as a proxy fails: ranking the 215 issue and sub-issue cells with at least 100 narratives by their share of flagged complaints, the cells that capture half of flagged complaints have a precision of 1.5%, barely above the 0.83% base rate. Without narratives, these complaints are invisible.

Narratives were also disappearing before the formal stop (Figure 4). In 2019, 45% of complaints were published with a narrative. The share drifted to about 27% in mid-2025, fell to 11% in December 2025, and was 1.5–4.1% in each month of 2026. The CFPB announced on August 14, 2026 that it would cease publication.[4] The structured file now holds 1,534,993 complaints received since August 1, 2026, none with a public narrative. The archive keeps past narratives available for now, as monthly bulk files rather than a searchable database.[5]

Figure 4. Share of complaints published with a narrative, by month received



Source: CFPB full CSV (denominator) and narratives archive (numerator). Narratives appear only with consumer consent and after company response, so recent months are always lower; the 2026 drop is far larger than that lag.

4 Discussion

The problem consumers report is access, not intelligence. Concern about AI in banking often centers on biased models. The complaints point to something more ordinary: phone systems and decision engines that stand between a consumer and their money, with no person available to override them. That is a design and staffing choice, and it is fixable.

The share is small, but the volume is not. Even the detected floor implies over 11,800 such complaints across products since 2019, and the full estimate for bank products alone is around 17,000. Complaints are themselves a small fraction of consumers who experience a problem.

Flat is not reassuring. A stable rate through a period of rapid chatbot adoption could mean automation is working, that consumers blame the firm rather than the machine, or that consumers do not know whether they are dealing with a bot, which is the gap Consumer Reports found.[2] Complaint data cannot separate these explanations, and after August 2026 the public cannot test them at all.

5 Recommendations

R1 Add structured fields for automation. The CFPB complaint form should ask whether the problem involved a chatbot, automated phone line or automated decision, and whether the consumer could reach a person. A two-question addition would make these complaints countable without narratives.

R2 Publish counts. The CFPB should report those fields in its public data and annual Consumer Response report, by product and firm, as it does for issues today.

R3 Keep narratives available for research. If public narrative search does not return, the CFPB should keep the bulk archive online, continue it for new complaints with consent, and offer a researcher access program with redaction, so independent analysis like this one remains possible.

R4 Guarantee a path to a human. Firms should ensure every chatbot and phone menu offers a working route to a person, at minimum for disputes, fraud and frozen funds, and should test that route regularly.

R5 Monitor their own complaints for automation. Firms should tag complaints that involve automated channels or decisions, track them as an operational risk measure, and give model risk and customer experience teams the results.

R6 Treat access as a baseline protection. Congress and state regulators should consider requiring disclosure when a consumer is dealing with an automated agent, and a right to escalate to a person for account access and dispute decisions.

6 Limitations and reproducibility

Unverified, self-selected narratives. The CFPB does not verify narratives, and they are published only with consent, so results describe published complaints, not consumers generally. **Consumer attribution.** We code what consumers say; a complaint that does not mention automation is not counted even if automation was involved, and a consumer may misattribute a human decision to a system. **AI coding.** The sample was coded by AI agents following a written codebook, with high agreement between independent agents; this is not a substitute for human coding, and the coding sheet is released for verification. The plan called for human coders; this is a deviation. **Sampling error.** The missed-case estimate rests on 100 random non-hits, and precision is assumed constant across years and firms within each stratum. **Deviations from the pre-analysis plan** are listed in the replication package: the Fargo correction, AI rather than human coding, 2026 treated as non-comparable, and an all-product regression fitted on a 1-million-complaint random sample for memory.

Replication package. Scripts (screening, sampling, analysis, validation, figures), the dated pre-analysis plan and codebook, the lexicon, the coding sheet with codes, complaint-level flags without narrative text, and aggregate tables accompany this paper. The CFPB source files are public and are not redistributed; their file names, sizes and download dates are listed.

Conflict of interest. The author is employed full time by a U.S. bank in an AI and data analytics role. The author's employer appears in the CFPB data analyzed but is not among the ten firms named in this paper, and the naming rule in Section 3.5 was fixed before analysis so that every firm is treated the same way. This research was conducted independently of the author's employer, using only public CFPB data; the employer did not fund, review or approve it, and the views expressed are the author's own.

References

1. Consumer Financial Protection Bureau. *Chatbots in consumer finance* (issue spotlight), June 2023. https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/
2. Chien, J. Memo to the CFPB on chatbots in consumer finance, with results of Consumer Reports surveys (Feb 2024, n = 2,035; Oct 2024, n = 2,019). Consumer Reports, 6 Dec 2024.
3. CFPB. *Interpretive Rules, Policy Statements, and Advisory Opinions; Withdrawal*. Federal Register document 2025-08286, 12 May 2025. https://www.federalregister.gov/documents/2025/05/12/2025-08286/interpretive-rules-policy-statements-and-advisory-opinions-withdrawal
4. CFPB. *The CFPB to Cease Discretionary Publication of Complaint Narratives and Visualizations*, 14 Aug 2026. https://www.consumerfinance.gov/about-us/newsroom/the-cfpb-to-cease-discretionary-publication-of-complaint-narratives-and-visualizations/
5. CFPB. *CFPB Consumer Complaint Database Narratives Archive*, FOIA electronic reading room (page created 14 Aug 2026; last modified 14 Sep 2026). https://www.consumerfinance.gov/foia-requests/foia-electronic-reading-room/cfpb-consumer-complaint-database-narratives-archive/
6. Wang, Zhu and Chen (2026). arXiv:2606.22664 (machine learning prediction of monetary relief from CFPB complaint narratives).
7. Vaishnav et al. (2024). arXiv:2407.06399 (machine learning applied to the CFPB complaint database).
8. Shin, Shin and Kim (2024). *The Adoption and Efficacy of Large Language Models: Evidence From Consumer Complaints in the Financial Industry.* SSRN 5004194.
9. CFPB. *Consumer Complaint Database* and field reference. https://www.consumerfinance.gov/data-research/consumer-complaints/; https://cfpb.github.io/api/ccdb/fields.html