



Article

From Delivery Delays to AI-Mediated Escalation Failures: A BERTopic Analysis of Complaints About Risk and Trust in E-Commerce Marketplaces (2019–2025)

Munise Hayrun Sağlam 

Department of Business Administration, Faculty of Economics and Administrative Sciences, Yıldız Technical University, Davutpasa Campus, 34220 Istanbul, Türkiye; mhsaglam@yildiz.edu.tr

Abstract

Automated customer service and algorithmic governance are common in digital marketplaces, yet trust can erode when logistics, refunds, and escalation fail. Complaint-based risk and trust narratives in Turkey's e-commerce marketplaces are analyzed for January 2019–December 2025 using 118,173 de-identified Turkish and English texts from Şikayetvar, a leading Turkish online consumer-complaint portal, and reviews of official marketplace apps on Google Play and the Apple App Store. BERTopic is implemented in Python with multilingual transformer embeddings, UMAP, HDBSCAN, and c-TF-IDF representations. The selected model identifies 35 micro-topics grouped into five macro-themes: fulfillment disruptions, remediation frictions, product-integrity risks, escalation failures, and governance threats. Monthly probability-weighted prevalence is estimated, and marketplace differences are evaluated with divergence measures, permutation tests, and multinomial regression controlling for time and language. Change-point tests indicate a shift toward fulfillment grievances in April 2020, rising governance threats from June 2022, and increasing escalation failures linked to automated support from February 2023. These patterns suggest that barriers to human escalation convert operational incidents into platform-level trust judgments, offering monitoring signals for service recovery, marketplace governance, and AI oversight. By isolating escalation failures as a distinct complaint domain, the study links service automation to procedural justice mechanisms that translate operational breakdowns into platform-level trust and risk judgments.

Keywords: e-commerce; consumer complaints; trust; perceived risk; BERTopic; topic modeling; AI customer service; procedural justice; platform governance; multilingual text mining



Academic Editor: Hua Pang

Received: 14 February 2026

Revised: 27 March 2026

Accepted: 3 April 2026

Published: 9 April 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Digital marketplaces promise convenience, selection, and price transparency. At the same time, they shift key elements of exchange away from direct interaction between buyers and sellers and into platform-managed processes. Trust becomes a prerequisite because consumers must rely on distant sellers, complex logistics networks, and platform dispute systems. Research on online transactions shows that trust reduces uncertainty and supports purchase intentions, whereas perceived risk discourages participation [1–3]. These mechanisms remain central in marketplace settings, where consumers evaluate not only sellers but also the platform that defines the rules and enforces them.

In 2024, Türkiye's e-commerce volume reached TRY 3.162 trillion (≈USD 90 billion), a 61.7% year-on-year increase, with 5.91 billion transactions, and e-commerce accounted

for 6.5% of GDP [4]. Consumer adoption is substantial but still below mature European markets: 51.7% of internet users in Türkiye reported purchasing goods or services online in the previous 12 months in 2024 [5], compared with 77% in the EU [6]. Against this backdrop, this study uses longitudinal complaint texts to map how consumer risk and trust grievances are framed, how their salience shifts from 2019 to 2025, and how these patterns differ across major marketplaces.

Consumer complaints provide a direct record of where expectations collapse. A complaint is rarely only a negative sentiment [7]. It is a narrative that assigns responsibility, describes harm, and demands a remedy [8]. Service recovery research shows that failures and recoveries shape loyalty, word of mouth, and relationship quality [9,10]. Complaint texts often contain richer detail than closed-form survey items because consumers describe specific breakdowns such as delayed delivery, missing parcels, refund delays, counterfeit items, and account compromise. Complaint narratives also reveal perceived justice [11]. When consumers view procedures as inconsistent or unresponsive, they infer unfairness and reduce trust even if outcomes are eventually corrected [12].

The platform economy adds another layer. Marketplaces govern access, visibility, and remediation through policies and technical systems. Platform research argues that governance shapes incentives and behavior across buyers, sellers, and intermediaries [13,14]. As marketplaces scale, governance becomes more standardized and data-driven. Rule enforcement, seller vetting, and dispute routing may rely on algorithmic classification and queue management. This creates a tension: scale encourages automation, while legitimacy depends on perceived fairness, transparency, and accountability [15,16]. Complaints, therefore, capture more than operational issues; they also document governance frictions.

Customer service automation intensifies this tension. Firms deploy chatbots, scripted flows, and self-service portals to reduce response time and cost. Service scholarship treats AI as a structural shift in service design because machines can perform routine tasks and mediate interaction quality [17–19]. Automation can improve speed, but it can also introduce new failure modes. Consumers may get trapped in repetitive scripts, fail to reach a human agent, or lose confidence that their case is being evaluated. These experiences align with procedural justice principles, where voice, consistency, and correctability shape acceptance of decisions [20]. One resulting complaint category is escalation failure. It reflects a perceived barrier to remedy rather than a slow outcome.

Despite extensive work on trust, risk, and service recovery [21–23], evidence remains limited on how consumer risk narratives change over time in marketplace environments. Many studies rely on surveys or experiments that hold context constant and measure perceptions at a single point in time [24–28]. These designs clarify mechanisms, yet they are less suited to capturing shifts driven by logistics shocks, policy changes, or service automation rollouts. Qualitative studies provide depth but often cover narrow time windows or small samples [29]. Longitudinal, high-granularity measurement is needed to track changes in complaint composition, not only changes in complaint volume [30].

Text-as-data methods enable the study of these dynamics at scale. Large review and complaint corpora have been used to identify product attributes, diagnose service pain points, and infer consumer priorities [31,32]. Topic modeling is particularly useful because it extracts recurring themes from unstructured text without requiring an exhaustive hand-coded scheme. Classical probabilistic models such as latent Dirichlet allocation are widely used for discovering themes in large corpora [33]. However, complaint texts are often short, noisy, and multilingual, which can reduce coherence and interpretability in bag-of-words approaches. Embedding-based methods address these limitations by clustering dense semantic representations derived from transformer models [34,35]. BERTopic follows this

approach by combining transformer embeddings, density-based clustering, and class-based TF-IDF to improve interpretability and stability in applied text corpora [36].

Building on this line of work, the present study develops and applies a BERTopic pipeline to map consumer risk and trust complaints in Turkish e-commerce between 2019 and 2025. The corpus comprises de-identified complaint texts drawn from two public channels: posts published on Şikayetvar [37] and user reviews of official marketplace mobile applications on Google Play and the Apple App Store. Using this corpus, the analysis (i) derives an interpretable topic map of complaint content, (ii) estimates topic prevalence over time and detects turning points, and (iii) tests whether topic-prevalence profiles differ across marketplaces after accounting for corpus composition.

This study makes three contributions. First, it offers a longitudinal, marketplace-level measurement of the consumer “complaint agenda” around risk and trust, showing how the relative prominence of grievance categories changes across time. Second, it extends platform trust and perceived risk research by identifying escalation failure as a distinct complaint category that speaks to perceived fairness and trust in marketplace governance, particularly in service and support pathways. Third, it contributes methodologically by demonstrating a multilingual, stability-oriented BERTopic workflow for short and noisy complaint texts, with robustness checks designed for applied settings. The findings also translate into actionable governance implications for refund service levels, authenticity controls, and the design of AI-mediated escalation paths. The main objectives of the study, along with their corresponding research questions, are outlined below.

RQ1: What are the dominant consumer risk and trust complaint topics in e-commerce, and how robust are these topics across modeling choices and languages (Turkish vs. English)?

RQ2: How do complaint topics evolve, and are there identifiable structural shifts or turning points in topic prevalence?

RQ3: Do topic-prevalence profiles differ systematically across marketplaces after accounting for corpus composition (source mix and language share)?

2. Theoretical Background

2.1. Trust, Perceived Risk, and Platform Assurances in E-Commerce Marketplaces

Trust is a prerequisite for exchange in e-commerce marketplaces because buyers pay and share personal data before product quality, delivery performance, or dispute outcomes are observable. This exposure is amplified by platform scale, where small frictions can affect large numbers of transactions [38]. In digital platform research, structural assurance and provider image or reputation show strong effects on trusting intention, underscoring the institutional character of trust in platform settings [39]. A common definition frames trust as a willingness to accept vulnerability based on positive expectations about another party’s intentions or behavior [40]. In marketplace commerce, the trust target is plural. Consumers evaluate at least two trustees: the seller who fulfills the order and the platform that designs and enforces the transaction environment [41–43].

This multi-target structure matters because platforms do more than host listings. They specify refund and return rules, set evidence requirements, route complaints, and monitor seller compliance. These institutional features can substitute for direct interpersonal trust when transactions are uncertain [44,45]. The credibility of platform assurances depends on whether remedies are delivered predictably and contestably. Policy work on online dispute resolution emphasizes fairness, transparency, and secure information transfer as design requirements for digital procedures [46]. As support and enforcement become more automated, governance quality also depends on whether decision processes remain explainable and appealable, since opacity can weaken perceived fairness and legitimacy [47].

Complaint narratives can reveal which trust target is salient because texts often assign responsibility and describe what remedy pathways were available.

Perceived risk is the conceptual counterpart to trust and refers to subjective expectations of loss and uncertainty around outcomes [48]. Recent reviews continue to treat risk as multidimensional, covering financial, performance, time, and privacy risks, and they show that these risks remain central drivers of online purchase decisions [49]. In marketplace settings, these facets map naturally onto complaint domains. Delivery delays align with time and performance risk. Refund delay aligns with financial risk. Counterfeit goods align with performance and authenticity risk, which is closely linked to information asymmetry [50]. Unauthorized transactions align with financial and privacy risk. Cross-border contexts can intensify these exposures because fulfillment and remedy processes are harder to verify [2]. This mapping supports a theory-guided interpretation of topics: topic prevalence can be read as a changing composition of risk facets and assurance failures expressed by consumers.

2.2. Service Failure, Recovery, and Procedural Justice in Human and AI-Mediated Support

Complaints rarely describe a single breakdown. They document a sequence: an initial failure, the platform's response, and the consumer's attempt to obtain a remedy. Service recovery research shows that response quality shapes satisfaction, loyalty, and negative word of mouth [9,10]. In marketplace settings, recovery also functions as a credibility test of the platform's promises, because the platform designs procedures that determine whether the consumer can reach a fair outcome.

Justice theory explains why similar outcomes can be evaluated very differently. Organizational justice is often decomposed into distributive, procedural, and interactional components [20]. In e-commerce, distributive justice is expressed through refunds, replacements, and compensation. Procedural justice is experienced through workflow design: evidence standards, time limits, return steps, appeal options, and the ability to reopen or correct a decision. Interactional justice concerns respectful treatment and adequate explanations, which can be delivered by a human agent or embedded in interfaces and automated messages. Procedural justice matters strongly for legitimacy judgments, even when outcomes are imperfect, because it captures voice, neutrality, and correctability [51].

Platform governance intensifies these dynamics. Marketplaces regulate seller behavior and dispute handling through technical systems that route tickets, rank cases, and enforce policies at scale [52,53]. Algorithmic governance can create accountability gaps when decisions are opaque or difficult to contest, which shifts attention from "what happened" to "how the platform decides" [15,16,54]. In complaint texts, this appears as perceived arbitrariness, blocked appeals, and "black box" outcomes.

Automation makes procedural fairness more fragile. Chatbots and scripted self-service can reduce waiting times for routine requests, yet they can also introduce distinct failure modes: misclassification, loss of context, looping replies, ticket auto-closure, and barriers to human handoff [29,55]. Recent service research shows that AI-mediated recovery is evaluated through fairness cues and communication quality, not just speed, and that context and failure severity condition whether AI responses are accepted [21,24,25,27]. These mechanisms motivate "AI escalation failure" as a theoretically coherent topic category rather than a narrow operational issue.

Complaining is also a form of voice and control restoration [56]. The complaint corpus therefore reflects voiced grievances where a remedy is perceived as possible. Topic prevalence should be interpreted as the changing composition of expressed justice and governance frictions, not as a census of all negative experiences.

3. Methodology

The study analyzes a longitudinal, publicly available corpus of consumer complaint texts about three major e-commerce marketplaces operating in Turkey. The corpus combines two public channels, posts published on Şikayetvar [37] and user reviews of the official marketplace mobile applications on Google Play and the Apple App Store, covering January 2019 to December 2025. Complaints were de-identified prior to analysis, and results are reported only in aggregate form. For brevity, the three marketplaces are denoted as Marketplace A, B, and C. All computational analyses were conducted in Python (version 3.11).

Complaint texts capture how consumers frame failures, assign blame, and define what counts as unacceptable, offering a direct window into trust, risk, and justice mechanisms. They also serve as legitimacy tests of platform governance and dispute resolution [51]. Accordingly, recurring complaint narratives can be treated as latent grievance categories whose prevalence tracks salience and friction points at scale, even if it does not equal incident rates.

To identify these grievance categories and trace their salience over time and across marketplaces, BERTopic is used. Complaint corpora are often short and noisy, which limits the applicability of classical topic models [33]. BERTopic leverages transformer embeddings to cluster semantically similar complaints despite lexical variation [34,35] and produces interpretable topics via class-based TF-IDF [36]. Topic quality and interpretability were evaluated using coherence and diversity diagnostics, stability checks, and structured human labeling [57,58].

3.1. Data Sources and Sampling Frame

The corpus comprises publicly available consumer complaint texts about three major e-commerce marketplaces operating in Turkey: Trendyol [59], Hepsiburada [60], and Amazon (Turkey) [61]. Data were collected from two public channels: (i) posts published on Şikayetvar and (ii) user reviews of the official marketplace mobile applications on Google Play [62–64] and the Apple App Store [65–67]. The sampling window spans January 2019 to December 2025. The unit of analysis is the complaint text; each record includes the text, platform label, timestamp, and a source indicator. Prior to analysis, complaints were de-identified by removing direct identifiers (e.g., phone numbers, emails, addresses, order or tracking codes), and results are reported only in aggregate form. For brevity in figures and tables, Trendyol is denoted as Marketplace A, Hepsiburada as Marketplace B, and Amazon Turkey as Marketplace C. These labels are used for readability only and do not indicate anonymization.

Data collection followed a targeted sampling frame. An e-commerce complaint was defined as a text describing a marketplace transaction or a platform-mediated service process, including order fulfillment, returns, refunds, customer support, seller conduct, and account security. During collection, a keyword filter was applied to reduce irrelevant content. Keywords covered core failure domains, such as delivery, refund, return, counterfeit, fraud, support, account, and privacy, in both Turkish and English. Texts were retained if at least one marketplace brand was referenced or if a marketplace transaction marker such as “order,” “seller,” “marketplace,” or “shipment” was present. The corpus was constructed using a targeted keyword frame designed to capture marketplace-related complaint narratives, followed by manual screening to remove clearly irrelevant, duplicate, or non-substantive entries. These steps were intended to improve topical relevance rather than to estimate population incidence.

The raw dataset contained 158,420 texts. After preprocessing, language filtering, and deduplication, the final analytical corpus comprised 118,173 texts. Only Turkish and

English texts were retained (Turkish: 71%, English: 29%). Deduplication was performed in two stages: (i) hash-based exact-match removal of identical texts after normalization (lowercasing, whitespace standardization, and removal of URLs and identifier-like strings such as order/tracking codes), retaining the first occurrence; and (ii) near-duplicate removal using character n-gram TF-IDF cosine similarity, collapsing records with similarity ≥ 0.95 and keeping the earliest timestamped text within each cluster. In total, 18,735 texts were removed through deduplication, and a further 21,512 texts were excluded due to language filtering and low-information/spam screening. Figure 1 summarizes the data collection and filtering pipeline, and descriptive characteristics of the corpus are reported in Table 1.

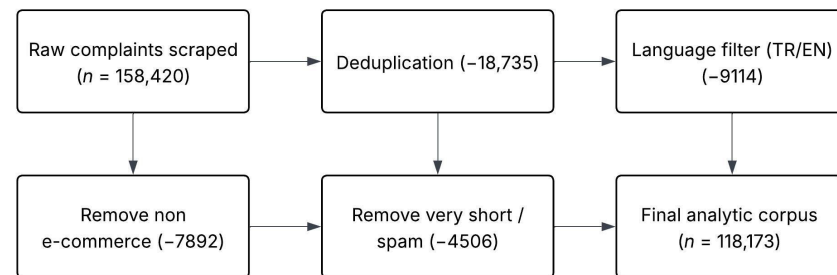


Figure 1. Data collection and filtering pipeline. Raw complaints ($n = 158,420$) were deduplicated ($-18,735$), restricted to Turkish and English (-9114), screened to remove non-e-commerce texts (-7892), and cleaned by excluding very short or spam-like entries (-4506), yielding a final analytic corpus of 118,173 complaints.

Table 1. Corpus summary.

Item	Value
Time window	January 2019 to December 2025 (84 months)
Raw texts collected	158,420
Final analytic corpus	118,173
Language coverage	Turkish (71%), English (29%)
Median text length (tokens)	84
Sources	Complaint portal posts; app-store reviews
Marketplaces covered	3 major marketplaces + complaint portal aggregator
Unit of analysis	Text entry (one record per complaint portal post or app-store review)

Although the texts are publicly available, complaint narratives can contain personal identifiers or transactional details. A de-identification protocol aligned with established guidance for ethical internet research was applied prior to analysis and storage [68]. No attempt was made to re-identify individuals, profile users, or link records to external datasets.

Redaction was rule-based and conservative. Emails, phone numbers, addresses, order IDs, tracking codes, and bank references containing partial card numbers were removed, along with personal names when they appeared in common name-surname patterns. Only de-identified text and derived features (topic probabilities and topic prevalence) were stored. Results were reported at aggregated levels (topic, month, platform) to reduce disclosure risk.

3.2. Text Preprocessing

Complaint texts are short, noisy, and heterogeneous, which can degrade clustering stability when preprocessing is inconsistent [69]. A standardized preprocessing pipeline was applied. Text was normalized through lowercasing, Unicode normalization, whitespace cleanup, and URL removal. Identifier redaction was implemented via regular expressions consistent with the de-identification protocol. Automatic language identification was

performed using the fastText lid.176 language identification model to retain Turkish and English texts only [70]. Empty entries were removed, and texts shorter than 15 whitespace tokens were excluded to avoid unstable representations driven by sparse content.

Stopwords were removed for topic representation only. Turkish and English stopword lists were used, and domain stopwords such as “order” and “platform” were added to reduce uninformative high-frequency terms in c-TF-IDF representations. Stopwords were not removed before embedding because transformer encoders can benefit from contextual function words in short texts [35].

Turkish morphology can inflate lexical variation [71]. As a robustness check, the full pipeline was repeated with Turkish lemmatization using an established Turkish morphological analyzer (e.g., Zemberek [72]), and the stability of topic solutions and temporal patterns was evaluated.

3.3. Topic Modeling with BERTopic

BERTopic was used to extract complaint topics from the multilingual complaint corpus. The method clusters dense semantic document representations and constructs interpretable topic descriptors using class-based TF-IDF (c-TF-IDF) [36]. The embedding step relied on a multilingual sentence-transformer suitable for Turkish and English (paraphrase-multilingual-MiniLM-L12-v2) [35]. Transformer embeddings capture semantic similarity beyond surface word overlap, which is valuable in complaint corpora where similar issues are described using varied phrasing [34].

The modeling pipeline proceeded in four stages. First, each complaint was encoded into a dense vector using the pretrained transformer model [35]. Second, UMAP was applied for dimensionality reduction prior to clustering [73]. Third, reduced embeddings were clustered with HDBSCAN, which accommodates clusters of varying density and assigns outliers to a noise class [74]. Fourth, c-TF-IDF was computed within each cluster to generate representative terms [36], and term lists were refined using a KeyBERT-inspired, diversity-aware representation to reduce redundancy among top words. Outliers were assigned by HDBSCAN to a noise class; low-confidence non-noise assignments were additionally excluded using a probability threshold ($p < 0.15$). Among the non-noise texts, 20,531 (19.0%) had topic-assignment probabilities below the 0.15 threshold and were therefore excluded from prevalence-based summaries. Manual screening was limited to removing residual spam or boilerplate content using pre-defined rules.

Model specifications are summarized in Table 2 to support reproducibility. Parameter settings were fixed prior to topic interpretation to limit the researcher degrees of freedom. Because UMAP includes stochasticity, stability checks re-estimated the pipeline across 10 UMAP random seeds while holding the remaining parameters constant.

Table 2. BERTopic configuration and key parameters.

Component	Specification
Embedding model	paraphrase-multilingual-MiniLM-L12-v2
UMAP	n_neighbors = 15; n_components = 5; min_dist = 0.0; metric = cosine
HDBSCAN	min_cluster_size = 250; min_samples = 10; metric = Euclidean
Topic representation	c-TF-IDF + KeyBERT-inspired representation; top_n_words = 15
Outliers	Assigned by HDBSCAN to a noise class; low-confidence non-noise assignments were additionally excluded using a probability threshold ($p < 0.15$). Manual screening was limited to removing residual spam/boilerplate using pre-defined rules.

Note. c-TF-IDF = class-based TF-IDF; UMAP = Uniform Manifold Approximation and Projection; HDBSCAN = Hierarchical Density-Based Spatial Clustering of Applications with Noise.

3.4. Topic Quality, Robustness, and Validation

Topic quality and robustness were assessed using complementary quantitative diagnostics, cross-run stability checks, multilingual consistency checks, and structured human validation. Quantitative diagnostics focused on semantic coherence (c_v) and topic diversity, defined as the share of unique words across topic descriptors [75]. In addition, topic intrusion and word intrusion tests were conducted on a subsample to assess interpretability [56].

Sensitivity to topic granularity: Candidate solutions were estimated for $k = 10$ –60 topics, and c_v coherence and topic diversity were compared across this range (Figure 2). Coherence peaked around $k = 35$ while topic diversity remained high. Although larger k values further increased diversity, they were associated with a noticeable decline in coherence. Accordingly, $k = 35$ was selected for the main analysis.

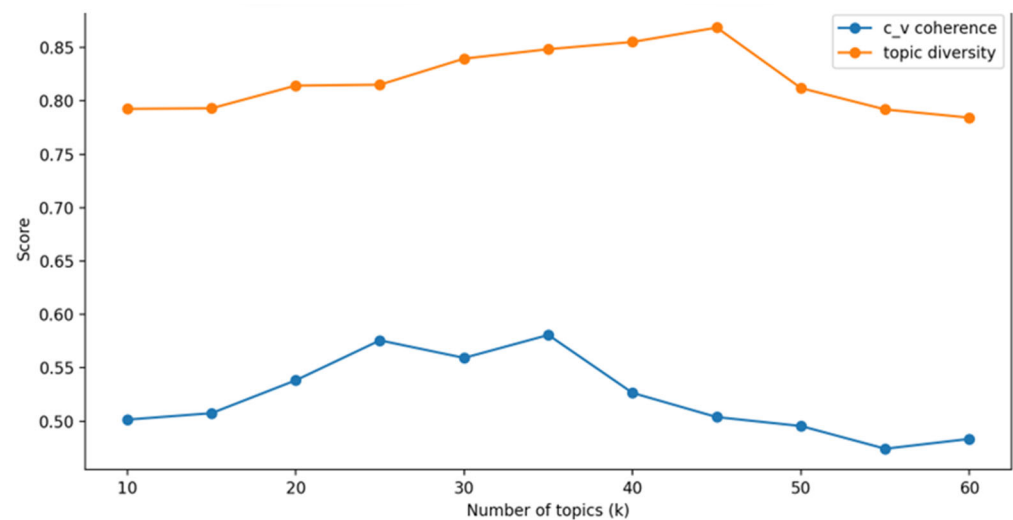


Figure 2. Model selection diagnostics across candidate topic counts (k): c_v coherence and topic diversity. Coherence peaks around $k = 35$, while diversity increases with k and peaks around $k = 45$; selected $k = 35$ as the main specification as it provides the strongest coherence while maintaining high diversity.

Stability across runs and preprocessing variants: To assess whether the identified topics were artifacts of a single stochastic run, the model was re-estimated across 10 UMAP random seeds while holding the embedding model and clustering settings fixed (Table 2). In addition, sensitivity to preprocessing variants was evaluated, including the Turkish-specific preprocessing described in Section 3.2.

Topic solutions were treated as stable when they showed high overlap in top terms (Jaccard similarity ≥ 0.80) and high similarity in topic representations (cosine similarity ≥ 0.70) across runs/specifications. Stability diagnostics are summarized in Table S1 (Supplementary Material).

Multilingual validation (Turkish vs. English): Because the corpus contains both Turkish and English complaints, the consistency of the topic structure across languages was evaluated. The topic model was re-estimated on Turkish-only and English-only subsets, and topics were aligned based on similarity in their representations (e.g., top-term overlap and representation similarity). Alignment was then assessed by comparing (i) the degree of topic correspondence across languages and (ii) whether the dominant macro-themes were preserved. Alignment summary statistics are provided in Table S2a, and topic-level language-stratified summaries and mappings are provided in Table S2b (Supplementary Material).

Human validation and topic labeling: Two independent coders labeled topics using (i) the top c-TF-IDF terms and (ii) 10 representative complaints per topic sampled from high-confidence assignments. Coders followed a shared codebook defining label rules and boundary conditions between closely related topics. Inter-coder agreement was assessed using Cohen's kappa ($\kappa = 0.81$) [76]. Disagreements were resolved through discussion, resulting in the final topic labels used in Section 4. For transparency, the full topic inventory (label, top terms, prevalence, and representative texts) is reported in Table S3a (Supplementary Material). For transparency, representative complaints sampled from high-confidence assignments are provided in Table S3b (Supplementary Material). To formalize the macro-theme grouping procedure, Table S3d (Supplementary Material) reports the topic-to-macro-theme assignment rules, boundary conditions for adjacent categories, and the sampling logic used for the 10 representative complaints per topic. After topic labeling, the macro-theme grouping was finalized through coder-guided adjudication using predefined rules. Because this step was consensus-based rather than conducted as a separate, independent coding round, a distinct macro-theme-level agreement statistic was not computed.

3.5. Measures and Analysis Strategy

BERTopic assigns each complaint to a topic and, when available, provides assignment probabilities. Monthly topic prevalence was estimated in two ways. The primary measure was probability-weighted prevalence for topic k in month t , defined as

$$\text{Prev}_{kt} = \frac{1}{N_t} \sum_{i \in t} p_{ik} \quad (1)$$

where N_t denotes the number of complaints in month t and p_{ik} is the model-based probability that complaint i belongs to topic k . As a robustness check, prevalence was re-estimated using modal assignments as the share of complaints in month t whose most likely topic was k .

Platform topic profiles were computed as topic distributions within each marketplace over the full window and within each year. Cross-platform differences in topic distributions were summarized using Jensen–Shannon divergence and assessed using permutation tests.

RQ1 identified dominant grievance categories. Topic inventories were reported with labels, representative terms, and prevalence. Interpretation was anchored in perceived risk facets and justice concerns [20,47]. Topic-solution quality, stability, and multilingual validation diagnostics are reported in Section 3.4.

Because the corpus is bilingual, language-related sensitivity was assessed at two levels. First, semantic robustness was evaluated by re-estimating the topic model separately on Turkish-only and English-only subsets and aligning topics across languages in Table S2a,b (Supplementary Material). Second, differences in macro-theme prevalence across Turkish and English complaints were tested using a Pearson chi-square test in Table S2c (Supplementary Material). To assess whether these differences persisted after adjustment, language effects were also examined within the multinomial logistic framework used for platform comparisons in Table S2d (Supplementary Material).

RQ2 assessed the temporal change in topic salience. Structural shifts in the monthly prevalence series were detected using PELT changepoint estimation [77]. Pre–post differences were assessed with the Mann–Whitney test and summarized with Cliff's delta [78,79]. Monotonic trends were evaluated with the Mann–Kendall test and Theil–Sen slopes [80]. False discovery rates were controlled using the Benjamini–Hochberg procedure [81]. Sensitivity checks assessed whether inferences were robust to temporal dependence in monthly series. Turning-point contrasts were summarized using fixed six-month pre/post windows

to provide a common reporting frame across themes. As a sensitivity check, the same comparisons were re-estimated using four-month and eight-month windows, together with a dependence-aware monthly comparison. As shown in Table S8 (Supplementary Material), the direction of the main shifts remained unchanged, with the strongest breaks proving robust across specifications, while the decline in Remediation Frictions was interpreted more cautiously.

To triangulate the interpretation of the post-2023 shift in escalation-related narratives, a bilingual dictionary of automation markers was constructed (Table S4, Supplementary Material), and each text was flagged if it contained ≥ 1 marker. Monthly marker prevalence was computed as the share of flagged texts. Pre- and post-2023-02 windows were then compared using the same six-month windowing and Mann–Whitney framework used for macro-theme turning points, with Cliff’s δ reported as the effect size. Results are reported for the full corpus and within Escalation Failures.

RQ3 evaluated platform differences net of composition. Multinomial logistic regression was estimated with modal topic assignment as the outcome and platform as the focal predictor, including month fixed effects and language controls. Results were reported as relative risk ratios with robust standard errors [82], supplemented by Jensen–Shannon divergence and permutation-based p -values.

Finally, substantive conclusions were required to persist under the modal-assignment prevalence estimates. Findings are interpreted as shifts in expressed risk and justice concerns under evolving operational and governance conditions, not as direct estimates of incident rates.

4. Results

4.1. Topic Structure and Macro-Themes

The final BERTopic specification ($k = 35$) yielded 35 interpretable topics and a noise (outlier) class accounting for 8.6% of documents. A targeted manual review of the HDBSCAN noise class likewise did not indicate a coherent omitted emerging-risk category; reviewed cases were predominantly weak fits to existing topics, mixed-issue texts, or low-specificity complaints (Table S3c, Supplementary Material). The noise class corresponds to texts assigned to the HDBSCAN noise cluster during clustering and is reported separately; unless otherwise noted, prevalence-based summaries are computed on the retained, high-confidence set of non-noise documents (i.e., after excluding low-confidence assignments as defined in Section 3.3). Topic labels were assigned based on top keywords and representative complaints and refined using coder definitions. For exposition, the 35 micro-topics were grouped into five macro-themes reflecting dominant risk and trust mechanisms in complaint narratives based on coder-defined conceptual similarity and semantic proximity in the intertopic map. The macro-theme mapping is reported in Table 3.

To address RQ1, the dominant grievance categories are synthesized into topics and macro-themes. The intertopic distance structure corroborates this organization (Figure 3). Fulfillment and remediation topics (T1–T16) occupy a contiguous region, while governance-oriented risks (T30–T35) form a more distinct area characterized by vocabulary around security, consent, and policy-based claims. Escalation-related topics (T24–T29), including automated and scripted support experiences, appear between these regions, consistent with escalation barriers functioning as a procedural justice concern that can convert operational failures into trust judgments [20,50]. In complaints about delays or refunds, the narrative often shifts when the dispute process is described as unresponsive or non-correctable, aligning with procedural fairness as a mechanism shaping legitimacy perceptions.

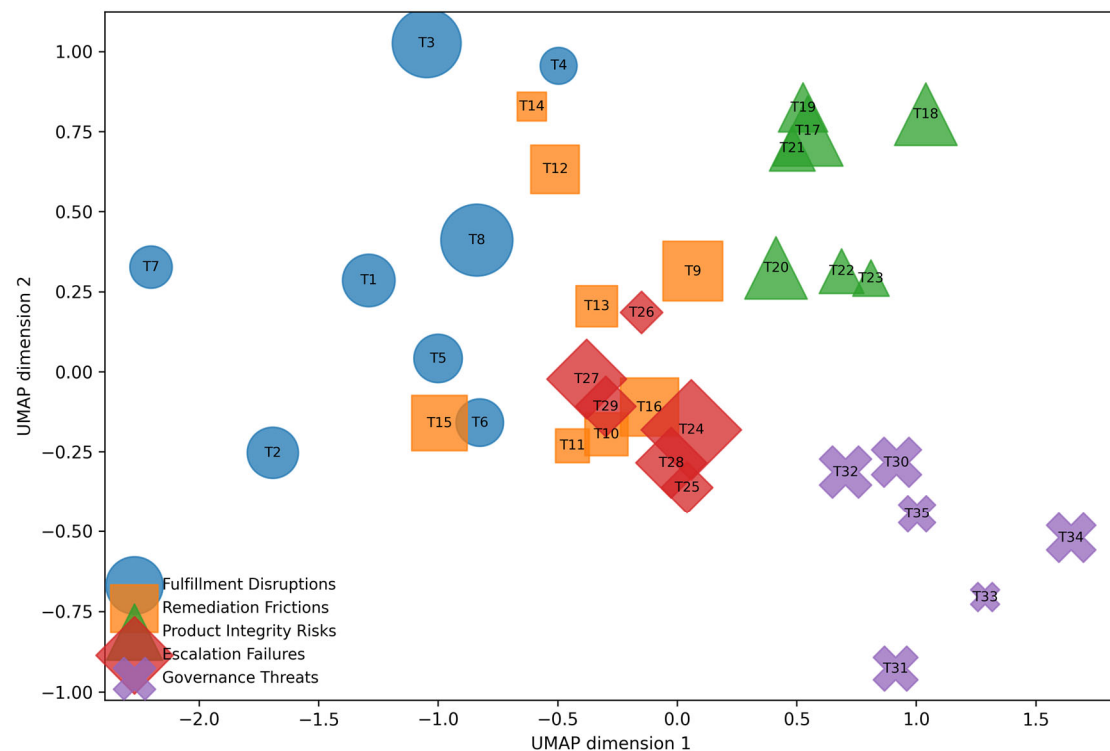


Figure 3. Intertopic map (UMAP projection of topic embeddings). Points represent micro-topics and are labeled by topic codes. Relative proximity reflects semantic similarity in the embedding space. Colors and marker shapes indicate macro-theme membership. Bubble size is proportional to topic prevalence. The outlier (noise) class is excluded from the map.

These distinctions also clarify the boundary between three adjacent grievance domains. Remediation Frictions refer to failures in the execution of restitution once a remedy pathway is available; the focal issue is whether refunds, returns, credits, or reversals are processed in a predictable, transparent, and timely manner. Escalation Failures arise when access to meaningful review is itself impaired; the central problem is not only delayed recovery, but also blocked voice, weak correctability, absent human handoff, or repetitive case handling that prevents the dispute from being substantively heard. Governance Threats differ from both because they concern risks attributed to the platform’s institutional safeguards and rule system, including account security, fraud prevention, privacy, consent, and manipulative interface design.

Cross-language robustness checks indicate that the pooled topic inventory is semantically stable across Turkish and English complaints in Table S2a,b (Supplementary Material). However, macro-theme prevalence differs significantly by language ($\chi^2(4) = 3124.84$, $p < 0.001$, Cramér’s $V = 0.189$; Table S2c, Supplementary Material). English-language complaints are relatively more concentrated in Remediation Frictions and Governance Threats, whereas Turkish-language complaints are relatively more concentrated in Fulfillment Disruptions and Product Integrity Risks.

Table 4 reports the most prevalent micro-topics in the analytic corpus. The distribution is not dominated by a single issue category: salience is spread across fulfillment problems and monetary remediation, with sizeable contributions from product integrity concerns and escalation failures. This pattern is consistent with institution-based trust in marketplaces, where platform-level structural assurances are evaluated through the reliability of delivery, refund execution, and dispute resolution [83].

Table 3. Macro-themes and constituent micro topics.

Macro-Theme	Definition (Used for Grouping)	Constituent Micro Topics
Fulfillment Disruptions	Failures in delivery execution and shipment control, including delays, loss, and misdelivery.	T1 Delivery delays, T2 Missed slots, T3 Tracking failures, T4 Lost parcels, T5 Wrong/partial delivery, T6 Damaged package on arrival, T7 Courier conduct complaints, T8 International shipping/customs issues
Remediation Frictions	Frictions in monetary remediation and return workflows, including delays, denials, and fee disputes.	T9 Refund pending, T10 Partial refund disputes, T11 Chargeback guidance and reversals, T12 Return pickup and label issues, T13 Return denied, T14 Restocking and hidden fees, T15 Wallet or store credit blocks, T16 Promotional refund and points disputes
Product Integrity Risks	Concerns about authenticity and information asymmetry, including misleading listings and quality mismatch.	T17 Counterfeit/inauthentic goods, T18 Used item sold as new, T19 Listing-photo mismatch, T20 Specification mismatch, T21 Wrong variant/size/color, T22 Seller misrepresentation and bait listings, T23 Warranty eligibility disputes at purchase stage
Escalation Failures	Breakdowns in customer support access and case handling, including chatbot loops and blocked escalation.	T24 Chatbot loop and scripted replies, T25 Human agent unreachable, T26 Escalation failure and handoff barriers, T27 Ticket auto-closure without review, T28 Slow response and backlog, T29 Inconsistent resolutions across contacts
Governance Threats	Risks tied to account integrity, fraud, data handling, and manipulative interface practices.	T30 Unauthorized transactions, T31 Account takeover and lockouts, T32 OTP and verification failures, T33 Phishing and social engineering incidents, T34 Privacy consent and data sharing concerns, T35 Dark patterns and subscription traps

Table 4. Top 12 most prevalent micro-topics.

Topic	Label	Prevalence (%)	Representative Keywords
T1	Delivery delays	7.80	delay, late, promised, courier, slot, tracking
T9	Refund pending	7.10	refund, pending, days, waiting, bank, reversal
T17	Counterfeit/inauthentic goods	6.40	fake, counterfeit, authenticity, replica, brand, serial
T12	Return pickup and label issues	5.60	return, label, pickup, courier, barcode, shipment
T19	Listing-photo mismatch	5.10	photo, description, misleading, different, mismatch, image
T24	Chatbot loop and scripted replies	4.70	chatbot, script, same, automated, loop, template
T4	Lost parcels	4.30	lost, missing, not delivered, depot, investigation, claim
T25	Human agent unreachable	4.10	agent, unreachable, call, wait, line, support
T30	Unauthorized transactions	3.90	unauthorized, card, payment, fraud, security, charge
T13	Return denied	3.60	denied, policy, rejected, reason, inspection, condition
T11	Chargeback guidance and reversals	3.40	chargeback, bank, dispute, reversed, proof, process
T27	Ticket auto-closure without review	3.20	ticket, closed, resolved, no review, status, system

Note. Prevalence values are percentages of the analytic corpus used for prevalence estimation (i.e., computed on the retained, high-confidence non-noise set as defined in Sections 3.4 and 3.5); only the top 12 topics are shown, and values are rounded. Keywords are shown in English for readability; bilingual descriptors and representative texts for all topics are reported in Supplementary Table S3a,b.

Figure 3 presents an intertopic map based on the UMAP projection of topic embeddings. Relative proximity indicates semantic similarity in complaint narratives. The noise (outlier) class is excluded from the map.

4.2. Theme Narratives

Theme narratives are based on topics estimated from the pooled corpus combining Şikayet-var posts and app-store reviews; source differences are addressed via composition controls (Section 3.5) and robustness checks reported in Tables S1 and S2 (Supplementary Material).

4.2.1. Fulfillment Disruptions

Fulfillment Disruptions captures complaint narratives in which the marketplace promise fails at the point of physical execution. The reported breakdown typically concerns delivery delays, missed delivery windows, stalled tracking updates, lost parcels, wrong or partial delivery, damage on arrival, courier conduct, or international shipping and customs problems (T1–T8). The language in these texts is concrete and sequential, often built around promised dates, scan events, delivery slots, depot holds, and handoff points that signal where control was perceived to be lost.

Fulfillment has long been treated as a core dimension of electronic service quality because it connects online ordering to the outcome that matters most to consumers: receiving the correct item, in usable condition, within the promised time frame [84–86]. Complaints in this theme do more than register inconvenience. They reveal performance and time risk as experienced by consumers, where uncertainty is amplified by limited visibility into last-mile processes and by dependence on platform-mediated coordination [47]. These narratives also function as a test of institution-based trust. When delivery execution appears unreliable, structural assurances provided by the marketplace, such as guarantees [87], dispute systems [88], and policy claims [89], are evaluated as less credible because the platform is perceived as unable to ensure dependable fulfillment at scale [43,44]. Sensitivity to system-level strain is expected in this theme because logistics disruptions can propagate and amplify bottlenecks across networks [90], creating clustered periods of failure rather than isolated incidents [91].

A marked temporal shift is observed for Fulfillment Disruptions. A turning point is detected in 2020-04, with the mean monthly prevalence increasing from 15.16% in the six months before the break to 18.45% in the six months after ($\Delta = 3.28$ percentage points; Mann–Whitney $p = 0.002$; Cliff's $\delta = 0.75$) (Table 5). This change indicates a broad reweighting toward fulfillment-related grievances in early 2020, consistent with a period in which execution reliability became a dominant trust-relevant signal in complaint narratives. Platform profiles also differ. Fulfillment Disruptions accounts for 18.24% of complaints on Marketplace A, compared with 14.42% on Marketplace B and 12.36% on Marketplace C (Figure 4), indicating that fulfillment breakdowns occupy a larger share of the expressed complaint agenda within Marketplace A.

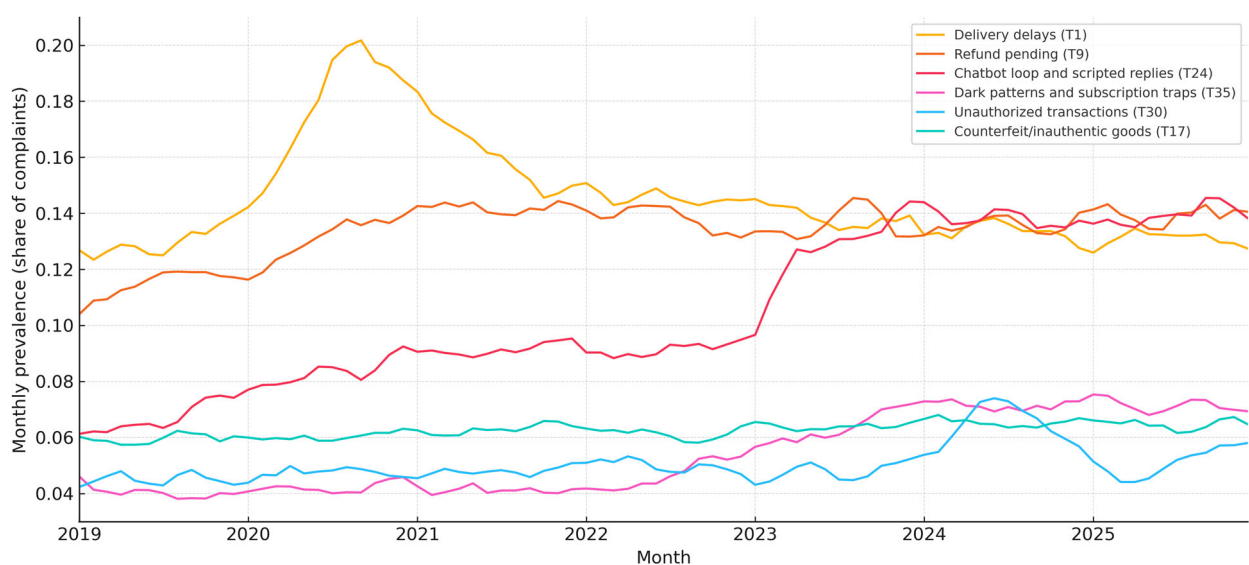


Figure 4. Prevalence of selected micro-topics over time (three-month rolling mean).

Table 5. Turning points in macro-theme prevalence.

Macro-Theme	Estimated Break	Pre Mean (%)	Post Mean (%)	Δ (pp)	Mann–Whitney p	Cliff's δ
Fulfillment Disruptions	2020-04	15.16	18.45	3.28	0.002	0.75
Remediation Frictions	2021-12	24.44	23.06	−1.38	0.010	−0.63
Product Integrity Risks	2020-04	24.45	22.12	−2.32	<0.001	−0.88
Escalation Failures	2023-02	10.11	12.45	2.34	<0.001	0.93
Governance Threats	2022-06	22.72	24.16	1.44	0.002	0.75

4.2.2. Remediation Frictions

Remediation Frictions captures complaint narratives in which the primary dispute concerns the return of money or value after a transaction has gone wrong. In this theme, the breakdown is framed through refund timelines that extend for days or weeks, partial refunds that do not match expectations, chargeback instructions and reversals, failed return pickups, label or barcode problems, return denials, and disputes about restocking fees, wallet credits, or promotional points. The emphasis is placed on financial exposure and on the uncertainty of recovery, aligning closely with the financial risk facet in e-services adoption research [92,93]. In service recovery scholarship, this stage functions as a pivotal moment because perceived recovery quality shapes satisfaction, negative word of mouth, and relationship outcomes after failure [94,95]. Justice theory clarifies why these complaints are often read as more than administrative friction [96]. When procedures are experienced as slow, inconsistent, opaque, or difficult to correct, trust can deteriorate even when some reimbursement is eventually issued [97,98]. In digital channels, recovery quality is also tied to compensation and contact options, which are frequently mediated through interfaces and scripted workflows rather than direct interpersonal exchange [84]. In marketplace settings, remediation complaints therefore act as a test of institution-based trust: platform policies, guarantees, and dispute mechanisms are evaluated as structural assurances when sellers and logistics partners are not directly accountable to buyers [99].

A substantive share of platform-level complaint agendas is allocated to Remediation Frictions. In the marketplace profiles, this macro-theme accounts for 28.91% to 32.78% of complaints, with the highest concentration observed for Marketplace B (Figure 4). A turning point is detected in 2021-12, where mean monthly prevalence declines from 24.44% to 23.06% ($\Delta = -1.38$ percentage points; Mann–Whitney $p = 0.010$; Cliff's $\delta = -0.63$) (Table 5). This breakpoint indicates a measurable reweighting of the complaint agenda after late 2021: refund and return frictions remain central, but they occupy a smaller share of monthly complaints as attention shifts toward other risk and trust concerns in later periods.

4.2.3. Product Integrity Risks

Product Integrity Risks captures complaint narratives in which the central allegation is that the delivered item does not match what was promised at the point of purchase. These complaints focus on authenticity and information-quality breakdowns, including counterfeit or inauthentic goods, used items sold as new, listing and photo mismatches, specification mismatches, wrong variants (size, color, model), bait listings, and warranty eligibility disputes raised at the purchase stage (T17–T23). At its core, this theme reflects an information asymmetry problem: consumers are asked to commit before direct inspection [100], so credibility hinges on the reliability of listing claims and on the institutional safeguards that are expected to discipline sellers and correct misrepresentation [101]. When those safeguards are perceived as weak or inconsistently enforced, the complaint is rarely confined to product dissatisfaction [102,103]. Instead, it becomes a trust-relevant statement about the marketplace as an institution, because structural assurances, governance rules, and enforcement capacity define whether promises made in the interface can be treated as

dependable [104]. This framing aligns with perceived risk scholarship, where authenticity and misrepresentation are experienced as consequential risks that erode willingness to transact even in the presence of convenience and price benefits [105–107].

A temporal reweighting is observed for this macro-theme. A turning point is detected in 2020-04, with mean monthly prevalence declining from 24.45% in the six months before the break to 22.12% in the six months after ($\Delta = -2.32$ percentage points; Mann–Whitney $p < 0.001$; Cliff's $\delta = -0.88$) (Table 5). The direction and magnitude of this change indicate that integrity-related disputes remained prominent but occupied a smaller share of the complaint agenda after early 2020, coinciding with periods in which operational disruption and recovery frictions became more salient in complaint narratives. Platform profiles also differ meaningfully: Product Integrity Risks represents 19.87% of complaints on Marketplace A, 21.11% on Marketplace B, and 24.98% on Marketplace C (Figure 4), indicating that authenticity and listing-accuracy disputes constituted a larger portion of the expressed complaint agenda within Marketplace C.

4.2.4. Escalation Failures

Escalation Failures captures complaint narratives in which the core problem is not only the initial service failure, but the inability to obtain meaningful case handling afterward. These texts describe being trapped in chatbot loops, receiving templated replies that do not address the claim, failing to reach a human agent, and encountering ticket closures that are experienced as automatic rather than deliberative. Additional subtopics include prolonged backlogs and inconsistent outcomes across repeated contacts (T24–T29). In service recovery terms, the complaint is framed around recovery access and recovery quality: when contact channels do not provide a credible route to remedy, the failure is reinterpreted as institutional neglect rather than a fixable operational mistake [108].

This theme maps closely onto procedural justice. The narratives emphasize lack of voice, weak correctability, and inconsistency across interactions, which are classic conditions under which legitimacy judgments deteriorate even when monetary or logistical outcomes remain theoretically possible. In platform-mediated commerce, escalation is also a governance function because the platform defines who can be reached, which claims are routed, how evidence is evaluated, and whether exceptions can be granted [109,110]. The expansion of automated support systems can improve speed for routine inquiries [111], yet it can also introduce failure modes that are specific to AI-mediated service [112]: repeated scripts, misclassification, blocked handoffs, and opaque queueing that is perceived as non-responsive [54]. Escalation Failures therefore represents a mechanism through which operational incidents are converted into trust-relevant judgments about fairness, accountability, and institutional reliability.

A durable reweighting of complaint attention is observed for this macro-theme. A turning point is detected in 2023-02, with mean monthly prevalence increasing from 10.11% in the six months before the break to 12.45% in the six months after ($\Delta = 2.34$ percentage points; Mann–Whitney $p < 0.001$; Cliff's $\delta = 0.93$) (Table 5). To triangulate whether this shift is accompanied by an increasing presence of scripted/automated support in user narratives, the monthly prevalence of bilingual automation markers was examined (Table S4, Supplementary Material). Table 6 summarizes the pre–post comparison around 2023-02 for the full corpus and within Escalation Failures.

Platform profiles also differ materially: Escalation Failures account for 13.68% of complaints on Marketplace A, compared with 9.74% on Marketplace B and 8.21% on Marketplace C (Figure 5). In the model-based analysis, a higher relative likelihood of Escalation Failures is retained for Marketplace A compared with Marketplace B after month fixed effects and language controls are included (RRR = 1.11, $p = 0.002$). The

post-2023 rise in escalation-related complaints is consistent with increased friction around automated support and blocked handoff, although the supplementary evidence supports an association rather than a definitive causal interpretation.

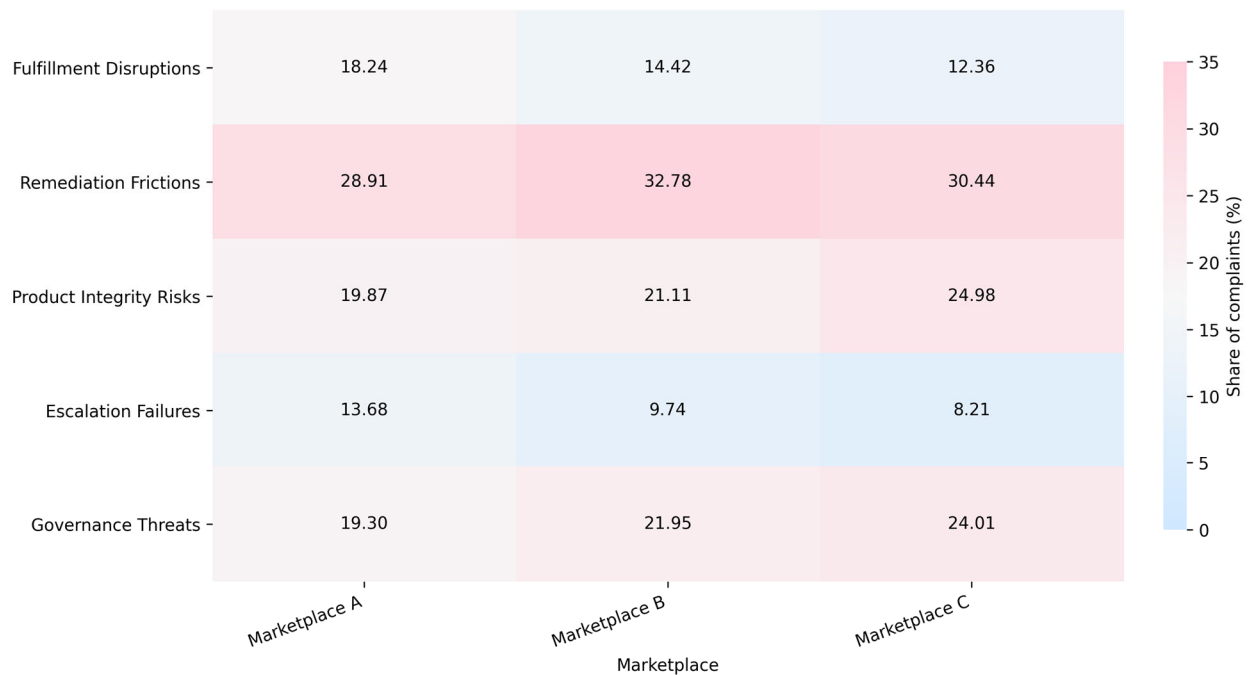


Figure 5. Macro-theme prevalence profiles across marketplaces. Cell values indicate the share of complaints within each marketplace (%); columns sum to 100%. Note. Marketplace A = Trendyol; Marketplace B = Hepsiburada; Marketplace C = Amazon Turkey. The A/B/C labels are used for presentation clarity only.

Table 6. Automation-marker triangulation (pre–post around 2023-02).

Sample	Pre Mean Monthly Share (%)	Post Mean Monthly Share (%)	Δ (pp)	Mann–Whitney p	Cliff's δ
Full corpus	5.2	7.9	+2.7	<0.001	0.83
Escalation Failures only	9.6	14.8	+5.2	<0.001	0.92

Note. Monthly shares are the percentage of texts containing ≥ 1 bilingual automation marker. Pre and post windows cover six months before and after 2023-02 ($n = 6$ months per period). Two-sided Mann–Whitney tests are applied to monthly shares, and Cliff's δ is reported as the effect size.

4.2.5. Governance Threats

Governance Threats captures complaint narratives in which harm is attributed to the platform's governance and safeguards rather than to a single transaction mishap. The emphasis is placed on account of integrity, fraud exposure, and control over personal data. Typical texts describe unauthorized transactions, account takeover and lockouts, failures in one-time password or verification flows, phishing and social engineering incidents, privacy and consent disputes, and manipulative interface practices such as subscription traps and misleading enrollment screens (T30–T35). Unlike fulfillment failures, these complaints are often read as violations of baseline institutional expectations: secure access, predictable authentication, transparent consent, and credible protection against misuse.

This macro-theme is tightly connected to institution-based trust in platform-mediated exchange because it concerns the integrity of the system that enables transactions in the first place [113,114]. Information security research highlights that incentives and externalities can lead to underinvestment in protection unless governance mechanisms are credible and consistently enforced [115]. Privacy scholarship similarly shows that trust is shaped

by perceived control, notice, and the alignment between stated practices and experienced outcomes, especially when information disclosure is unavoidable for participation [116,117]. When complaints in this theme document verification failures, account recovery dead ends, unexplained charges, or consent disputes, the platform is evaluated as an institutional actor responsible for security controls and procedural safeguards, not as a neutral intermediary [98,118]. Dark-pattern narratives sharpen this governance framing by portraying interface design as a mechanism that steers decisions in ways users experience as deceptive or difficult to reverse [119,120].

Governance Threats occupies a growing share of the complaint agenda. A turning point is detected in 2022-06, with mean monthly prevalence rising from 22.72% in the six months before the break to 24.16% in the six months after ($\Delta = 1.44$ percentage points; Mann–Whitney $p = 0.002$; Cliff's $\delta = 0.75$) (Table 5). Platform profiles also differ: Governance Threats account for 19.30% of complaints on Marketplace A, 21.95% on Marketplace B, and 24.01% on Marketplace C (Figure 5). In the model-based analysis, Governance Threats remains more likely on Marketplace C than on Marketplace B after month fixed effects and language controls are included (RRR = 1.28, $p < 0.001$). This pattern indicates that governance and security concerns represent a distinct and consequential component of expressed risk, with salience that increases after mid-2022 and varies systematically across marketplace ecosystems.

4.3. Temporal Dynamics and Turning Points

To address RQ2, the evolution of complaint topics over time is examined, and identifiable turning points in topic prevalence are assessed. Figure 4 plots monthly prevalence for six salient micro-topics, shown as a three-month rolling mean to reduce month-to-month noise. Delivery-delay complaints rise sharply in early 2020 and then decline toward a lower plateau. Refund-pending complaints remain elevated after the early-2020 disruption period. Chatbot-loop and scripted-reply complaints increase after early 2023 and stay higher through the end of the window. Dark-pattern complaints rise from mid-2022, security-related complaints increase around early 2024, and counterfeit/inauthentic-goods complaints remain comparatively steady with gradual drift rather than abrupt shifts.

Turning points were estimated on monthly macro-theme prevalence series and validated using distributional pre–post comparisons. Table 5 reports the estimated break month, mean prevalence in the six months before and after the break, the level change in percentage points, and nonparametric evidence. The largest increase is observed for Fulfillment Disruptions (break = 2020-04; $\Delta = 3.28$ pp). Escalation Failures show a further upward shift after early 2023 (break = 2023-02; $\Delta = 2.34$ pp), and Governance Threats rise after mid-2022 (break = 2022-06; $\Delta = 1.44$ pp). Declines are observed for Remediation Frictions (break = 2021-12; $\Delta = -1.38$ pp) and Product Integrity Risks (break = 2020-04; $\Delta = -2.32$ pp). Effect sizes are large for the strongest shifts (absolute Cliff's $\delta \geq 0.75$), indicating that the detected breaks reflect broad changes in the monthly prevalence distributions rather than isolated spikes. As a source-composition robustness check, the major pooled breakpoints were also re-examined separately for complaint-portal texts and app-store reviews. The results, reported in Table S7 (Supplementary Material), show that the main shifts remain directionally consistent across source types, although their magnitude varies.

The most pronounced temporal break occurs in 2020-04, which coincides with the onset of the COVID-19 disruption period. The sharp increase in Fulfillment Disruptions, together with the concurrent decline in Product Integrity Risks, suggests that complaint attention shifted toward delivery execution, delay, and logistics reliability during the early pandemic period. Given the observational design, this pattern is interpreted as pandemic-consistent rather than as a strict causal effect.

As a triangulation check, the monthly prevalence of texts containing ≥ 1 bilingual automation marker is computed (see Table 6); the dictionary, monthly series, and pre–post tests are provided in Tables S4–S6 (Supplementary Material).

4.4. Platform Differences

To address RQ3, macro-theme distributions were compared across marketplaces. Figure 5 shows systematic reweighting of complaint agendas across platforms. Marketplace A exhibits higher shares of Fulfillment Disruptions (18.24%) and Escalation Failures (13.68%) than Marketplace B (14.42%; 9.74%) and Marketplace C (12.36%; 8.21%). Marketplace B allocates a larger share to Remediation Frictions (32.78%) than Marketplace A (28.91%). Marketplace C shows higher shares of Product Integrity Risks (24.98%) and Governance Threats (24.01%) than Marketplace A (19.87%; 19.30%), indicating greater salience of authenticity, information quality, and governance-related concerns in that ecosystem.

These differences are supported by divergence-based and model-based tests. Pair-wise Jensen–Shannon divergence is largest between Marketplace A and Marketplace C (JSD = 0.0132), followed by Marketplace A and Marketplace B (JSD = 0.0056), with Marketplace B and Marketplace C more similar (JSD = 0.0029); permutation tests indicate statistically reliable separation (Table 7). In multinomial logistic regression with month fixed effects and language controls, platform effects are jointly significant ($\chi^2(8) = 214.6$, $p < 0.001$). With Fulfillment Disruptions as the reference category, Escalation Failures are more likely on Marketplace A than on Marketplace B (RRR = 1.11, $p = 0.002$). Product Integrity Risks and Governance Threats are more likely on Marketplace C than on Marketplace B (RRR = 1.38, $p < 0.001$; RRR = 1.28, $p < 0.001$), while Remediation Frictions are more likely on Marketplace B than on Marketplace A (RRR = 1.43, $p < 0.001$). Full model estimates are reported in Table S9 (Supplementary Material). Given the large corpus, the multinomial results are interpreted primarily in terms of effect direction and relative magnitude rather than statistical significance alone. In substantive terms, the largest platform differences are observed for Product Integrity Risks, Governance Threats, and Remediation Frictions, whereas the Escalation Failures contrast between Marketplaces A and B is statistically reliable but comparatively modest in size.

Table 7. Pairwise Jensen–Shannon divergence of macro-theme distributions.

Comparison	Jensen–Shannon Divergence	Permutation p
Marketplace A vs. Marketplace B	0.0056	0.0001
Marketplace A vs. Marketplace C	0.0132	0.0001
Marketplace B vs. Marketplace C	0.0029	0.0001

Note. Permutation p -values are reported to four decimals; values of 0.0001 indicate $p \leq 0.0001$.

Language also remained jointly significant in the adjusted multinomial specification (Joint Wald $\chi^2(4) = 148.72$, $p < 0.001$; Table S2d, Supplementary Material). Relative to Turkish complaints, English complaints were more likely to fall into Remediation Frictions (RRR = 1.52, 95% CI [1.45, 1.59], $p < 0.001$) and Governance Threats (RRR = 1.86, 95% CI [1.77, 1.96], $p < 0.001$), and less likely to fall into Product Integrity Risks (RRR = 0.82, 95% CI [0.78, 0.86], $p < 0.001$) and Escalation Failures (RRR = 0.93, 95% CI [0.89, 0.98], $p = 0.006$), relative to Fulfillment Disruptions. This adjusted pattern is substantively consistent with the unadjusted comparison and indicates that language differences reweight macro-theme prevalence without overturning the pooled topic structure.

5. Discussion

An implication to consider for trust and perceived risk theory is that complaint salience tracks the risk facets that are most exposed at a given time. The sharp increase in Fulfillment Disruptions at the 2020-04 breakpoint aligns with heightened time and performance risk, where consumers face uncertainty about delivery execution and limited visibility into last-mile processes [47]. In marketplace settings, these risks are evaluated through institution-based trust, since platforms act as structural assurance providers through rules, guarantees, and dispute systems [43,44]. When fulfillment appears unreliable, the credibility of platform assurances is tested more strongly, even if the marketplace is not the direct carrier. The parallel decline in Product Integrity Risks after the same breakpoint suggests agenda substitution: authenticity and misrepresentation remain important, yet a system-wide disruption can re-prioritize what consumers choose to voice in public complaints. This supports a compositional view of risk narratives: the most salient grievance category is not necessarily the most frequent underlying failure, but the one that feels most consequential or least controllable in that period.

A second and more novel contribution concerns the emergence and growth of Escalation Failures as a distinct macro-theme. Service failure and recovery research emphasizes that post-failure response quality shapes satisfaction and downstream relationship outcomes [9,10]. Justice theory clarifies why escalation narratives often sound more severe than the initiating incident: lack of voice, weak correctability, and inconsistency in procedures are conditions under which legitimacy judgments deteriorate [20,50]. The post-2023 increase in Escalation Failures suggests a shift in where breakdowns are perceived, from “slow resolution” to “blocked access to resolution.” This pattern is consistent with users increasingly describing front-line support as scripted or automated, which can reduce friction for routine cases while also being associated with failure modes such as looping responses, misclassification, and delayed or blocked handoff to humans [12,29,54,121]. In platform contexts, escalation is not a peripheral service attribute; it is part of governance because the platform defines routing, evidence requirements, and the conditions under which exceptions can be granted [122]. The results extend platform trust research by showing that automation and procedural constraints reshape grievances via response speed and perceived limits on voice and correctability.

A third insight is the salience and growth of Governance Threats, which concentrates risks tied to account integrity, fraud exposure, privacy, consent, and manipulative interface practices. These complaints target the platform as an institutional actor responsible for the integrity of the transaction environment, not just an intermediary connecting buyers and sellers. This is aligned with institution-based trust frameworks, where the platform’s safeguards substitute for direct interpersonal trust under uncertainty [43,44]. The rise in Governance Threats after the 2022-06 breakpoint is consistent with a broader shift toward system-level risk narratives, including security and privacy concerns. Research on the economics of information security highlights how incentives and externalities can lead to underinvestment unless governance mechanisms are credible and consistently enforced [115]. Privacy scholarship similarly emphasizes control, notice, and alignment between stated practices and experienced outcomes as foundations of trust [116,117]. Dark-pattern complaints sharpen this framing by positioning interface design as a governance instrument that steers choice in ways users experience as deceptive or difficult to reverse [120,123,124]. The results suggest that governance and security are not “edge” topics in the complaint ecosystem; they represent a stable cluster that gains prominence over time.

Platform comparisons add an additional layer by showing that grievance compositions differ systematically across marketplaces even after accounting for time variation and bilingual corpus composition. Marketplace A’s higher shares of Fulfillment Disruptions

and Escalation Failures are consistent with an agenda shaped by operational execution and post-failure procedural access, while Marketplace B's higher Remediation Frictions indicate a stronger concentration on monetary recovery workflows. Marketplace C's higher Product Integrity Risks and Governance Threats indicate greater salience of authenticity, information quality, and system integrity narratives in that ecosystem. These differences align with the multi-target nature of trust in marketplaces, where consumers evaluate sellers and platform governance simultaneously [40,42]. Complaint narratives make these targets visible because texts assign responsibility, describe what remedy pathways were available, and reveal which assurances were perceived as credible. The platform-level patterning of topics provides an empirical signature of governance and service design differences that is hard to capture with cross-sectional surveys.

Methodologically, the study shows the value of multilingual text-as-data for theory-driven measurement of risk and trust narratives. Topic modeling has been used to extract themes from large corpora without exhaustive hand coding [33], and recent work demonstrates how embeddings improve semantic coherence in short, noisy texts [34,35]. The BERTopic workflow strengthens interpretability through class-based TF-IDF representations and clustering of dense semantic vectors [36]. The additional stability checks, multilingual consistency checks, and structured human labeling improve confidence that the identified categories correspond to meaningful grievance structures rather than artifacts of a single run or a language-specific vocabulary [56,75]. This strengthens the bridge between computational measurement and theory: topic prevalence can be interpreted as changing salience of risk facets and justice concerns, rather than a purely descriptive taxonomy.

Interpretation should remain bounded by what complaint data can support. Complaint corpora capture voiced dissatisfaction rather than the full distribution of negative experiences, and complaint behavior reflects opportunity, motivation, and perceived efficacy of voicing [55]. The analysis therefore speaks most directly to the composition of expressed grievances in public channels, not to incident rates. Platform differences can reflect differences in user base, channel mix, and reporting norms as well as differences in operational performance. Turning points identify structural changes in prevalence series but do not, on their own, establish causal drivers. These boundaries do not weaken the central contribution, which is to provide a high-granularity, longitudinal view of how risk and trust narratives are framed and reweighted in marketplace complaints under evolving operational conditions and governance regimes.

The main theoretical message is that marketplace trust is increasingly shaped by procedural access and governance integrity, as well as fulfillment and refunds. Complaint agendas reveal where consumers perceive control, fairness, and accountability to break down and how these perceptions shift when platforms scale, automate, and adjust governance. This sets up a clear foundation for the separate implications section by identifying the mechanisms that move complaint narratives from operational incidents to institutional trust judgments.

Although the empirical setting is Turkey, the mechanisms highlighted by the findings are not unique to Turkey. Fulfillment reliability, remedy predictability, escalation access, and governance integrity are core trust levers in many marketplace environments. What is likely to vary across countries is the relative salience of these themes, depending on logistics maturity, platform structure, consumer protection enforcement, and the extent of service automation. For this reason, the complaint-to-action framework should be read as a transferable diagnostic template rather than a Turkey-specific ranking of interventions.

6. Managerial Implications

The findings translate into practical guidance for marketplace operators because the complaint agenda shifts across failure types and across platforms. The point is not to treat complaint prevalence as incident rates, but to use changes in topic salience as a high-signal view of where customers perceive reliability, recovery, escalation access, and governance to be breaking down. Table 8 summarizes a compact “complaint-to-action” cheat sheet that links each macro-theme to a priority lever and a minimal KPI set. The actions listed in Table 8 should therefore be read as plausible managerial responses inferred from complaint salience patterns, not as tested causal remedies or validated intervention effects.

Table 8. Complaint-to-Action Cheat Sheet.

Theme (Macro)	Sentinel Topics (Examples)	What to Do First (Priority Lever)	What to Monitor (1–2 KPIs)
Fulfillment Disruptions	T1 Delivery delays; T4 Lost parcels	Add proactive exception handling for stalled shipments + clearer tracking states	On-time delivery; scan-latency/stalled-shipment rate
Remediation Frictions	T9 Refund pending; T12 Return pickup/label	Publish refund/return SLAs + status transparency; separate “auto-approve” vs. “investigate”	Refund cycle time; % refunds within SLA
Product Integrity Risks	T17 Counterfeit; T19 Listing-photo mismatch	Fast quarantine/takedown for high-risk listings + repeat-seller enforcement	Time-to-takedown; repeat-offender rate
Escalation Failures	T24 Chatbot loop; T27 Auto-closure	Guarantee human handoff rules + audit and reduce auto-closures for unresolved cases	Time-to-human; recontact rate/% reopened cases
Governance Threats	T30 Unauthorized transactions; T35 Dark patterns	Harden account recovery and fraud response + audit subscription/consent flows	Account recovery success; cancellation completion rate
Cross-cutting	Watchlist of 6–10 topics	Weekly topic dashboard with thresholds + rapid root-cause drills	Topic z-score vs. baseline; time-to-mitigate

Topic prevalence can be operationalized as a monitoring layer. Platforms may move from reactive case handling to proactive control by tracking a small watchlist of sentinel topics (Table 7) at weekly or monthly cadence. The dashboard should include thresholds relative to each platform’s own baseline, plus a short “time-to-mitigate” measure to prevent recurring spikes from becoming normalized. This creates a repeatable routine: detect a shift, diagnose the likely process bottleneck, deploy a short-horizon fix, and confirm improvement through the same topic signals.

Fulfillment may be better treated as a trust lever rather than a logistics metric. Fulfillment Disruptions remain prominent and show sharp turning-point behavior, which makes them suitable for early-warning monitoring. The quickest wins often come from reducing “last-mile ambiguity”: clearer tracking states, automatic triggers for stalled scans, proactive messages that specify what the platform will do next, and exception workflows that do not require repeated customer contact. These interventions target performance and time risk as customers experience it [48] and protect institution-based trust when the platform is the visible accountability holder [43,44].

Remediation can be made more predictable by productizing refund and return SLAs. Remediation Frictions represent uncertainty of recovery, which customers often interpret as ongoing financial exposure. The managerial lever is predictability: explicit SLAs, transparent case status, and a clean separation between cases that require investigation and cases that can be auto-approved. Where wallet credits, promotional refunds, or partial refunds drive disputes, platforms may reduce avoidable complaints by clarifying convertibility and eligibility rules before customers initiate a return.

Escalation paths can be designed around procedural access. Escalation Failures are actionable because they reflect not only a service breakdown but a breakdown in getting heard. Procedural justice research highlights why these cases become trust-damaging: lack of voice, weak correctability, and inconsistent handling undermine legitimacy judgments [20,50]. If automated support is part of the service model, the core safeguard is a reliable handoff rule: when the system cannot resolve a claim with evidence, customers need a clear route to human review, case ownership, and protection against non-deliberative auto-closure. This reduces repeat contacts and prevents “looping” interactions from becoming a stable source of grievances.

Greater investment in governance integrity becomes important where risks become system-level. Governance Threats point to risks that customers attribute to the platform’s safeguards: account takeover, unauthorized transactions, verification failures, consent disputes, and manipulative interface practices. These are not “edge” issues because they frame the platform as responsible for system integrity. Security economics and privacy research emphasize the need for credible, consistently enforced protections and transparent data practices [115–117]. Dark-pattern complaints are especially reputationally costly because they imply intent; routine audits of enrollment, cancellation, defaults, and consent flows can prevent issues that customer support cannot easily “fix” after the fact [121,124].

Finally, platform-level differences in macro-theme profiles can guide prioritization. A platform with a complaint agenda weighted toward escalation should not spend its primary improvement capacity on minor UI refinements; it needs escalation architecture and case governance. A platform weighted toward integrity disputes needs seller enforcement and listing governance. Using the macro-theme profile as a resource-allocation tool helps ensure that investments match the most trust-relevant pain points visible in customer narratives.

7. Robustness, Limitations, and Future Research

The main findings are not tied to a single modeling or measurement choice. Topic prevalence was estimated using probability-weighted assignments and re-estimated using modal assignments; substantive patterns (theme ordering, turning-point timing, and platform reweighting) were required to hold under both constructions. Topic solutions were selected using coherence–diversity diagnostics and checked for stability across UMAP random seeds and preprocessing variants, with topic consistency evaluated through overlap in top terms and similarity in topic representations. Multilingual validity was examined by re-estimating models on Turkish-only and English-only subsets and aligning topics across languages to verify that the dominant macro-themes were preserved. Temporal inferences were further supported by distributional pre–post comparisons around estimated breakpoints, effect-size reporting, and multiple-testing control.

The corpus reflects voiced dissatisfaction in public channels rather than the full distribution of negative experiences. Complaint behavior depends on opportunity, motivation, and perceived efficacy of voicing, so topic prevalence should be interpreted as the salience of expressed grievances rather than incident rates [55]. Source mix also matters: complaint-portal posts and app-store reviews differ in length, context, and posting incentives, which can shape what gets articulated and how. Platform comparisons can likewise reflect differ-

ences in user composition, channel usage, and reporting norms in addition to operational or governance differences. Turning points identify structural changes in prevalence series, but they do not, on their own, establish causal drivers; multiple real-world changes can coincide with a breakpoint. Finally, topic models provide structured summaries, yet boundaries between closely related topics are not always sharp, and some nuances in complaint narratives (especially rare or highly specific issues) may be absorbed into broader categories or the noise class.

Several extensions would deepen inference and improve actionability. First, linking complaint topics to operational and policy data (logistics KPIs, refund processing logs, seller enforcement actions, support queue metrics, and product or interface changes) would help identify which mechanisms plausibly drive observed turning points. Second, stronger causal designs could be pursued around discrete shocks or rollouts, using quasi-experimental strategies (e.g., difference-in-differences around policy changes, staggered adoption of support automation, or regional variation in logistics constraints). Third, richer cross-channel triangulation, adding social media, call-center transcripts, or in-app support chats, could test whether the same themes appear when incentives to complain differ. Fourth, the escalation-failure mechanism merits dedicated study: future work could map where automated support improves resolution versus where it produces “blocked access” experiences, and which design choices reduce perceived unfairness. Finally, cross-country replication would clarify which patterns are specific to the Turkish marketplace context and which generalize to other platform ecosystems with different governance regimes and consumer protection environments.

Cross-country comparative research would be especially valuable for testing how patterns such as chatbot loops, blocked handoff, and ticket auto-closure vary across institutional settings. These complaint patterns may differ in countries with stronger consumer protection rules, clearer appeal rights, or lower and higher levels of AI adoption in customer service. Such comparisons would help distinguish grievance structures that are broadly general from those that are shaped more directly by national regulatory and service environments.

8. Conclusions

This study examined 118,173 de-identified Turkish and English complaint texts from Turkey’s e-commerce marketplace environment between 2019 and 2025 to address three research questions. First, it identified a stable complaint architecture composed of 35 micro-topics grouped into five macro-themes: Fulfillment Disruptions, Remediation Frictions, Product Integrity Risks, Escalation Failures, and Governance Threats. Second, it showed that complaint salience changed over time, with a marked shift toward Fulfillment Disruptions in 2020-04, rising Governance Threats after 2022-06, and increasing Escalation Failures after 2023-02. Third, it showed that complaint profiles differ systematically across marketplaces even after accounting for time and language composition.

The main contribution of the study is to show that marketplace trust is shaped not only by delivery, refunds, or product integrity but also by what happens after failure. In particular, the macro-theme of Escalation Failures emerges as a distinct grievance domain, indicating that blocked or automated support can transform operational incidents into broader judgments about fairness, accountability, and platform trust. More broadly, the findings suggest that complaint texts can serve as an early warning signal of changing consumer risk and trust narratives. At the same time, the results should be interpreted within the limits of the data on voiced dissatisfaction, the source composition, the targeted text selection, and observational inference. For managers and platform designers, these findings point to plausible priorities for trust protection, including more reliable fulfillment

and remediation, more credible escalation paths, stronger authenticity controls, and more visible governance safeguards.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jtaer21040116/s1>, Table S1: Stability diagnostics for the BERTopic solution ($k = 35$); Table S2a: Language-stratified topic summaries and Turkish–English topic alignment; Table S2b: Topic-level alignment; Table S2c: Macro-theme distribution by language and unadjusted statistical test; Table S2d: Adjusted language effect from multinomial logistic regression; Table S3a: Full topic inventory (label, top terms, prevalence, and representative texts); Table S3b: Representative complaints; Table S3c: Targeted manual review of texts assigned to the HDBSCAN noise class; Table S3d: Condensed codebook for topic-to-macro-theme assignment; Table S4: Bilingual automation-marker dictionary used for triangulation; Table S5: Monthly prevalence of automation markers around 2023-02; Table S6: Pre–post tests for automation-marker prevalence (six-month windows); Table S7: Source-type decomposition of major temporal shifts; Table S8: Temporal sensitivity checks for macro-theme turning points; Table S9: Full multinomial regression reporting for platform comparisons.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from publicly accessible third-party sources (Şikayetvar; Google Play; Apple App Store) and are available from these platforms, subject to their terms of service.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Ghosh, M. Meta-analytic review of online purchase intention: Conceptualising the study variables. *Cogent Bus. Manag.* **2024**, *11*, 2296686. [CrossRef]
2. Ma, D.; Dong, J.; Lee, C.-C. Influence of perceived risk on consumers' intention and behavior in cross-border e-commerce transactions: A case study of the Tmall Global platform. *Int. J. Inf. Manag.* **2025**, *81*, 102854. [CrossRef]
3. Wang, R.; Wang, H.C.; Li, S.G. Predicting the determinants of consumer complaint behavior in e-commerce live-streaming: A two-staged SEM-ANN approach. *IEEE Trans. Eng. Manag.* **2025**, *72*, 1027–1038. [CrossRef]
4. Republic of Türkiye Ministry of Trade. Türkiye'de E-Ticaretin Görünümü Raporu Yayınlandı. Available online: <https://ticaret.gov.tr/duyurular/turkiyede-e-ticaretin-gorunumu-raporu-yayinlandi-06-05-2025> (accessed on 11 February 2026).
5. Turkish Statistical Institute (TurkStat). Survey on Information and Communication Technology (ICT) Usage in Households and by Individuals, 2024. Available online: <https://data.tuik.gov.tr/Bulten/Index?dil=2&p=Survey-on-Information-and-Communication-Technology-%28ICT%29-Usage-in-Households-and-by-Individuals-2024-53492> (accessed on 2 February 2026).
6. Eurostat. E-Commerce Statistics for Individuals. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce_statistics_for_individuals (accessed on 11 February 2026).
7. Fan, L.; Li, S.S.; Wang, C.; Zhang, X.P. Is a chatbot more effective? Investigating the effect of service recovery agents and consumer loss on consumer forgiveness. *J. Theor. Appl. Electron. Commer. Res.* **2026**, *21*, 35. [CrossRef]
8. Chaparro-Peláez, J.; Hernández-García, Á.; Urueña-López, A. The role of emotions and trust in service recovery in business-to-consumer electronic commerce. *J. Theor. Appl. Electron. Commer. Res.* **2015**, *10*, 77–90. [CrossRef]
9. Smith, A.K.; Bolton, R.N.; Wagner, J. A model of customer satisfaction with service encounters involving failure and recovery. *J. Mark. Res.* **1999**, *36*, 356–372. [CrossRef]
10. Tax, S.S.; Brown, S.W.; Chandrashekar, M. Customer evaluations of service complaint experiences: Implications for relationship marketing. *J. Mark.* **1998**, *62*, 60–76. [CrossRef]
11. Hazarika, B.B.; Gerlach, J.; Cunningham, L. The role of service recovery in online privacy violation. *Int. J. E-Bus. Res.* **2018**, *14*, 1–27. [CrossRef]
12. Su, Z.; Ha, H.-Y. Longitudinal impact of perceived fairness after service failures: Evidence from online travel agencies. *Int. J. Hosp. Manag.* **2025**, *128*, 104177. [CrossRef]

13. Ma, Y.; Guo, X.; Su, W.; Fu, G. The evolution of price discrimination in e-commerce platform trading: A perspective of platform corporate social responsibility. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1907–1921. [CrossRef]
14. Sun, M.H.; Zhao, J.C. Behavioral patterns beyond posting negative reviews online: An empirical view. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 949–983. [CrossRef]
15. Gillespie, T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*; Yale University Press: New Haven, CT, USA, 2018.
16. Pasquale, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*; Harvard University Press: Cambridge, MA, USA, 2015.
17. Gao, H.; Liu, H.S.; Zhang, H.; Xu, W.T. Artificial intelligence service and customer loyalty: A moderated mediation role of satisfaction and contextual factors in human–AI interaction. *SAGE Open* **2026**, *16*, 1–13. [CrossRef]
18. Rajesh, J.I.; Rajaguru, R.; Thanigan, J.; Siriwardana, S.; McMurray, A.; Pare, V. Master–servant–partner interaction styles with artificial intelligence chatbots and perceived value co-creation. *Asia Pac. J. Mark. Logist.* **2025**; in press. [CrossRef]
19. Hamidi, H. A model for generative artificial intelligence in customer decision-making process using social interaction. *Telemat. Inform. Rep.* **2025**, *19*, 100237. [CrossRef]
20. Colquitt, J.A. On the dimensionality of organizational justice: A construct validation of a measure. *J. Appl. Psychol.* **2001**, *86*, 386–400. [CrossRef]
21. Fürst, A.; Trissler, L.; Friedrich, R.; Wirtz, J. Service recovery by AI or human agents: Do failure and strategy context matter? *J. Serv. Manag.* **2025**, *36*, 390–418. [CrossRef]
22. Shi, Y.X.; Zhang, B.; Zhang, R.M.; Yu, L.L. Robot service failure: Interaction effect of robot language style and customers' sense of humor on service failure recovery. *J. Serv. Theory Pract.* **2026**, *36*, 145–170. [CrossRef]
23. Ayyildiz, A.Y.; Ayyildiz, T.; Koc, E. The use of ChatGPT in service recovery: Compensating customers. *Technol. Soc.* **2026**, *84*, 103058. [CrossRef]
24. Kwon, A.; Chung, L.N.; Namkung, Y. Can service technologies make you feel justice like human agents? *J. Serv. Mark.* **2025**, *39*, 864–879. [CrossRef]
25. Leong, M.K.; Sidhu, S.K.; Khoo, K.L. AI chatbot service recovery quality and customer experience: A moderated mediation model of continuance usage intention. *J. Consum. Mark.* **2026**, *43*, 195–208. [CrossRef]
26. Feng, Y.; Kim, H.J. Decoding the trust matrix: Unraveling key predictors of consumer trust in AI-generated personalized advertising. *J. Interact. Advert.* **2025**, *25*, 123–138. [CrossRef]
27. Hao, X.Y.; Dong, D.P.; Zhang, Y.X.; Demir, E. When customers know it's AI: Experimental comparison of human and LLM-based communication in service recovery. *J. Mark. Commun.* **2025**; in press. [CrossRef]
28. Lv, J.H.; Chen, C.-M.; Kumari, S.; Li, K.Q. Resource allocation for AI-native healthcare systems in 6G dense networks using deep reinforcement learning. *Digit. Commun. Netw.* **2025**, *11*, 2016–2029. [CrossRef]
29. Ozuem, W.; Ranfagni, S.; Willis, M.; Salvietti, G.; Howell, K. Chatbots, service failure recovery, and online customer experience through lenses of frustration–aggression theory and signaling theory. *J. Serv. Mark.* **2025**, *39*, 493–512. [CrossRef]
30. Pavone, G.; Desveaud, K. Gendered AI in fully autonomous vehicles: The role of social presence and competence in building trust. *J. Consum. Mark.* **2025**, *42*, 240–254. [CrossRef]
31. Archak, N.; Ghose, A.; Ipeirotis, P.G. Deriving the pricing power of product features by mining consumer reviews. *Manag. Sci.* **2011**, *57*, 1485–1509. [CrossRef]
32. Tirunillai, S.; Tellis, G.J. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Mark. Res.* **2014**, *51*, 463–479. [CrossRef]
33. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
34. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186. [CrossRef]
35. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PN, USA, 2019; pp. 3982–3992. [CrossRef]
36. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF–IDF procedure. *arXiv* **2022**, arXiv:2203.05794. [CrossRef]
37. Şikayetvar. Consumer Complaints and Reviews Portal. Available online: <https://www.sikayetvar.com/> (accessed on 16 December 2025).
38. Ji, F.; He, Z.; Cheng, G. Manufacturing enterprise digital platform capabilities: Structural dimensions and scale development. *Technol. Anal. Strateg. Manag.* **2025**, 1–15. [CrossRef]
39. Oesterreich, T.D.; Anton, E.; Hettler, F.M.; Teuteberg, F. What drives individuals' trusting intention in digital platforms? An exploratory meta-analysis. *Manag. Rev. Q.* **2025**, *75*, 3615–3667. [CrossRef]

40. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An integrative model of organizational trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734. [CrossRef]
41. Gefen, D.; Karahanna, E.; Straub, D.W. Trust and TAM in online shopping: An integrated model. *MIS Q.* **2003**, *27*, 51–90. [CrossRef]
42. Kim, D.J.; Ferrin, D.L.; Rao, H.R. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decis. Support Syst.* **2008**, *44*, 544–564. [CrossRef]
43. Pavlou, P.A. Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *Int. J. Electron. Commer.* **2003**, *7*, 101–134. [CrossRef]
44. McKnight, D.H.; Choudhury, V.; Kacmar, C. Developing and validating trust measures for e-commerce: An integrative typology. *Inf. Syst. Res.* **2002**, *13*, 334–359. [CrossRef]
45. Pavlou, P.A.; Gefen, D. Building effective online marketplaces with institution-based trust. *Inf. Syst. Res.* **2004**, *15*, 37–59. [CrossRef]
46. Organisation for Economic Co-operation and Development (OECD). *OECD Online Dispute Resolution Framework*; OECD Public Governance Policy Papers, No. 59; OECD Publishing: Paris, France, 2024. Available online: https://www.oecd.org/en/publications/oecd-online-dispute-resolution-framework_325e6edc-en.html (accessed on 12 February 2026). [CrossRef]
47. Papagiannidis, E.; Mikalef, P.; Conboy, K. Responsible artificial intelligence governance: A review and research framework. *J. Strateg. Inf. Syst.* **2025**, *34*, 101885. [CrossRef]
48. Featherman, M.S.; Pavlou, P.A. Predicting e-services adoption: A perceived risk facets perspective. *Int. J. Hum.-Comput. Stud.* **2003**, *59*, 451–474. [CrossRef]
49. Phamthi, V.A.; Nagy, Á.; Ngo, T.M. The influence of perceived risk on purchase intention in e-commerce: Systematic review and research agenda. *Int. J. Consum. Stud.* **2024**, *48*, e13067. [CrossRef]
50. Akerlof, G.A. The market for “lemons”: Quality uncertainty and the market mechanism. *Q. J. Econ.* **1970**, *84*, 488–500. [CrossRef]
51. Tyler, T.R. *Why People Obey the Law*; Yale University Press: New Haven, CT, USA, 1990.
52. Gawer, A. Bridging differing perspectives on technological platforms: Toward an integrative framework. *Res. Policy* **2014**, *43*, 1239–1249. [CrossRef]
53. Tiwana, A. *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*; Morgan Kaufmann: Waltham, MA, USA, 2014.
54. Diakopoulos, N. Accountability in algorithmic decision making. *Commun. ACM* **2016**, *59*, 56–62. [CrossRef]
55. Huang, M.-H.; Rust, R.T. Artificial intelligence in service. *J. Serv. Res.* **2018**, *21*, 155–172. [CrossRef]
56. Singh, J. Consumer complaint intentions and behavior: Definitional and taxonomical issues. *J. Mark.* **1988**, *52*, 93–107. [CrossRef]
57. Chang, J.; Boyd-Graber, J.; Gerrish, S.; Wang, C.; Blei, D.M. Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*; Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2009; pp. 288–296.
58. Roberts, M.E.; Stewart, B.M.; Tingley, D.; Lucas, C.; Leder-Luis, J.; Gadarian, S.K.; Albertson, B.; Rand, D.G. Structural Topic Models for Open-Ended Survey Responses. *Am. J. Polit. Sci.* **2014**, *58*, 1064–1082. [CrossRef]
59. Trendyol. Türkiye’nin Online Alışveriş Sitesi. Available online: <https://www.trendyol.com/> (accessed on 17 December 2025).
60. Hepsiburada. Hepsiburada.com. Available online: <https://www.hepsiburada.com/> (accessed on 17 December 2025).
61. Amazon.com.tr. Available online: <https://www.amazon.com.tr/> (accessed on 17 December 2025).
62. Google Play. Trendyol—Online Alışveriş. Available online: <https://play.google.com/store/apps/details?hl=tr&id=trendyol.com> (accessed on 17 December 2025).
63. Google Play. Hepsiburada: Online Alışveriş. Available online: <https://play.google.com/store/apps/details?hl=tr&id=com.pozitr.on.hepsiburada> (accessed on 17 December 2025).
64. Google Play. Amazon.com.tr Mobile Alışveriş. Available online: <https://play.google.com/store/apps/details?hl=tr&id=com.amazon.mShop.android.shopping> (accessed on 17 December 2025).
65. Apple App Store. Trendyol: Online Shopping. Available online: <https://apps.apple.com/tr/app/trendyol-online-shopping/id524362642> (accessed on 17 December 2025).
66. Apple App Store. Hepsiburada: Online Shopping. Available online: <https://apps.apple.com/tr/app/hepsiburada-online-shopping/id481035064> (accessed on 17 December 2025).
67. Apple App Store. Amazon.com.tr Mobile Alışveriş. Available online: <https://apps.apple.com/tr/app/amazon-com-tr-mobile-al%C4%B1%C5%9Fveri%C5%9F/id297606951?l=tr> (accessed on 17 December 2025).
68. Franzke, A.S.; Bechmann, A.; Ess, C.M.; Zimmer, M.; Association of Internet Researchers. *Internet Research: Ethical Guidelines 3.0*; AoIR (The International Association of Internet Researchers): Chicago, IL, USA, 2020. Available online: <https://aoir.org/reports/ethics3.pdf> (accessed on 8 February 2026).
69. Hong, L.; Davison, B.D. Empirical Study of Topic Modeling in Twitter. In *Proceedings of the 1st Workshop on Social Media Analytics (SOMA ’10)*, Washington, DC, USA, 25 July 2010; pp. 80–88. [CrossRef]

70. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*; Association for Computational Linguistics: Valencia, Spain, 2017; pp. 427–431.
71. Oflazer, K. Two-level Description of Turkish Morphology. *Lit. Linguist. Comput.* **1994**, *9*, 137–148. [[CrossRef](#)]
72. Akin, M.D.; Akin, A.A. Türk Dilleri İçin Açık Kaynaklı Doğal Dil İşleme Kütüphanesi: Zemberek. *Elektr. Mühendisliği* **2007**, *431*, 38–44.
73. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [[CrossRef](#)]
74. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [[CrossRef](#)]
75. Röder, M.; Both, A.; Hinneburg, A. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*, Shanghai, China, 2–6 February 2015; pp. 399–408. [[CrossRef](#)]
76. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]
77. Killick, R.; Fearnhead, P.; Eckley, I.A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **2012**, *107*, 1590–1598. [[CrossRef](#)]
78. Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol. Bull.* **1993**, *114*, 494–509. [[CrossRef](#)]
79. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
80. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13*, 245–259. [[CrossRef](#)]
81. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]
82. Agresti, A. *An Introduction to Categorical Data Analysis*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2018.
83. Mondal, C.; Giri, B.C. Analyzing strategies in a green e-commerce supply chain with return policy and exchange offer. *Comput. Ind. Eng.* **2022**, *171*, 108492. [[CrossRef](#)]
84. Parasuraman, A.; Zeithaml, V.A.; Malhotra, A. E-S-QUAL: A multiple-item scale for assessing electronic service quality. *J. Serv. Res.* **2005**, *7*, 213–233. [[CrossRef](#)]
85. Wolfinbarger, M.; Gilly, M.C. eTailQ: Dimensionalizing, measuring and predictingetail quality. *J. Retail.* **2003**, *79*, 183–198. [[CrossRef](#)]
86. Zeithaml, V.A.; Parasuraman, A.; Malhotra, A. Service quality delivery through web sites: A critical review of extant knowledge. *J. Acad. Mark. Sci.* **2002**, *30*, 362–375. [[CrossRef](#)]
87. Zhou, Q.; Gümüş, M.; Miao, S. E-commerce order fulfillment problem with limited time window. *Oper. Res.* **2025**, *73*, 6. [[CrossRef](#)]
88. Patel, A.; Ranjan, R.; Kumar, R.K.; Ojha, N.; Patel, A. Online dispute resolution mechanism as an effective tool for resolving cross-border consumer disputes in the era of e-commerce. *Int. J. Law Manag.* **2025**, *ahead of print*. [[CrossRef](#)]
89. Huang, J.W.; Mei, S.; Zhong, W.J. Optimal introduction of price guarantee considering consumer returns in e-commerce platform promotion. *Asia Pac. J. Mark. Logist.* **2026**, *ahead of print*. [[CrossRef](#)]
90. Russo, I.; Masorgo, N.; Gligor, D. Examining the impact of service recovery resilience in the context of product replacement: The roles of perceived procedural and interactional justice. *Int. J. Phys. Distrib. Logist. Manag.* **2022**, *52*, 638–672. [[CrossRef](#)]
91. Tang, C.S. Perspectives in supply chain risk management. *Int. J. Prod. Econ.* **2006**, *103*, 451–488. [[CrossRef](#)]
92. Benaroch, M.; Appari, A. Pricing e-service quality risk in financial services. *Electron. Commer. Res. Appl.* **2011**, *10*, 534–544. [[CrossRef](#)]
93. Tan, E.; Lau, J.L. Behavioural intention to adopt mobile banking among the millennial generation. *Young Consum.* **2016**, *17*, 18–31. [[CrossRef](#)]
94. Owusu, P.; Li, Z.; Mensah, I.A.; Omari-Sasu, A.Y. Consumer response to e-commerce service failure: Leveraging repurchase intentions through strategic recovery policies. *J. Retailing Consum. Serv.* **2025**, *82*, 104137. [[CrossRef](#)]
95. Fang, D.; Zhang, X. The Protective Effect of Digital Financial Inclusion on Agricultural Supply Chain during the COVID-19 Pandemic: Evidence from China. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 3202–3217. [[CrossRef](#)]
96. Lin, H.-H.; Wang, Y.-S.; Chang, L.-K. Consumer responses to online retailer's service recovery after a service failure: A perspective of justice theory. *Manag. Serv. Qual.* **2011**, *21*, 511–534. [[CrossRef](#)]
97. Jafarzadeh, H.; Tafti, M.; Intezari, A.; Sohrabi, B. All's well that ends well: Effective recovery from failures during the delivery phase of e-retailing process. *J. Retail. Consum. Serv.* **2021**, *62*, 102602. [[CrossRef](#)]
98. Zhang, Y.; Huang, J.; Pang, Q. Turning setbacks into smiles: Exploring the role of self-mocking strategies in consumers' recovery satisfaction after e-commerce service failures. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 183. [[CrossRef](#)]
99. Nugroho, A.; Wang, W.-T. Applying justice theory to investigate the effects of consumer complaints and opportunistic intention on brand reputation and consumer repurchase behavior. *J. Electron. Commer. Res.* **2024**, *25*, 18–40.

100. Nam, J.; Yeon, J.; Jung, Y.; Lee, D.; Kim, J.; Lee, J. Platformization of taxi-hailing services: Reducing information asymmetry by examining consumer preferences from the cognitive fit perspective. *Int. J. Hum.-Comput. Interact.* **2025**, *41*, 9930–9946. [CrossRef]
101. Ye, D.D.; Huang, X.M.; Cheng, G.L.; Hossain, M.S. Multi-dimensional contract design for energy-efficient wireless sensing in consumer-centric e-commerce systems. *IEEE Trans. Consum. Electron.* **2024**, *70*, 6883–6891. [CrossRef]
102. Dong, Y.-J.; Zhang, X.; Zhou, J.-Z.; Song, J.-Y. Coordination mechanisms for traceable products in the context of live streaming e-commerce from the perspective of platform, anchor, and consumer. *Asia-Pac. J. Oper. Res.* **2025**, *42*, 2440013. [CrossRef]
103. Yang, Y.X.; Zhao, Q.H.; Yang, Y.F.; Zhou, J.E. Recommended, purchased, disappointed? Exploring the divergent effects of algorithmic product recommendations on sales and satisfaction. *Appl. Econ.* **2025**, ahead of print. [CrossRef]
104. Singh, N.; Misra, R.; Quan, W.; Radic, A.; Lee, S.-M.; Han, H. An analysis of consumer's trusting beliefs towards the use of e-commerce platforms. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 899. [CrossRef]
105. Wang, S.S.; Peng, K.-L.; Huang, Z.L.; Ma, L.J. AI-generated videos: Influencing trustworthiness, awe, and behavioral intention in space tourism e-commerce. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 307. [CrossRef]
106. Yu, J.T. Examining the interplay of affect and cognition in online information disclosure in e-commerce: Insights from two empirical studies. *Electron. Commer. Res. Appl.* **2025**, *73*, 101531. [CrossRef]
107. Le, T.T.; Phan, D.N.; Ngo, T.T.T.; Le, N.T. Website quality's impact on Gen Z's eWOM behavior and online purchase intentions: The mediating role of trust in online shopping. *Asia Pac. J. Mark. Logist.* **2025**, ahead of print. [CrossRef]
108. Chen, J.Y.; Du, P.F. Under the dark side of online trust: How and when livestreamers' online expressive coping strategy impacts the livestreaming e-commerce failure recovery process. *Internet Res.* **2025**, ahead of print. [CrossRef]
109. Basili, M.; Rossi, M.A. Platform-mediated reputation systems in the sharing economy and incentives to provide service quality: The case of ridesharing services. *Electron. Commer. Res. Appl.* **2020**, *39*, 100835. [CrossRef]
110. Dann, D.; Teubner, T.; Adam, M.T.P.; Weinhardt, C. Where the host is part of the deal: Social and economic value in the platform economy. *Electron. Commer. Res. Appl.* **2020**, *40*, 100923. [CrossRef]
111. Luo, Z.; Li, D.; Wang, G.; Wang, S.; Cheng, M.; Li, T.; Lei, Z. Service orchestration of customized production with automated workflow systems. *Cluster Comput.* **2025**, *28*, 602. [CrossRef]
112. Cheng, X.; Nam, I. Rethinking picky shoppers and store reputation: Effective online service recovery strategies for products with minor defects. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 259. [CrossRef]
113. Alharbi, K.; Alkhalifah, A. Examining the role of trust and privacy effects through online reviews in social commerce using an integrated model and hybrid approach analysis. *IEEE Trans. Eng. Manag.* **2024**, *71*, 10943–10965. [CrossRef]
114. Alkhalifah, A. Exploring trust formation and antecedents in social commerce. *Front. Psychol.* **2022**, *12*, 789863. [CrossRef]
115. Anderson, R.; Moore, T. The economics of information security. *Science* **2006**, *314*, 610–613. [CrossRef]
116. Acquisti, A.; Brandimarte, L.; Loewenstein, G. Privacy and human behavior in the age of information. *Science* **2015**, *347*, 509–514. [CrossRef] [PubMed]
117. Malhotra, N.K.; Kim, S.S.; Agarwal, J. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Inf. Syst. Res.* **2004**, *15*, 336–355. [CrossRef]
118. Meng, H.; Xiao, Q.; Na, Y.P. Warmhearted cues: A study of the impact of social mindfulness on trust repair by intelligent customer service in service recovery. *Int. J. Hosp. Manag.* **2025**, *128*, 104131. [CrossRef]
119. Shao, Z.; Zhang, L.; Li, X.T.; Zhang, R. Understanding the role of justice perceptions in promoting trust and behavioral intention towards ride-sharing. *Electron. Commer. Res. Appl.* **2022**, *51*, 101119. [CrossRef]
120. Gray, C.M.; Kou, Y.; Battles, B.; Hoggatt, J.; Toombs, A.L. The dark (patterns) side of UX design. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Montreal, QC, Canada, 21–26 April 2018. [CrossRef]
121. Rese, A.; Witthohn, L. Recovering customer satisfaction after a chatbot service failure—The effect of gender. *J. Retail. Consum. Serv.* **2025**, *84*, 104257. [CrossRef]
122. Wang, S.; Fatima, N.; Shahbaz, M.; Asif, M. Building user trust in AI chatbots for customer service through human-like cues and perceived reliability. *Sci. Rep.* **2026**, *16*, 7860. [CrossRef] [PubMed]
123. International Consumer Protection and Enforcement Network (ICPEN). *ICPEN Dark Patterns in Subscription Services Sweep: Public Report 2024*; ICPEN: Bogota, CO, USA, 2024. Available online: <https://www.icpen.org/sites/default/files/2024-07/Public%20Report%20ICPEN%20Dark%20Patterns%20Sweep.pdf> (accessed on 21 February 2026).
124. Brennecke, M. Regulating Dark Patterns. *arXiv* **2023**, arXiv:2310.00340. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.