

Automatic Mining of Customer Pain Points from Open Reviews: The “Review to Pain Matrix” Workflow

Konstantin Zhuchkov

Private Consulting Practice Result ED, New York, NY, USA
Email: kzhuchkovusa@gmail.com

How to cite this paper: Zhuchkov, K. (2025). Automatic Mining of Customer Pain Points from Open Reviews: The “Review to Pain Matrix” Workflow. *American Journal of Industrial and Business Management*, 15, 1215-1228.
<https://doi.org/10.4236/ajibm.2025.159060>

Received: August 1, 2025

Accepted: September 2, 2025

Published: September 5, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper examines the automatic extraction of customer pain points from open reviews using the “Review to Pain Matrix” pipeline. The objective of this study is to develop a systematic approach for extracting customer pains from unstructured reviews, ensuring the reproducibility, transparency, and scalability of the analysis, while preserving the contextual usage and emotional valence of the statements. The relevance of this work is driven by the growing volume of user reviews and the need for product teams to rapidly focus on real customer issues without manual processing, which is constrained by subjectivity and the labor-intensive handling of large datasets. The novelty of the proposed “Review to Pain Matrix” workflow lies in the integration of product discovery and delivery stages within a single automated pipeline: from text cleaning and normalization, deduplication, thematization and aspect-level analysis, through to the generation of a canonical table of pain points with measurable attributes of frequency, intensity, impact on key metrics and effort required for resolution. Of particular importance is the ability to cyclically reprocess successive batches of reviews for prompt evaluation of product release effects and for updating marketing and service scripts. The outcome is a Pain Matrix: a machine- and human-readable matrix of verifiable pain-point formulations linked to audience segments and customer-journey stages, with priorities computed as a weighted combination of frequency, intensity, impact, and effort. This format enables the transformation of fragmented user feedback into a governed decision-making system, accelerates the conversion of insights into backlog items and experiments, and optimizes communication among product managers, designers, engineers, and front-office teams. This paper will be of interest to researchers and practitioners in product management, UX research, and data analytics.

Keywords

Automatic Review Mining, Customer Pain Points, Pain Matrix, Product Discovery, CJM, Semantic Analysis, Prioritization

1. Introduction

Open reviews represent customers' natural utterances, in which specific difficulties, usage context, and emotional tone are recorded, without the constraints of a survey or interview script. Systematic extraction of pains from such sources enables the formulation of testable product hypotheses, the refinement of problem statements in customers' own words, and the preparation of guides for Customer Development. Where manual analysis quickly reaches limits of scale and subjectivity, an automated pipeline ensures reproducibility, transparent normalization rules, and comparability of analysis waves over time, which is crucial for tracking the effects of changes.

The "Review to Pain Matrix" process occupies the intersection of product discovery and delivery. In discovery, it functions as a continuous problem radar, linking reviews to customer-journey stages and segments, helping to highlight root causes rather than symptoms. The output of discovery is an ordered list of pains with substantiated priorities, which then feeds into delivery as input for backlog formation, solution design, and experiment planning. By supporting cyclic operation, each new review batch is rerun through the pipeline, providing rapid feedback on the impact of releases on pain points, and guiding marketing, sales, and support regarding which messages and scripts require updates.

The result is the Pain Matrix, a machine-readable and human-readable table of canonical pain formulations annotated with frequency, intensity, impact on product and business metrics, effort required to resolve, and mapping to segments and CJM stages. Such a matrix becomes a common language for product managers, researchers, designers, engineers, marketers, and front-office teams, reducing ambiguity in terminology and prioritization. Based on this, it is easier to construct three prototypical segment avatars—typical user, motivated user, and skeptic—to prepare interviews, test hypotheses about value and barriers, and link each pain point to a task, expected outcome, and success criteria. Altogether, this transforms disparate review noise into a governed decision-making system aligned with the product roadmap and release cadence.

2. Materials and Methodology

The foundational literature comprised regulatory acts and guidance on personal-data protection, namely EU Regulation 2016/679 (GDPR) (European Commission, 2016) and the California Consumer Privacy Act (CCPA) (CLI, 2018); the technical standard for robot-exclusion protocols RFC 9309 (Koster et al., 2022) and case law in *hiQ Labs vs LinkedIn* before the U.S. Ninth Circuit (Wallace et al., 2022) to sub-

stantiate legal and ethical aspects of data collection; research on duplicate-detection based on shingles and fingerprints—MinHash and SimHash (Manku et al., 2007)—for selecting algorithms to eliminate duplicate reviews; and an industry report on the sentiment-analysis tools market, demonstrating market trends and growth forecasts (BRI, 2025).

A content analysis was conducted to study these works, following these steps systematically: collection of full texts of documents and articles; identification of key thematic blocks (legal frameworks, data collection and normalization methods, deduplication algorithms, semantic-analysis approaches); development of a coding scheme; independent coding by two researchers; reconciliation of codes and consolidation into a final set of categories. This approach enabled the identification of common principles and differences in methods, the determination of theme frequencies, and the tracing of technology and regulatory evolution from 2007 to 2025.

Subsequently, a comparative analysis of the identified categories was performed, including the juxtaposition of GDPR and CCPA in terms of de-identification and data-minimization requirements; comparison of technical deduplication methods concerning accuracy and performance; evaluation of the role of robots.txt protocols and legal precedents in the construction of ethical data-collection pipelines; and comparison of market approaches to sentiment-analysis software regarding functional capabilities and growth rates.

3. Results and Discussion

Within this study, the Pain Matrix is understood as a normalized set of customer pains, presented in both machine-readable and human-readable forms, where each pain is succinctly and testably formulated, linked to an audience segment and a customer-journey stage, and enriched with measurable attributes for prioritization. This format is advantageous because it translates heterogeneous reviews into a unified decision-making language, simplifies comparison across analysis waves over time, provides traceability from the original citation to the canonical formulation and back, and serves as a bridge between research and development, thereby reducing subjectivity and accelerating the translation of insights into experiments and releases.

The evaluation axes are selected to simultaneously reflect the customer's voice and the team's operational constraints. Frequency records the proportion of reviews in which a pain appears within a given wave or period, helping to distinguish systemic issues from isolated cases. Pain intensity reflects the severity of a negative experience and its behavioral consequences, such as app abandonment or payment refusal. It is convenient to assess on a calibrated scale with a limited number of levels and to calibrate against examples. Impact on metrics links pain to product or business dynamics, such as activation, conversion, retention, or revenue, enabling hypothesis testing both qualitatively and quantitatively (Malik & Bilal, 2024). The effort to resolve describes the labor intensity of addressing the

pain, including engineering complexity, support costs, and risks; this parameter is crucial for balancing quick wins against long-term investments. Two additional dimensions render the matrix applicable for segmented work: audience segment (e.g., SMB teams, enterprise administrators, individual users) and customer-journey stage (e.g., advertising, landing, onboarding, first value, regular use, payment, support). For practical use, it is convenient to compute an overall priority as a weighted combination of frequency, intensity, impact, and effort, with coefficients chosen according to strategy and release-cycle constraints, and with the formula kept transparent for discussion within cross-functional teams.

The canonical pain formulation must be concise, specific, and solution-neutral, while preserving customer lexicon and permitting linkage to multiple sources. Typical examples in a matrix-friendly format include: “app crash on startup on certain devices during onboarding”; “non-obvious team-permission settings causing a new administrator to fail to grant access on first attempt”; “service invoice issued without itemization causing accounting to reject payment”; “card payment declined without clear reason at the 3DS step leading the user to abandon the attempt”; “video tutorials interrupted by unstable connection preventing achievement of first value”. These formulations are easily matched to source citations, aggregated into canonical entries when synonyms are encountered, and placed in the Pain Matrix for subsequent scoring, segmentation, and inclusion in experimental plans. A summary of pain-point prioritization is presented in **Figure 1**.

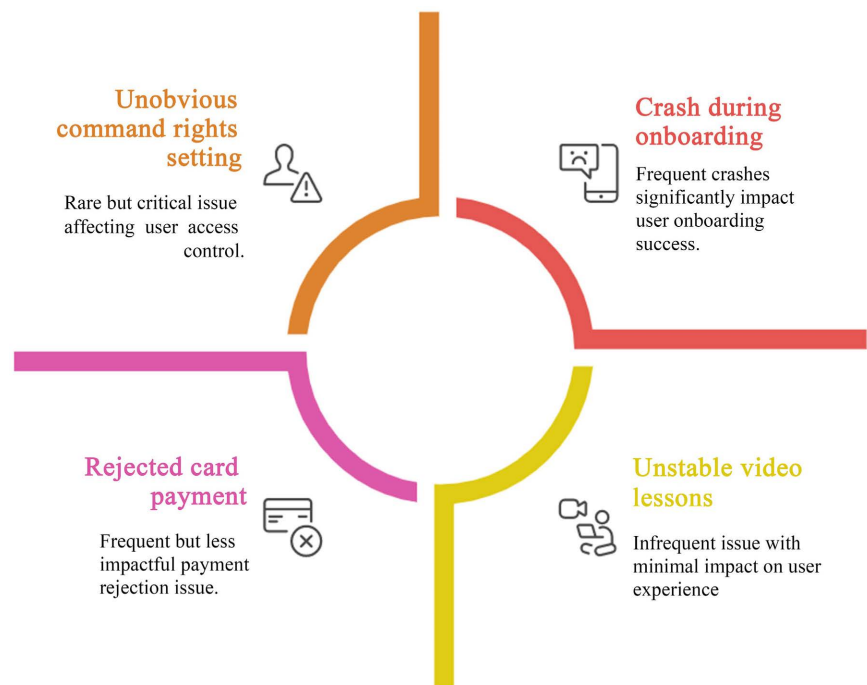


Figure 1. Prioritization of customer pain points (compiled by author).

Sources of open reviews should be collected across multiple classes of platforms: application marketplaces and SaaS directories, social networks and specialized fo-

forums, public feedback forms and issue trackers, as well as reviews of competitors' products. It is advisable to agree in advance on the sampling framework—specifically, the time window, list of channels, and minimum text length—to minimize source bias and retain the original text without editing. Before collection, one must verify the platforms' terms of use and automated-access policies, observe robots.txt rules as an industry standard for coordination with the resource owner, and recognize that these rules do not constitute legal authorization (Koster et al., 2022). Jurisprudence concerning publicly available data is uneven, so teams typically combine technical measures that respect the site with legal risk assessments. The HiQ Labs v. LinkedIn line of cases in the U.S. Ninth Circuit is illustrative (Wallace et al., 2022). These considerations do not preclude the study of open reviews, but they require careful methods for collecting and storing primary materials.

A minimal set of metadata renders the review corpus reproducible and suitable for aggregation. For each record, capture a unique identifier, the source with a link, the date and time of the event in UTC, the language of the original text, the product version and platform (e.g. iOS, Android, Web), as well as the presumed segment, if inferable from context, and the customer-journey stage for subsequent CJM mapping. It is recommended to store an immutable snapshot of the raw field and a normalized version for analysis, maintain a unified reference list of platforms and versions, and apply deterministic rules for identifier generation to facilitate deduplication and traceability from canonical pain to citation.

Legal and ethical aspects define the boundaries of permissible data operations. Under GDPR, personal data encompasses any information relating to an identifiable natural person, and pseudonymization is treated as processing that reduces risk but does not exempt data from the regulation; processing must rely on a lawful basis, such as legitimate interests, and be accompanied by protective and minimization measures (European Commission, 2016). Under the CCPA, personal information is defined broadly, and de-identification requires that re-identification be reasonably impossible, given the organizational and technical safeguards in place (CLI, 2018). In practice, this means that user names, email addresses, telephone numbers, and other identifiers should either not be collected or be transformed into stable hashes for internal linking. Only data necessary for source verification and reproducibility should be retained. Additionally, procedures for data-subject deletion requests must be in place.

Data preparation begins with cleaning and normalization. Typical steps include removal of overt spam and off-topic content according to rules, normalization of case and whitespace characters, unification of encodings and markup, language detection with retention of the original text, and, if needed, machine translation into the corpus's base language for unified analysis. It is essential to maintain the connection to the original and its language, so that, in the event of interpretive discrepancies, one can refer back to the primary source. At this stage, it is also helpful to annotate the filters' confidence levels, enabling rapid selection of records for manual quality review later.

Deduplication removes duplicate and near-duplicate messages, reducing the overestimation of specific formulations’ frequency. A practical approach relies on shingles and probabilistic fingerprints, such as MinHash and SimHash, which preserve textual similarity under minor edits and enable the rapid identification of clusters of similar reviews at scale (Manku et al., 2007). In web-scale environments, these techniques have been utilized for document similarity assessment and duplicate elimination during indexing, making them well-suited for the Review to Pain Matrix pipeline, where it is crucial to retain formulation variability while eliminating mechanical duplication.

Tag-based enrichment links raw text to usage context and future prioritization. At this step, each record is labeled with tariff or license type, region, and time zone, acquisition channel, and CJM stage (e.g., advertising, landing, onboarding, first value, regular use, payment, support). Tariff and platform tags are most often extracted from the source or client fields. The CJM stage is assigned based on explicit textual markers or rules, and if needed, with model assistance. The result is recorded as a hypothesis with a confidence value for subsequent validation. To uphold minimization principles, sensitive attributes (e.g., gender, ethnicity, religion) should not be inferred unless necessary for analysis; any derived attributes must be reproducible by rule rather than manual interpretation.

The Review to Pain Matrix workflow begins with goal setting and hypothesis formulation: the team defines research questions that link open reviews to measurable outcomes, as shown in **Figure 2**.

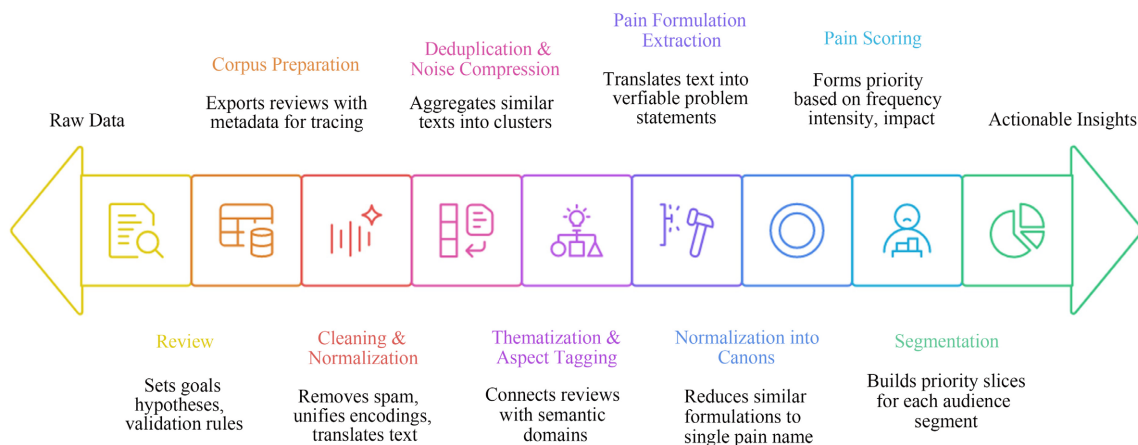


Figure 2. Pain matrix workflow stages (compiled by author).

At this stage, anticipated audience segments and customer-journey stages are recorded, and operational definitions of pain and validation rules are established—such as which textual markers indicate action abandonment, how to distinguish symptoms from causes, and which product metrics will be used for subsequent verification. Precise framing enhances comparability across analysis waves, facilitates interpretation, and reduces the likelihood of confirmation bias.

Next, the review corpus is assembled and exported in a tabular format, including immutable raw text, the source link, and minimal metadata. The schema includes event date and time in a unified time zone, original text language, product version, platform, and, where inferable, the presumed segment and CJM stage. This layer serves as the unit of record and traceability: each row can later be mapped to a canonical pain formulation while preserving the backward link to the source.

Cleaning and normalization convert text to an analyzable form while retaining the original reference. Typical operations include rule-based spam removal, case and whitespace normalization, encoding unification, language detection with confidence logging, and, where necessary, machine translation into the corpus's base language. It is crucial to separate automated transformations from manual edits, log both processes, and store raw and normalized versions so that, in contentious cases, one can return to the client's original phrasing.

After normalization, deduplication, and noise reduction are performed to eliminate duplicate and near-duplicate messages without sacrificing formulation diversity. Similar texts are clustered, with all original identifiers preserved as links; cluster counts are used for accurate frequency estimation. Frequency smoothing across clusters prevents overestimation of individual formulations due to technical copying and pasting, or cross-posting, thereby improving prioritization stability between waves (Li et al., 2025).

Thematization and aspect annotation associate reviews with semantic domains and quality-analysis units, utilizing an Aspect-Based Sentiment Analysis model to extract, for each aspect (e.g., loading speed, payment, support), the target entity and sentiment polarity (negative, neutral, positive) along with confidence.

Currently, pre-trained great language models might be the base for carrying out this stage. The transformation of such a model toward the particular job is accomplished using organized prompts that direct the model to get all the needed aspects and their sentiment from the text.

As concerns language coverage, since the models are naturally multilingual, this enables the treatment of reviews in several languages. Validation of the model's performance goes hand-in-hand with the "human-in-the-loop" principle being executed in practice as described above under quality control. Manual verification on a representative sample from each wave of reviews checks the quality of aspect and sentiment extraction (and thereby the quality of the model for this particular task) and accuracy.

Aspect tags are supplemented with CJM-stage mappings—advertising, landing, onboarding, first value, regular use, payment, and support—enabling the identification of bottlenecks in transitions and hypothesis generation about causal relationships (e.g., the impact of payment errors on retention). Meanwhile, the sentiment-analysis software market was valued at USD 2.1 billion in 2024 and is projected to reach USD 6.85 billion by 2033, at a compound annual growth rate of 14.1%, as shown in **Figure 3** (BRI, 2025).

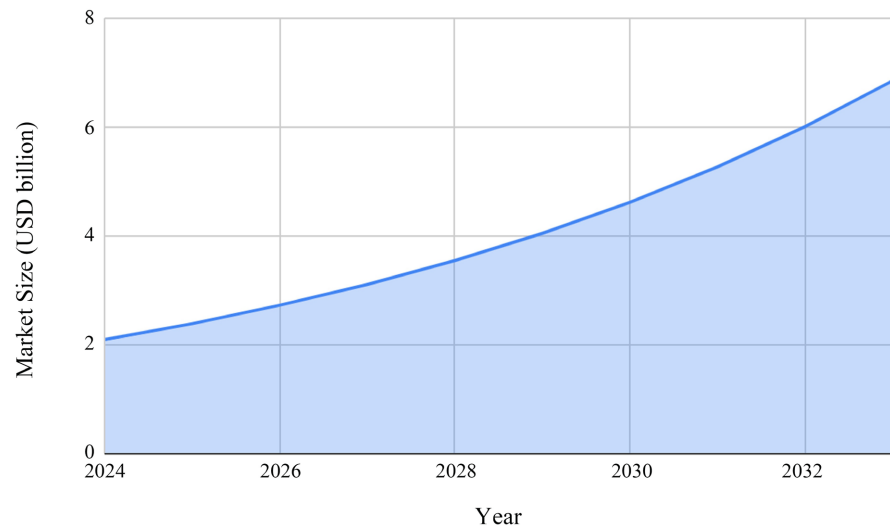


Figure 3. The global sentiment analysis software market size (BRI, 2025).

Extraction of pain-point formulations converts free text into concise, testable statements that describe the customer’s problem without imposing a solution. Practically, these take the form of expressions specifying who, where, and under what conditions cannot perform the target action, or receives a result that hinders value realization, with explicit linkage to context. Each formulation retains a reference to the citation and its associated aspects, ensuring reproducibility and auditability, while also simplifying subsequent segmentation and quantitative validation.

Normalization into canonicals consolidates synonymous or similar formulations under a single point of reference, while maintaining a reference dictionary version and change log. A canonical is phrased neutrally, succinctly, in the customer’s language, allows linkage to multiple CJM stages and segments, and includes a list of typical triggers and exclusions to reduce the risk of overgeneralization. Consolidation is performed by established rules and under the guidance of experts. For further quality control, some mappings are reviewed by an independent evaluator.

Pain scoring establishes a transparent priority for decision-making. Its core components include occurrence frequency, pain intensity as measured by emotional and behavioral expressiveness, impact on key product metrics—such as conversion, retention, and revenue—and effort to resolve, including engineering complexity, timeline, and operational risks. The final priority is calculated as a weighted combination of these components.

The determination of weight coefficients results from a cross-functional team effort, as stated in the article. In discussion with the team, it defines the current strategic focus—for example, between growth, retention, or reduction of technical debt—and how significance is allocated among the four factors. This method ensures that priorities are keyed to the current business objectives.

A sensitivity check is also in this discussion. They see how the ranking of major

issues would change if the weight of one component were to be changed, which speaks to the robustness of the final priority list. This method of open discussion is a practical implementation of principles underlying well-known prioritization frameworks and ensures transparency and consensus in decision-making (Šmíd & Přibán, 2023).

Weights are calibrated cross-functionally based on strategy and release-cycle constraints, and the priority sensitivity to weight selection is tested on a control sample to ensure the ranking is robust to reasonable variations.

Segmentation by audience and CJM stage translates the Pain Matrix from a general view into an applied one: for each relevant segment, a tailored priority slice is constructed, enabling the differentiation of conflicts and alignments. For example, a pain point critical for new users during onboarding may carry low weight for long-term customers, and vice versa. Consequently, solutions are planned in bundles, aligned by impact and effort, with consideration of the effects on adjacent segments and channels.

Visualization of the Pain Matrix combines tabular presentation and a quadrant map. The table contains the canonical pain statement, frequency, intensity, impact, effort, segment, and stage, as well as links to sources and those responsible for verification and implementation. The quadrant map, with axes of frequency and intensity, marker size representing impact, and, if desired, hue representing effort, provides a compact overview of the problem landscape, enabling the team to quickly select a set of candidate tasks for experimentation and development. At the same time, all decisions remain traceable to table rows.

Embedding the Pain Matrix into the CustDev cycle closes the research loop, as guides for interviews are generated based on top priorities. This testing of assumptions includes examining usage context, barriers, willingness to change behavior and pay, and perceived value of the proposed improvement. Simultaneously, three prototypical segment avatars—typical user, motivated user, and skeptic—are created from the same pain points, facilitating the design of questions and the interpretation of responses. Interview results are fed back into the matrix, where weights and formulations are updated, and hypotheses and success criteria are refined. Afterward, priorities are converted into backlog tasks and experiment plans. Subsequent waves of reviews provide quantitative feedback on the impact of releases on the pain landscape.

The storage architecture must support the entire Review to Pain Matrix cycle, specifically accumulating raw materials, ensuring the reproducibility of transformations, and yielding collaborative artifacts. Practically, this entails separate layers for narrative wave reports, a tabular master database, a knowledge base with hypotheses and statuses, and visual maps for planning discussions. Google Docs is used to create concise reports for each wave, documenting the collection window, sources, methodological decisions, and priority summaries. The document includes links to rows in the master table and to the sources. Google Sheets or Excel serves as the systemic source of truth for the Pain Matrix, housing normal-

ized fields and computed metrics, as well as keys linking to tasks and experiments. Notion functions as a long-term knowledge base, where for each segment and hypothesis, there are cards with context, assumptions, current results, and next steps, and where interview notes and avatars are conveniently stored. Miro or FigJam are employed for problem and CJM maps, enabling the team to visually discuss quadrants and solutions, with each board element linked to a Pain Matrix row, thereby preserving traceability.

The master file's table schema is straightforward and sufficient for daily work. The Reviews sheet holds basic fields: unique identifier, source link, event date and time in UTC, local date if needed, language, numeric rating if present in the source, product version, platform, channel, raw text and normalized text in separate columns, hypothesis on segment and CJM stage, and processing flags such as spam flag, language-detection confidence level and machine-translation flag. The Aspects sheet records aspect and sentiment annotations linked to review_id; for each aspect-review pair, sentiment polarity, confidence value, and a short citation are stored, and optionally, the fragment's position in the text for re-review. The PainsCanon sheet contains the canonical dictionary, including pain_id, concise neutral pain statement in customer words, description, typical triggers and exclusions, CJM-stage and segment mappings, links to multiple representative review_ids, and a dictionary version with a change log. The Scores sheet stores scoring results, listing the occurrence frequency in the current wave, pain-intensity score, impact score on key product and business metrics, effort-to-resolve score, computed priority, weight-version, and timestamp of computation for each pain_id. This allows for a transparent explanation of the ranking. The RoadmapLinks sheet associates pain IDs with solution ideas and experiments, containing a brief hypothesis description, status, owner, expected metric, and target change, as well as a link to the tracker task and an approximate timeline. This linkage makes the Pain Matrix a working tool rather than an archive.

Automation is built from simple to complex. In a no-code variant, uploading a CSV to Sheets, using formulas for frequency counts and normalization, and utilizing Apps Script scenarios to call the model via API, record aspect annotations, normalize formulations into canonicals, and calculate priority using the specified formula with weight versioning suffice. Zapier or Make would work well for planned source and table updates; they can periodically pull new reviews, initiate processing, and notify everyone when the wave is complete. A better stack includes embeddings for semantic similarity search, clustering to group nearly identical formulations, quick identification of similar reviews using cosine similarity and approximate nearest neighbor indexing, as well as a topic drift watch module that checks for new cluster growth or changes in aspect spreads on a weekly basis. For prioritized visualization, a bubble chart with axes of frequency and intensity, marker size indicating impact, and hue representing effort to resolve, provides a concentrated view of the pain landscape for a specific segment and wave.

Quality control and validation are built around the human-in-the-loop princi-

ple. In each wave, a representative sample of reviews is selected for manual verification of cleaning, normalization, and pain-point extraction. All discrepancies are recorded, and adjustments are made to the rules or model as needed. Inter-annotator agreement is monitored on a portion of the corpus where two or more experts independently annotate aspects and canonicals according to the guidelines; the results are compared, discrepancies discussed, and integrated into updated guidelines and examples, thereby reducing interpretive drift. A separate loop is devoted to drift monitoring: distributions of sources, regions, platforms, and channels are compared between waves; emergence of new topics is tracked; the canonical dictionary is periodically reviewed by version; obsolete canonicals are flagged; new ones are introduced with examples and source links. Such discipline preserves the reproducibility and explainability of results, maintaining the Pain Matrix's utility as the product and its audience evolve.

The application of findings closes the loop regarding operational decisions. In marketing, headlines, subheadlines, and offers are rewritten in customer language to address segment-priority pains; teams record changes in the hypothesis card and correlate them with subsequent response shifts. In the product, each `pain_id` is linked to a backlog task, specifying the expected metric impact and measurement method. After release, the review wave is rerun, providing quantitative feedback on the pain field. In sales and support, responses and scripts are prepared that can acknowledge the pain and tactfully address it, either temporarily via workarounds or in the context of the roadmap. Current formulations are stored in the knowledge base and synchronized with the matrix. As a result, the Pain Matrix becomes not a static showcase but a live decision-making system, connecting open reviews with hypotheses, experiment plans, and impact metrics.

A case study was implemented to validate this approach. The comments of users left open to the public were analyzed to identify the main pain points for the audience and test the hypothesis: Will it be true that new students are motivated not by acquiring new skills but by the real opportunity to change their lives and build a new career?

The process started with the collection of data. A huge amount of comments and posts have been collected from ten popular YouTube channels about how to get into the IT industry, five thematic Telegram groups, and relevant forums, at the author's methodological materials' recommendation. The initial dataset was then made more valuable through performing data cleaning and deduplication.

In the next stage, the audience was decomposed by these frameworks. It is this segment that happens to be not only large but acute—office workers between 27 and 45 years of age suffering from professional burnout and career stagnation. Their statement analysis brought out ten major queries, of which three were the most frequent, as well as emotionally loaded—investment of so much time and money without any guarantee for a job later on; uncertainty regarding which programming language to begin with; and anxiety about whether one can indeed enter into IT after turning 30 or even 40.

This became the Pain Matrix, where the principal canonical pain was formulated as “fear of unemployment after course completion because of either age or lack of experience.” Relevant conversations attested to a high frequency for this pain, and, indeed, it was rated high in intensity. Validation came through the use of such affective language as “anxiety” and “dead-end”, which marked marketing that directly impacts conversion rates. It takes medium effort to resolve through adjustment via marketing and add a career support module.

A user avatar was created and used in the simulated dialogue for final validation of the hypothesis. The interaction confirmed that for this segment, the proper deliverable is not a certificate of completion but rather employability assurance and a structured career transition plan. Based on these conclusions, marketing offers were redesigned to emphasize a “career track” and success stories of mature graduates, which subsequently led to a measurable increase in conversion for the next student cohort.

4. Limitations

Though it is of practical value, the methodology has several limitations that have to be considered. The first is that the choice of platforms from which to collect reviews can introduce sampling bias since users on different channels may have different motivations and experiences. As a partial offset to this effect, the methodology, as articulated in this article, implements mitigation through the development of a clear sampling framework and then segmentation of the results by source to identify channel-specific patterns.

Second, working with multilingual corpora, the use of machine translation might distort the original meaning or sentiment of a review. As a way of reducing this risk, it is indicated that both texts, the original and the translated, are preserved so that they can always be referred back to the source for clarification in cases where ambiguity arises, more especially in issues rated high on the priority list.

Another risk is annotation drift, i.e., the change of interpretation of canonical formulations over time and input from different analysts. To check for this, a versioned dictionary of canons is kept, and regular inter-annotator agreement sessions are instituted to check how far the analysis remains consistent and reproducible across different waves of reviews.

5. Conclusion

In the present study we introduce a systematic approach for automatically extracting customer “pain points” from open reviews via the “Review to Pain Matrix” pipeline, which integrates the stages of product discovery and delivery, ensuring reproducibility, transparency and scalability of the analysis, while preserving linkage to the usage context and the emotional valence of the statements. In the Introduction, we justify the need to move from manual review analysis, which is limited by volume and subjectivity, to an automated pipeline with clear normalization rules, enabling the tracking of pain-point dynamics over time and the eval-

uation of product changes' impact on user experience. In contrast, the Materials and Methodology section describes in detail all data-processing stages—from collection and normalization through deduplication, thematization, aspect analysis, and construction of canonical pain formulations.

As a result of the pipeline, a Pain Matrix is produced: a machine-readable and human-readable table in which each pain is represented by a short, testable statement, linked to an audience segment and a customer-journey stage, and annotated with measurable attributes of frequency, intensity, impact on key metrics, and effort required for resolution. This format translates disparate reviews into a unified decision-making language, simplifies comparison across analysis waves, ensures traceability from original citations to canons, and serves as a nexus between research and development, reducing subjectivity in prioritization and accelerating the transformation of insights into experimental tasks and releases.

Particular attention is paid to the integral priority, calculated as a weighted combination of frequency, intensity, impact, and effort, whose coefficients are tailored to the product strategy and release-cycle constraints, while maintaining complete transparency of the formula and its calibration within a cross-functional group. Segmentation by audience and CJM stages enables the adaptation of the Pain Matrix to meet the needs of different user groups, thereby enhancing the accuracy and relevance of proposed solutions.

The practical value of the approach is confirmed by its ability to close the CustDev loop: based on top pain priorities, interview guides, and segment avatars are developed, results are fed back into the matrix to update weights and formulations, and comparison with a new wave of reviews provides quantitative feedback on release effectiveness. An integrated tooling ecosystem, including Google Sheets, Notion, Miro, and other platforms, supports collaborative work, traceability, and preservation of all artifacts at each pipeline stage.

Thus, the proposed “Review to Pain Matrix” workflow transforms a chaotic volume of open reviews into a governed decision-making system aligned with the product roadmap and release rhythm, reducing risks of bias and subjectivity, accelerating the delivery of insights to development and front-office teams, and enabling marketing, sales, and support to update messages and scripts promptly. Further research may focus on expanding algorithmic clustering and semantic-search capabilities, adapting the pipeline to new languages and cultures, and integrating streaming review processing to provide even more immediate feedback on changes in user experience.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- BRI (2025). *Sentiment Analysis Software Market Report and Global*. BRI.
<https://www.businessresearchinsights.com/market-reports/sentiment-analysis-software-market-102912>

- CLI (2018). *California Consumer Privacy Act of 2018*. CLI.
https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&law-Code=CIV&part=4.&title=1.81.5
- European Commission (2016). *EU 2016/679*. European Commission.
<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- Koster, M., Illyes, G., Zeller, H., & Sassman, L. (2022). *RFC 9309: Robots Exclusion Protocol*. IETF Data Tracker. <https://datatracker.ietf.org/doc/html/rfc9309>
- Li, S., Wang, Z., Zhao, Z., Zhang, Y., & Li, P. (2025). Exploring Model Editing for LLM-Based Aspect-Based Sentiment Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39, 24467-24475. <https://doi.org/10.1609/aaai.v39i23.34625>
- Malik, N., & Bilal, M. (2024). Natural Language Processing for Analyzing Online Customer Reviews: A Survey, Taxonomy, and Open Research Challenges. *PeerJ Computer Science*, 10, e2203. <https://doi.org/10.7717/peerj-cs.2203>
- Manku, G. S., Jain, A., & Das Sarma, A. (2007). Detecting Near-Duplicates for Web Crawling. In *Proceedings of the 16th international conference on World Wide Web* (pp. 141-150). ACM. <https://doi.org/10.1145/1242572.1242592>
- Šmíd, J., & Přibáň, P. (2023). Prompt-Based Approach for Czech Sentiment Analysis. In *Proceedings of the Conference Recent Advances in Natural Language Processing—Large Language Models for Natural Language Processings* (pp. 1110-1120). INCOMA Ltd. https://doi.org/10.26615/978-954-452-092-2_118
- Wallace, J., Berzon, M., & Berg, T. (2022). *United States Court of Appeals for the Ninth Circuit*. U.S. Courts.
<https://cdn.ca9.uscourts.gov/datastore/opinions/2022/04/18/17-16783.pdf>