UHASSELT
KNOWLEDGE IN ACTION

Faculty of Business Economics

Master of Management

Master's thesis

How do user-friendly explanations influence customer reactions to algorithmic failures ?

B Samvav

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Data Science

SUPERVISOR :
Prof. dr. Claire DEVENTER

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2024
2025

Faculty of Business Economics

Master of Management

Master's thesis

How do user-friendly explanations influence customer reactions to algorithmic failures ?

B Samvav

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Data Science

SUPERVISOR :

Prof. dr. Claire DEVENTER

PREFACE:

This dissertation marks the culmination of my academic journey in the Master of Management program with a specialization in Data Science at Hasselt University. Undertaking this research has been both intellectually rewarding and personally transformative. I, B Samvav, would like to express my deepest gratitude to my supervisor Prof. dr. Claire Deventer, for her insightful guidance, critical feedback and steady encouragement throughout the development of this thesis. Her expertise and patience have helped shape this work in meaningful ways and I am sincerely thankful for her mentorship. I am also indebted to my professors, peers and the academic staff at Hasselt University who have contributed to my learning and growth over the past years. Special thanks go to the individuals who participated in the study and made this research possible. Finally, I wish to extend heartfelt appreciation to my family and close friends for their unwavering support, understanding and motivation. Their belief in me has been a constant source of strength during this challenging yet fulfilling process.

SUMMARY:

We have all probably seen it: an app pushes some aggressive "get thin fast" content to someone already suffering from eating disorder. Such mismatched recommendations are not rare. In fact, prior research shows that "performance differs substantially for distinct subgroups," often harming marginalized users the most (Buolamwini & Gebru, 2018, p. 11). While fixing the tech side of algorithms matters (retraining models, removing bias, etc.), numerous scholars argue that technicality alone does not suffice and users must also be supported through design and communication (Liao et al., 2021; Shin, 2021). Yet the human communication element is often ignored (Liao et al., 2021).

This research digs into how both cold technical and warm empathetic explanations shape people's reactions to algorithmic failure. Three core hypotheses were proposed to explore this: H1 predicted that user-friendly explanations (empathetic or technical) would increase customer satisfaction compared to no explanation. H2 suggested that empathetic explanations would generate higher perceived justice than technical or no explanations. H3 hypothesized that perceived justice would positively predict customer satisfaction. To find out, an online experiment was conducted. Participants saw a simulated offensive content recommendation, followed by different types of explanations (empathetic, technical, both, or neither). Their reactions were measured using 7-point Likert Customer Satisfaction, Interactional Justice and Credibility scales.

The results are telling that not everyone took the failure equally hard, but the kind of explanation offered afterward made a big difference. Participants responded more positively (feeling more satisfied and fairly treated) when the algorithm's failure was followed by some kind of explanation. The group that received both empathetic and technical explanations showed significantly higher satisfaction than the no-explanation group ($p = 0.047$), while technical-only explanations were marginally significant ($p = 0.056$), and empathetic-only explanations were not significant ($p = 0.120$). This suggests that while technical clarity can somewhat help, it is the combination with empathy that delivers the strongest effect.

And if that explanation came with a human touch (empathetic), it was even better. To break it down, first, we begin with basic Customer Satisfaction scores. When participants received no explanation for the offensive recommendation, they had low satisfaction. But when the platform followed up with either a technical or an empathetic response, satisfaction ticked upwards. The best case was the blend of both, when a message apologized with emotional warmth and offered a technical fix. Because this group had the highest satisfaction ratings. These findings offer partial support for H1: while both technical and combined explanations improved satisfaction compared to no explanation, only the combined condition reached full statistical significance.

In case of Interactional Justice, people reacted best when the explanation acknowledged their emotional experience. Purely technical statements were seen as better than nothing, but they lacked the personal care users seemed to crave. Technical explanations were marginally better than no explanation ($p = 0.071$) but empathy alone did not outperform either. This means that H2 was not supported: empathetic explanations did not significantly enhance justice compared to other types.

This suggests that technical content is necessary to generate fairness perceptions and empathy enhances (but does not substitute) this effect.

The regression analysis helped tie it all together. Interactional Justice wasn't just a nice feeling but it directly predicted Customer Satisfaction. This result was highly significant ($p = 0.0043$), indicating a strong relationship between justice and satisfaction. So, statistically, the justice factor significantly boosted satisfaction scores. This result strongly supports H3, confirming that perceived justice plays a key role in driving satisfaction outcomes.

As for internal validity, the credibility scale held up well (Cronbach's alpha = 0.70). That means users generally found the experimental setup believable and their responses were stable. The satisfaction and justice scales had slightly lower alpha values (around 0.60–0.63) which is acceptable given the exploratory nature of the research and the subjective nature of user perception (Field, 2013) .

Interestingly, the study also showed that people don't just care about what an explanation says but how it was said. In fact, participants who saw only a technical explanation still reported better reactions than those who got no response. On the other hand, empathy alone did not produce a statistically significant improvement which was contrary to initial expectations. This implies that technical explanations are foundational to restoring satisfaction and empathy is best used in conjunction with them.

So, when failures happen, the way they respond can make or break the user satisfaction. Brands that own up, explain clearly, and express genuine concern actually deepen user satisfaction. The study's biggest contribution is this: it reframes recovery from algorithmic errors not just as a technical issue but as a communicative one. Engineers should keep working on fairness and bias. But in the meantime, what systems say and how they say it can prevent real damage. On a broader level, this research shows that digital systems need to learn how to speak to us and not just calculate. This means combining clear technical explanations with messages that show care and understanding. Technical explanation is essential, but its impact can be amplified when delivered with empathy. As automation spreads into sensitive parts of our lives like identity, health, values, etc., so the designers must take empathy seriously.

This study has its fair share of limits though. First off, it's based on a simulation not a live social media platform. So while participants found the scenario believable, there's always a gap between imagined reactions and real-time emotions. Someone getting a problematic advertisement in the wild might respond with stronger emotion than they would in a study. Also, the belly fat scenario was chosen for its relatability and young audiences, but different offensive recommendations (example, gender, race, religion) could provoke different reactions and future studies should test more varied triggers. The sample was also demographically narrow, mostly students and young adults. These users are tech-savvy and may be more forgiving than older generations. The study also doesn't capture long-term behavioral outcomes. Did users forgive and move on? Did they come back to the brand? Those are questions for follow-up research.

Practically, though, the implications are strong. Brands should equip their AI with more than just logic. In other words, they need language that acknowledges harm and shows accountability. Developers and marketers need to work together here. A "cold" explanation might check the box, but an emotionally aware one might actually save the relationship. Because it's not just about helping the customer, but also about protecting the brand when smart systems make serious mistakes.

Table of Contents

1. INTRODUCTION

1.1. Problem Statement

When a social media algorithm displays a suggestion such as "Start shedding your belly fat today, no excuses?", it is considered as an Offensive Recommendation which is a form of service failure where the system delivers harmful or inappropriate outputs that violate user expectations or ethical norms. As noted by Rosenbaum et al., (2022), digital technology providers have an ethical duty to take steps to prevent unintended harm to consumers before launching their service technologies. Service Failures refer to instances where the system output causes confusion, mistrust or emotional harm due to a mismatch between system behaviour and user values. For example, showing a user recovering from an eating disorder advertisement about promoting extreme weight loss methods. In such cases, Service Recovery comes into play. It refers to deliberate communicative strategies used to address the failure, reassure the user and rebuild trust. Such as providing an immediate apology, explanation, and the option to adjust content preferences to prevent similar incidents. As Liao et al. (2021, p.6) pointed out, "explanation is not merely to provide transparency but support downstream user actions such as adapting interactions or acting on the decisions".

This study focuses on how user-friendly explanations influence customer reactions to algorithmic failures and how consumers react to such algorithmic failures, particularly in situations involving identity-sensitive product categories. Identity-sensitive categories include products or content that relate to race, gender, body image, or cultural values, in other words, areas where incorrect assumptions can feel highly personal or offensive. As Buolamwini and Gebru (2018, p. 11) note, "performance differs substantially for distinct subgroups," often resulting in the highest error rates for marginalized identities.

In case a social media post recommends content promoting unrealistic body standards to users who have previously searched for mental health support, it reflects an Algorithmic Failure. That means an outcome where automated decision-making generates results that violate ethical, contextual, or emotional expectations. Algorithmic failures occur when system outputs deviate from what users consider fair, accurate, or appropriate. This leads to confusion or harm. As Jones-Jang and Park (2023, p. 1) note, "AI can make mistakes and cause unfavorable consequences," highlighting the importance of addressing these decision-making errors. Such failures may stem from underlying system design or data limitations. This study focuses on how these breakdowns are communicated to users. Specifically, it examines how clear, user-friendly explanations can help mitigate the negative impact of these failures by acknowledging the error and ultimately restoring trust. "Explainability gives users assurance and confidence that AI systems work well, help developers understand why a system works a certain way, and safeguard against prejudice" (Shin, 2021, p.2). Such algorithmic failures have already led to corporate concern. "In a survey of chief marketing officers (CMOs), fielded by the CMO Council and Dow Jones Inc., most CMOs (78%) expressed concern about the threats to their brands' reputations from algorithm errors" (Srinivasan & Abi, 2021, p.1).

Algorithmic failures on social media such as offensive content recommendations provokes a strong negative reaction among users, particularly when these failures violate personal values, ethical expectations, or emotional well-being. For example, suggesting content related to extreme dieting to individuals recovering from eating disorders constitutes not just a service error but a trust breach that demands compassionate response. This challenge is especially relevant in social platforms where automated personalization drives engagement but can also result in lack of awareness of user context. As Srinivasan and Abi (2021) note, algorithmic errors can lead to brand harm crises as consumers increasingly expect fairness, empathy, and accountability in AI-driven interactions, especially when they have no control over the algorithmic decisions that affect them.

A related issue is the discomfort consumers experience when communicating with AI tools rather than humans. "As recent research suggests that 87% of consumers find human interaction in a service context more favorable than chatbots (Press, 2019). Customers also experienced discomfort when communicating with chatbots (Luo et al., 2019)" (Agnihotri & Bhattacharya, 2024, p.1).

Importantly, research emphasizes that users respond less negatively to failures when systems openly admit the mistake and respond with emotional intelligence. This is particularly true for minority or marginalized user groups who may experience disproportionate algorithmic harms (Agnihotri & Bhattacharya, 2024). Thus, recovery strategies should not only address technical flaws but also restore user dignity through empathetic and understandable communication.

Empirical cases also illustrate how algorithm errors can amplify user dissatisfaction. There had been some problems in Twitter timelines that resulted in the incorrect display of tweets for users. Some of these incorrectly displayed tweets had inappropriate content that had offended some Twitter users. In a scenario where Twitter's algorithm mistakenly displayed offensive tweets on user timelines, participants showed significantly more negative brand attitudes when the error was caused by a machine learning algorithm compared to a human or a standard algorithm (Srinivasan & Abi, 2021).

Lastly, consumers' willingness to remain loyal to a brand is deeply affected by the nature of the failure. "Participants' willingness to repeat the task would indicate a less negative response to the Qualtrics brand" (Srinivasan & Abi, 2021, p. 8). Consumers' willingness to re-engage with a brand dropped by over 22% when a service failure was attributed to human error versus algorithm error. 64.3% were willing to re-engage with the brand after an algorithmic mistake compared to 42.1% after a human mistake. (Srinivasan & Abi, 2021).

1.2. Literature Review

A focused review of relevant literature reveals both valuable insights and ongoing gaps regarding algorithmic failure recovery, particularly in social media contexts:

Agnihotri and Bhattacharya (2024) demonstrated that empathetic and anthropomorphic (human-like) chatbot behaviors enhanced consumer forgiveness and reduced negative word-of-mouth after service failures. However, these effects were highly dependent on the user's perception of the system's trustworthiness.

Westphal et al. (2023) highlighted the necessity of human-AI collaboration to reduce confusion caused by opaque system behaviors, which refer to situations where the system's internal logic or decision-making process is hidden or difficult for users to understand. They demonstrated that when users are given the ability to adjust or influence system judgments, allowing them to reinterpret or refine the AI's conclusions based on context or input, it can significantly enhance their understanding, trust, and willingness to comply with the system's output.

Srinivasan and Abi (2021) introduced the concept of mind perception of agency, which is, the extent to which users perceive an algorithm as having human-like intentions or the ability to act deliberately in response to algorithm errors. They found that when users believe algorithms lack agency, meaning, they see the system as an unthinking tool rather than a decision-making entity, they assign less blame for failures. However, when the algorithm appears human-like or emotionally aware (example, using empathetic tone or names like 'Erica' or 'Alexa'), user expectations and accountability rise which makes explanation quality or how clearly and convincingly the system communicates its reasoning more critical.

Explainable artificial intelligence (XAI) is the set of techniques developed primarily by machine learning researchers to mathematically or visually reveal how a model works. As stated by Adadi and Berrada (2018, p. 2), "XAI is a research field that aims to make AI systems results more understandable to humans." Liao et al. (2021) focused on the mismatch between algorithmic XAI methods and real-world design practices. Their study showed that user needs vary widely. XAI design (the way explanation tools are built into AI systems to make their decisions understandable) lacks actionable support and they proposed a question-driven framework. It is a design approach that starts by identifying the specific types of questions users ask about AI decisions, and then builds explanations tailored to those needs to bridge these gaps.

Shin (2021) examined how explainability (the extent to which an AI system can describe how a decision was made) and causability (the ability to justify that decision in terms users can causally understand) can influence user trust and perception in AI systems in the context of algorithmic recommendations, which are suggestions automatically generated by algorithms based on user data, behaviors, or preferences. The study found that causability and explainability shape trust through intuitive and systematic user evaluations. In other words, users rely both on gut reactions (intuition) and structured reasoning (systematic analysis) to judge whether they trust the AI's decisions, thereby enhancing AI acceptance and perceived performance. However, prior research has overlooked how users interpret AI explanations and how to evaluate the quality of those explanations.

Belanche et al. (2020) proposed a framework encompassing robot design (the physical and interactive attributes of the service robot, such as its appearance, speech capabilities, or level of automation), customer features (user traits such as expectations, prior experience with technology, or emotional state) and service encounter characteristics (the specific context in which the customer interacts with the robot, including the setting, task type, or timing) to guide both academic research and managerial decisions around service robots. A key gap identified was the lack of integrative frameworks for practically deploying service technologies in real-world user interactions. This framework is relevant to this study because it highlights how both user characteristics and interaction contexts affect user responses and insights that help explain how user-friendly explanations can shape customer reactions to algorithmic failures on social media platforms.

Collectively, these studies suggest a crucial gap: while technical interventions (example model retraining or data de-biasing) are well-studied, the communicative aspect of failure recovery particularly through user-friendly explanations tailored to emotional and ethical contexts in social media remains underdeveloped. One promising direction is the use of explainability-driven feedback loops: systems that provide clear, timely explanations of how and why a recommendation was made, followed by mechanisms that allow users to adjust or contest the logic. For example, if a social platform explains, "This content was suggested because of your recent engagement with fitness posts," and immediately offers the option to "Refine my preferences," the user may feel empowered rather than violated.

STUDY	FOCUS	KEY FINDINGS	GAPS IDENTIFIED
Liao et al. (2021)	To investigate gaps between algorithmic XAI methods and real-world design practices for explainable AI	User needs vary widely, XAI design lacks actionable support, and the study proposes a question-driven framework to bridge these gaps	XAI methods often fail to address diverse, real-world user questions and lack human-centered design support
Shin (2021)	To examine how explainability and causability influence user trust and perception in AI systems, especially in algorithmic recommendations	Causability and explainability shape trust through intuitive and systematic user evaluations, enhancing AI acceptance and perceived performance	Prior research overlooks how users interpret AI explanations and the quality metrics of explainability
Belanche et al. (2020)	To build a framework that guides both research and managerial decisions for implementing service robots.	A three-part framework, robot design, customer features, and service encounter characteristics that shapes successful service robot implementation	Lack of integrative frameworks guiding service robot research and practical deployment in service contexts.
Agnihotri and Bhattacharya (2024)	To study the impact of empathetic chatbot communication on service recovery	Empathetic, human-like chatbot behavior enhances forgiveness and reduces negative word-of-mouth; trust	Prior models lack focus on emotional design and user perception in AI-mediated recovery

		perception plays a moderating role	
Westphal et al. (2023)	Investigated how decision control and system explanations affect user trust, understanding, and compliance in human-AI collaboration.	Found that giving users control to adjust AI recommendations significantly improved trust, understanding, and compliance. Explanations, however, increased perceived task complexity and only helped when matched to users' cognitive abilities	Many explanations increase cognitive load and impair user outcomes unless tailored to the individual's cognitive ability. Most research fails to assess explanation quality or user fit.
Srinivasan and Abi (2021)	To assess how perceived agency affects blame attribution in algorithm errors	Users assign less blame to low-agency (non-human-like) algorithms, and more to human-like ones; emotional tone raises expectations	Limited frameworks addressing how algorithm design affects trust and perceived accountability

1.3. Research Question

This thesis investigates the following research question: How do user-friendly explanations influence customer reactions to algorithmic failures? This question will be addressed through an online experiment, which aims to simulate real-world user experiences and assess their responses to different types of algorithmic explanations following service failures. The study is motivated by the increasing presence of identity-sensitive algorithmic failures and the importance of emotionally and cognitively meaningful recovery strategies.

Chapter 2 presents the hypotheses development, beginning with a discussion on offensive product recommendations and their impact on user satisfaction. It then examines how service recovery theory and justice-based frameworks can be applied to algorithmic communication failures, leading to the formulation of three hypotheses.

Chapter 3 outlines the study design, including the structure of the online experiment, the development of the experimental scenario, the explanation conditions presented to participants, and the measurement scales used to assess satisfaction, perceived justice, and credibility.

Chapter 4 presents the findings of the experiment, including descriptive statistics, scale reliabilities, and hypothesis testing outcomes through t-tests and linear regression.

Chapter 5 discusses the practical and theoretical implications of the results and emphasizing the value of hybrid recovery messages (empathetic + technical) in restoring trust after algorithmic failures. It also reflects on the study's limitations, such as its simulation setting and demographically narrow sample and outlines key avenues for future research.

Chapter 6 concludes the thesis by summarizing the main findings. It emphasized the study's contributions to business practice and academic literature and also restating the significance of emotionally intelligent, transparent communication in algorithmic service recovery.

Chapter 7 contains the bibliography which listing all academic sources cited throughout the thesis.

2. HYPOTHESIS DEVELOPMENT

2.1. Offensive product recommendations and user satisfaction

The rise of personalized recommendation systems in digital services especially social platforms has brought both personalization benefits and significant risks, particularly when outputs conflict with users' values or identities. These systems analyze user data such as past searches, browsing habits, and purchase history to predict what the user might want next. They are powered by machine learning models, which means they learn from patterns in large datasets to make future predictions. However, when these systems make mistakes, such as offering body-shaming suggestions, racially insensitive products, or outdated gender-role items, they do more than fail at personalization. In other words, they disrupt expectations of fairness and empathy, triggering emotional and relational breaches. They violate the user's social and emotional expectations, potentially undermining trust in the platform (Burgoon, 1978). A growing body of research emphasizes the importance of empathetic, user-centered recovery strategies in such contexts.

According to Brand Congruence Theory, customers are more satisfied when a brand's message aligns with their self-concept (their internal view of who they are). This self-concept includes values, identity, lifestyle, and personal characteristics. Abosag et al. (2020) reinforces this by showing that when digital platforms make recommendations that contradict users' identity or values, such as when promoting stereotypical or offensive products, the users experience a sense of misalignment. This disconnect reduces the perceived relevance of the platform and creates emotional discomfort. For example, if a user who supports body positivity is shown a diet advertisement, it can feel like a personal attack or misunderstanding, leading to dissatisfaction. Therefore, when algorithms fail to reflect a user's identity appropriately, they damage brand-user alignment and undermine satisfaction.

2.2 Service recovery and perceived justice

When algorithmic errors occur, companies often have an opportunity to repair the emotional damage by explaining the situation. Sparks and McColl-Kennedy (2001, p.2) emphasize that "Service recovery techniques may include providing customers with explanations about the service failure, apologizing, empowering staff to resolve problems on the spot, making offers of compensation, and being courteous in the process (see for example, Bitner, 1990; Goodwin and Ross, 1990; Hoffman et al., 1995; Sparks and Callan, 1996; Blodgett et al., 1997)". Their findings indicate that the concern displayed by the service provider was especially important in terms of customers' reporting of satisfaction and likely re-use of the service product, highlighting the emotional impact of recovery communication. This study supports the idea that acknowledging customer concerns whether through empathetic statements or informative clarity plays a vital role in restoring confidence and rebuilding trust. By addressing the issue directly, the platform signals that it values the customer's experience, which can significantly improve satisfaction even when the failure cannot be undone.

The service recovery literature explains that successful complaint handling relies on three types of justice: procedural, distributive, and interactional, for example offering a clear refund process (procedural), providing fair compensation (distributive), and treating the customer with empathy and respect (interactional). Sparks and McColl-Kennedy (2001, p.2) found that interactional justice, which refers to how the customer is treated on a personal level, had the greatest impact on satisfaction. Specifically, "an employee's politeness and concern coupled with a strong effort to resolve the problem contributed to diffusing the problem in the mind of the consumer."

This aligns with Tax et al. (1998) justice-based model which emphasizes that customers judge complaint responses not just by outcome, but by the interpersonal quality of communication. As they explain, "a complaint is viewed as a conflict between the customer and the organization in which the fairness of (1) the resolution procedures, (2) the interpersonal communications and behaviors, and (3) the outcome are the principal evaluative criteria of the customer" (Tax et al., 1998, p.3)." Their findings underline that the most successful resolution strategies combine fair outcomes with emotional engagement, especially in cases where customers feel personally affected. This is further supported by the observation that "The employee or manager acting in a polite and empathetic manner, coupled with a strong effort to resolve the problem, contributed to diffusing customers' anger in many of the complaint incidents, whereas rude, uncaring behavior exacerbated the anger", and that "Firms that focus on providing dissatisfied customers with generous dispute settletnents are unlikely to reap the desired effects on satisfaction, commitment, and trust if the remuneration is delivered through unfair procedures or by uncaring employees" (Tax et al., 1998, p.15).

From a behavioral stendpoint, Zeithaml et al. (1996) found that perceived service quality, especially during recovery moments, influences not only immediate satisfaction but also long-term behavioral intentions like loyalty, advocacy, and complaint suppression. Their analysis showed that the strongest effects of both WP and OQ are on loyalty (.70 and .55) followed by switch (-.67 and -.47), pay more (.43 and .37) and external response (-.28 and -.21) in that order (Zeithaml et al., 1996), indicating that service quality has the greatest influence on whether customers stay with a brand or walk away. The study confirms that when users feel valued and emotionally understood, they are more likely to forgive and continue using the service even after failures. As the authors conclude, effective service recovery significantly improves all facets of behavioral intentions", and this includes reinforcing loyalty among customers receiving satisfactory resolution compared to those whose problems remained unresolved. Ultimately, the findings demonstrate the importance of strategies that can steer behavioral intentions in the right directions, including striving to meet customers' desired-service levels... and effectively resolving problems that do occur (Zeithaml et al., 1996).

These evidences suggest empathetic, user-friendly explanations are key to mitigating backlash from algorithmic failures. In addition to clarification, they humanize the system, restore trust, and preserve long-term user relationships, especially in emotionally sensitive contexts.

H1: The presence of a user-friendly explanation (either empathetic or technical) following an algorithmic failure increases customer satisfaction compared to no explanation.

The presence of a user-friendly explanation (either empathetic or technical) following an algorithmic failure increases customer satisfaction compared to no explanation. Perceived justice can be shaped by how respectfully and appropriately customers feel they are treated during a brand interaction. Grace and O'Cass (2005) emphasize that satisfaction is not just about resolving the issue, but also about whether the customer feels acknowledged and respected. Similarly, Chaudhuri and Holbrook

(2001) discuss how trust and honesty in brand communication influence consumers' evaluation of their experience. A user-friendly explanation either empathetic or technical, helps convey fairness and credibility. However, an empathetic explanation (by validating the customer's frustration) may evoke stronger affective responses than a purely technical one, which might be perceived as detached, or no explanation at all, which risks being interpreted as neglect. Therefore, both explanation types are likely to enhance satisfaction compared to no explanation, but empathetic framing may further boost the perception of fairness (Grace & O'Cass, 2005; Chaudhuri & Holbrook, 2001).

H2: Empathetic explanations lead to higher perceived justice compared to technical explanations or no explanation.

When a company responds to an algorithmic failure, how the explanation is delivered can influence not just satisfaction directly, but also how fairly the customer feels they were treated. Sparks and McColl-Kennedy (2001) highlight explanations showing emotional sensitivity, especially empathetic ones which enhance perceived fairness by acknowledging customer feelings. This fairness perception in turn plays a key role in how customers evaluate their overall experience. While a technical explanation may clarify the error, it may not generate the same sense of interpersonal care. Therefore, it is not the explanation alone, but the justice it conveys, that ultimately shapes satisfaction. In this way, perceived justice acts as a bridge between the explanation provided and how satisfied the customer feels afterward.

H3: Perceived justice increases customer satisfaction.

When users encounter algorithmic failures such as being shown harmful or insensitive content on social media, their reaction is shaped not only by the technical error but also by how fairly they feel they are treated in the aftermath. Perceived justice in this context, refers to the sense of fairness users experience across three key areas: whether the outcome feels fair (distributive justice), whether the process of resolving the issue seems reasonable (procedural justice) and whether they are treated with respect and dignity (interactional justice). "a complaint is viewed as a conflict between the customer and the organization in which the fairness of (1) the resolution procedures, (2) the interpersonal commu- nications and behaviors, and (3) the outcome are the princi- pal evaluative criteria of the customer" (Tax et el., 1998, p.3). Each of these dimensions plays a critical role in shaping user satisfaction after a service failure, as shown in the foundational study (Tax et al., 1998). Their findings reveal that when users perceive the platform's response as fair especially when it includes empathy and clear communication, they tend to report greater satisfaction and are more likely to stay loyal, even if the original mistake cannot be undone. In the case of algorithmic service recovery, i.e. the way a platform responds (prompt, honest, or compassionate) sends powerful signals about respect, care, and accountability. Emotional harm caused by AI-generated failures can often be softened when users feel genuinely acknowledged and fairly treated, highlighting the need for communication that feels more human, even when delivered by automated systems. In this way, perceived justice becomes a crucial psychological bridge, turning recovery efforts into renewed trust, satisfaction, and willingness to re-engage. Based on this, it is reasonable to hypothesize that perceived justice will have a positive impact on customer satisfaction, making it a vital part of emotionally meaningful recovery following algorithmic failures.

3. STUDY DESIGN

3.1. Online Experiment

In this study, online questionnaires were chosen as the primary method for data collection, rather than telephonic or physical interviews. This decision is grounded in the nature and purpose of the research. Telephonic or face-to-face interviews are typically employed in exploratory research, where the goal is to uncover themes, patterns, or insights in contexts that are not yet well understood. They are ideal for open-ended, qualitative inquiry, but less suited for studies where specific hypotheses are being tested. As noted by Malhotra (2010, p.108), "Exploratory research is characterized by flexibility and versatility with respect to the methods because formal research protocols and procedures are not employed. It rarely involves structured questionnaires, large samples, and probability sampling plans". The objective of conclusive research is to test specific hypotheses and examine specific relationships. Conclusive research is typically more formal and structured than exploratory research (Malhotra, 2010). In contrast, this research is theory-driven, with well-defined hypotheses derived from established frameworks. For example, H1 proposes that the presence of a user-friendly explanation (either empathetic or technical) following an algorithmic failure increases customer satisfaction compared to no explanation. H2 suggests that empathetic explanations lead to higher perceived justice compared to technical explanations or no explanation. H3 claims that perceived justice increases customer satisfaction. Online questionnaires are more appropriate in such contexts because they enable the collection of structured, standardized responses from a large sample which is essential for statistical testing. As per Kuphanga (2024), the questionnaire method ensures accessibility and facilitates the utilization of large sample sizes, thereby enhancing the reliability and robustness of research finding. The use of pre-validated scales and targeted items derived from the literature ensures that the data aligns closely with the theoretical constructs under investigation. As such, the questionnaire format supports the confirmatory nature of this study and facilitates meaningful quantitative analysis.

3.2. Scenario Development

This "belly fat" scenario was chosen because it represents a type of offensive recommendation that is likely to be widely relatable. Concerns about body image and weight are prevalent across diverse age groups, particularly among younger individuals. As Tiggemann (2004, pp.1-2) explains "Numerous studies have demonstrated considerable dissatisfaction with body size and shape among women, so much so that women's weight has been aptly described as 'a normative discontent' (Rodin, Silberstein, & Striegel-Moore, 1985)" and on addition to it "Although less well-documented among males, it appears that men and boys are also increasingly reporting body dissatisfaction, with a focus on muscularity (McCreary & Sasse, 2000)", indicating that such concerns are not limited to a specific demographic. For instance, consider a social media recommendation that reads: "Struggling with belly fat? Our 7-day cleanse is designed especially for busy students. Swipe to see how one university girl dropped 4kg in a week!" By using a diet-related product with a pushy tone and a visual comparison, the scenario simulates how algorithmic personalization can unintentionally trigger discomfort or reinforce harmful assumptions. This realism increases the likelihood that participants can empathize with the situation making it suitable for evaluating emotional and cognitive reactions to algorithmic failure, even if they have not experienced it directly.

	Empathetic	Non-Empathetic
Technical	"We're sorry this occurred. It was due to outdated data patterns in our system. We're working to correct it and respect all users."	"This was caused by a data error. We are updating our system."
Non-Technical	"We're truly sorry. This recommendation did not meet our standards for respectful personalization."	No explanation

To examine how different types of explanations influence user responses to this scenario, participants were presented with the following messages (above table).

3.3. Measurements

a) Customer Satisfaction (Gelbrich et. al, 2021)

This scale uses a 1 to 7 Likert format to assess overall satisfaction with a service. It includes three core items gauging satisfaction, happiness, and general approval of the service provided by the organization. An attention check item is embedded to ensure participant attentiveness.

On a scale of 1 to 7, to what extent do you agree with the following statements (1 = I don't agree at all; 7 = I totally agree):

I am satisfied with the service provided by the organization.
I am happy with the service provided by the organization.
I think the service provided by the organization is satisfactory.
This is an attention check, please select "I don't agree at all".

b) Interactional Justice (Sparks & McColl-Kennedy, 2001)

This scale measures perceived fairness in interpersonal treatment during customer service interactions. Rated on a 1 to 7 scale, items assess empathy, honesty, politeness, justification of the issue, and agent effort in handling the problem.

On a scale of 1 to 7, to what extent do you agree with the following statements (1 = I don't agree at all; 7 = I totally agree):

The customer service agent seemed very concerned about my problem.
I felt I was treated rudely.
The customer service agent did not appear to be telling me the truth.
I was given a reasonable account as to why the original problem occurred.
The customer service agent put a lot of positive energy into handling my problem.

c) Credibility (Tax et al., 1998)

This scale evaluates the perceived realism of the service scenario. Participants rate the believability of the employees, customers, and situations presented, and their ability to assume the role of the customer, using a 1 to 7 agreement scale.

On a scale of 1 to 7, to what extent do you agree with the following statements (1 = I don't agree at all; 7 = I totally agree):

I think there are employees like this in real life.
I think there are customers like this in real life.
I think there are service situations like this in real life.
I was able to adopt the role of the customer.

Together, these scales provide a reliable framework for assessing participants' satisfaction, justice, and the credibility of the service scenario. They support a structured understanding of customer responses.

4. FINDINGS:

Table 1. Descriptive statistics for customer Satisfaction.

	Mean	Standard Error	Minimum	Maximum	Cronbach's Alpha
CustomerSatisfaction_1	5.52	0.045	3	7	0.60
CustomerSatisfaction_2	5.39	0.048	2	7	0.60
CustomerSatisfaction_3	5.26	0.048	4	7	0.60

Table 2. Descriptive statistics for Interactional Justice

	Mean	Standard Error	Minimum	Maximum	Cronbach's Alpha
InteractionalJustice_1	5.12	0.041	3	7	0.63
InteractionalJustice_4	5.02	0.048	3	7	0.63
InteractionalJustice_5	5.21	0.055	1	7	0.63

Table 3. Descriptive statistics for Credibility

	Mean	Standard Error	Minimum	Maximum	Cronbach's Alpha
Credibility_1	5.28	0.049	3	7	0.70
Credibility_2	5.24	0.052	2	7	0.70
Credibility_3	5.27	0.058	2	7	0.70
Credibility_4	5.57	0.052	4	7	0.70

Table 4. Impact of Explanation Types

Explanation Type	CustomerSatisfaction	Interactional Justice	Credibility
No Explanation	5.28	5.08	5.24
Empathetic	5.41	5.15	5.38
Technical	5.44	5.21	5.48
Empathetic + Technical	5.44	5.01	5.28

Table 5. Consolidated Construct Statistics

	Mean	Standard Error	Minimum	Maximum	Cronbach's Alpha
CustomerSatisfaction	0.3731	0.0274	2	7	0.6055
InteractionalJustice	0.36997	0.0282	1	7	0.6410
Credibility	0.6694	0.0267	2	7	0.7616

In assessing the internal consistency and descriptive performance of the three main constructs used in this study, namely, Customer Satisfaction, Interactional Justice and Credibility, multiple statistical indicators were used and analyzed. All constructs were measured using multi-item scales on a 7-point Likert format, where higher values reflect stronger agreement and lower values reflect disagreement with the construct in question.

Customer Satisfaction was captured through three items. Descriptive results shown in Table 1 indicate relatively high means: CustomerSatisfaction_1 (M = 5.52, SE = 0.045), CustomerSatisfaction_2 (M = 5.39, SE = 0.048), and CustomerSatisfaction_3 (M = 5.26, SE = 0.048) with response ranges spanning from 2 to 7. The internal consistency across these items was moderate, with a Cronbach's alpha of 0.60.

Similarly, Interactional Justice was measured using three items, namely, InteractionalJustice_1, InteractionalJustice_4 and InteractionalJustice_5. Their means ranged from 5.02 to 5.21 (see Table 2) and a slightly stronger internal reliability of Cronbach's alpha = 0.63.

Credibility was assessed with four items, namely, Credibility_1, Credibility_2, Credibility_3, and Credibility_4. They also yielded favorable ratings, like, means ranging from 5.24 to 5.57 and standard errors ranging from 0.49 to 0.58 (see Table 3). This construct demonstrated the strongest internal consistency among the three, with a Cronbach's alpha of 0.70.

Comparative overview of mean scores across the four explanation conditions were presented through Table 4. Customer Satisfaction and Credibility were highest in the Technical and Empathetic + Technical groups (both M = 5.44 for satisfaction), suggesting that informative or blended explanations were most effective. While the Empathetic condition alone improved ratings relative to No Explanation, it did not outperform technical approaches. For Interactional Justice, the Technical explanation yielded the highest perceived fairness (M = 5.21), indicating that participants linked clarity with fair treatment more than emotional tone.

To evaluate reliability at the construct level, consolidated scores were calculated by averaging item responses per participant. As shown in Table 5, the consolidated reliability scores were Cronbach's alpha = 0.6055 for Customer Satisfaction, Cronbach's alpha = 0.6410 for Interactional Justice and Cronbach's alpha = 0.7616 for Credibility. According to Field (2013), the latter value for Credibility falls within an acceptable value for Cronbach's alpha. Field says (p. 2.037), "a value of .7 to .8 is an acceptable value for Cronbach's α; values substantially lower indicate an unreliable scale. Kline (1999) notes that although the generally accepted value of .8 is appropriate for cognitive tests such as intelligence tests, for ability tests a cut-off point of .7 is more suitable".

While the reliability values for Customer Satisfaction and Interactional Justice were slightly below the .7 threshold, they remain acceptable in the context of online questionnaires. Because they involve subjective and diverse respondent perspectives. As Field (2013, p. 2.037) also clarifies, "when dealing with psychological constraints values below even .7 can, realistically, be expected because of the diversity of the constructs being measured". Cortina (1993, as cited in Field) further emphasizes that Cronbach's alpha is influenced by both the number of items and the average inter-item correlation. For instance, "with more than 12 items, and fairly high correlations between items (r > .5), α can reach values around and above .7 (.65 to .84)" (p. 2.038). In this study, each construct was measured using relatively few items, namely, 3 for Satisfaction and Justice and 4 for Credibility, which justifies these slightly lower values.

Field (2013) illustrates with Cortina's example: "Cortina (1993) reports data from two scales, both of which have α = .8. The first scale has only three items, and the average correlation between items was a respectable .57; however, the second scale had 10 items with an average correlation between these items of a less respectable .28." (p. 2.038), concluding that the quality of inter-item correlation often matters more than the item count itself.

Results of Hypothesis Testing:

To evaluate the impact of user-friendly explanations on customer reactions, three hypotheses were tested using t-tests and regression analysis. Overall, the results offer partial support for the proposed relationships, with some conditions showing stronger effects than others.

H1(t-test)

Table 6. Satisfaction T-test.

	satisfaction of empathetic > satisfaction of no explanation	satisfaction of technical > satisfaction of no explanation	satisfaction of empathetic + technical > satisfaction of no explanation
T-statistic:	1.560	1.918	1.992
P-Value	0.120	0.056	0.047

H1 predicted that user-friendly explanations (either empathetic or technical) would improve customer satisfaction compared to offering no explanation. The results were mixed. The combination of both empathetic and technical explanations significantly increased satisfaction ($t = 1.992$, $p = 0.047$), indicating that layered communication which included both emotionally responsive and informative elements was the most effective. This is statistically significant at the $p < 0.05$ level, suggesting a meaningful improvement in satisfaction over providing no explanation.

Technical-only explanations were marginally significant ($t = 1.918$, $p = 0.056$), suggesting that clear information even without warmth, still had a noteworthy impact on satisfaction. Empathetic-only messages, on the other hand, did not significantly outperform no explanation ($t = 1.560$, $p = 0.120$). This is particularly interesting, as it implies that empathy alone is not sufficient to enhance satisfaction without additional context or reasoning.

H2(t-test)

Table 7. Justice T-test.

	justice of empathetic > justice of technical	justice of empathetic > justice of no explanation	justice of technical > justice of no explanation
T-statistic:	-0.780	0.975	1.811
P-Value	0.436	0.331	0.071

H2 expected empathetic explanations to be more effective than technical or no explanations in promoting perceived justice. This was not supported. The difference between empathetic and technical explanations was not statistically significant ($t = -0.780$, $p = 0.436$), nor was the difference between empathetic and no explanation ($t = 0.975$, $p = 0.331$). Interestingly, technical explanations were marginally significant in outperforming no explanation ($t = 1.811$, $p = 0.071$), hinting that participants may have equated technical clarity with procedural fairness more than with emotional validation.

Table 8. Interactional justice impacts on satisfaction.

Variable	Coefficient (Estimate)	Standard Error	T-Value	P-Value	95% CI
Intercept (a)	4.3978	0.1813	24.26	<0.0001	[4.0416, 4.7540]
InteractionalJustice (b)	0.2142	0.0748	2.86	0.0043	[0.0672, 0.3613]

H3 received strong support. A linear regression showed that perceived justice significantly predicted customer satisfaction, with a coefficient of $b = 0.2142$ and a p-value of 0.0043. This is highly significant ($p < 0.01$), confirming that when participants felt they were treated fairly (even if not emotionally comforted) they were more likely to feel satisfied with the service recovery.

To conclude, while empathy alone did not yield a significant improvement, the findings highlight the value of combining emotional intelligence with clear and technical reasoning. This hybrid approach appears most effective in repairing trust and maintaining customer satisfaction after an algorithmic failure.

5. DISCUSSION

This study set out to examine how user-friendly explanations, namely, empathetic, technical, or both shape customer reactions to algorithmic failures in a social media context. The findings offer refined insights into both business practice and academic discourse, revealing how service recovery communications can mitigate the damage of offensive recommendations and algorithmic errors. This section discusses the implications of the results for practitioners and scholars, the limitations of the current study and directions for future research.

5.1. Implications for Business Practice and Scientific Literature

Implications for Business Practice

From a business perspective, the results strongly suggest that providing an explanation message is better than offering no explanation at all following an algorithmic failure. Participants who received any form of explanation, be it technical, empathetic, or both, reported higher satisfaction and perceived fairness than those who received no explanation. This reinforces the idea that silence in the face of system failure is interpreted by users as neglect or indifference that can further erode trust in the platform.

However, the type of explanation matters. The most effective strategy was the combined empathetic and technical message because it significantly improved customer satisfaction ($p = 0.047$). This hybrid approach appears to resonate because it balances emotional warmth with informational clarity. Business practitioners should therefore consider designing recovery messages that both acknowledge user emotions and provide a credible and specific explanation for the failure. A well-crafted message might include an apology, an explanation of what went wrong (example, outdated data patterns) and a commitment to correction, which was reflected in the combined scenario used in the experiment: "We're sorry this occurred. It was due to outdated data patterns in our system. We're working to correct it and respect all users."

Purely technical explanations, while marginally effective ($p = 0.056$) were less impactful than the combined version and empathetic-only explanations were not statistically significant. This suggests that empathy without context may not be sufficient, possibly because users interpret vague apologies as evasive or pretentious. Businesses should be cautious not to rely solely on sentiment without actionable information. In sum, recovery communication must be emotionally intelligent and substantively informative to rebuild trust and retain users.

Contributions to Scientific Literature

The findings of this study offer important extensions to the literature reviewed in Section 1.2. Firstly, the results support Agnihotri and Bhattacharya's (2024) findings that empathetic communication improves consumer reactions but add nuance by showing that empathy is more effective when combined with technical clarity. This aligns with Sparks and McColl-Kennedy's (2001) emphasis on interactional justice and reinforces that politeness and emotional concern are vital but must be accompanied by resolution efforts to be fully effective.

Secondly, the study extends to Westphal et al. (2023) by illustrating how explanations influence trust and satisfaction even in non-collaborative, consumer-facing scenarios. Although Westphal et al. focused on user control and transparency in human-AI collaboration, the current results suggest that perceived fairness from system explanations alone can improve satisfaction, thereby extending their framework to automated service recovery without user input.

Moreover, the result that technical explanations increased justice perceptions marginally ($p = 0.071$) supports Shin's (2021) finding that explainability and causability influence trust. However, our study highlights a key gap in Shin's model: users do not always respond favorably to logical transparency alone. This reinforces the call by Liao et al. (2021) for explanations tailored to user values and emotional states rather than generic transparency.

Finally, the significant link between perceived justice and satisfaction ($p = 0.0043$) aligns with Tax et al. (1998) justice-based model and confirms the mediating role of interactional justice in service recovery. It emphasizes that even in digital contexts without human agents, users assess fairness based on how the system communicates which makes fairness perception an essential mechanism for recovery success.

In summary, this study extends existing literature by highlighting the value of hybrid explanation strategies and confirming the crucial role of perceived justice in algorithmic service recovery especially in emotionally sensitive identity-related contexts.

5.2. Limitations

Although the study yields valuable insights, it is essential to acknowledge several limitations that constrain the generalizability and interpretation of the findings.

First, the study was conducted through an online survey rather than on a live social media platform. While the credibility scale showed acceptable reliability (consolidated Cronbach's alpha = 0.76), there is an inevitable gap between participants' imagined responses and real-time emotional reactions. In natural settings users may respond with more intense affect or behaviorally disengage from the platform altogether. This phenomenon could not be fully captured in this setup.

Second, the study focused exclusively on one type of offensive recommendation: a body image-related message targeting "belly fat." Although this scenario was chosen for its relatability among younger demographics, it represents only a narrow slice of the broader spectrum of algorithmic failures. Offensive recommendations based on gender, race, religion or political beliefs may elicit different emotional responses and require different types of explanations. As such, the findings cannot be assumed to generalize to all forms of identity-sensitive errors.

Third, the participant sample was limited in diversity. Most respondents were university students or young adults, which introduces sampling bias. Younger users may be more forgiving of algorithmic failures or more familiar with AI interfaces, which could skew perceptions of fairness and satisfaction. Older users, individuals from marginalized groups or less tech-savvy populations might respond differently to the same explanations.

Fourth, while the scales for satisfaction and justice demonstrated acceptable internal consistency (Cronbach's alpha ≈ 0.60–0.63), these values fall slightly below the ideal threshold of 0.70. Although this is not uncommon in psychological research with subjective constructs (Field, 2013), it does suggest that measurement error may have reduced some effects, particularly in hypothesis H2 where empathetic explanations failed to achieve significance.

Finally, the study does not capture long-term behavioral consequences. While immediate reactions like satisfaction and perceived justice were measured, it remains unknown whether participants who received empathetic or technical explanations would remain loyal to the brand, reduce their usage or share their negative experience with others in real life.

In total, these limitations highlight the importance of treating the findings as contextually valid but not universally conclusive. Transparency in recognizing these constraints is vital for both scholarly rigor and practical application.

5.3. Directions for Future Research

To build on the findings and address the above limitations, several avenues for future research can be recommended.

First, future studies should explore a wider range of algorithmic failures, particularly those tied to other identity-sensitive dimensions such as race, gender identity, religion, or political orientation. Each context may provoke distinct emotional and cognitive responses and the optimal recovery message may vary accordingly. Comparative studies could reveal whether certain types of failures demand stronger empathy, more detailed technical explanations or other formats altogether.

Second, future researches should test the impact of explanation messages in live environments either through partnerships with platforms or longitudinal field studies. This would allow researchers to observe real-world behavioral responses such as reduced engagement, switching behavior or advocacy, thereby extending beyond the immediate subjective responses of satisfaction and justice.

Third, future work should engage diverse demographic samples including older adults, low-literacy users and underrepresented groups. These users may have different thresholds for offense, differing expectations of fairness and varying levels of trust in automated systems. Cross-cultural studies could also reveal important variances in response to explanations across different value systems and social norms.

Fourth, researchers might investigate multi-modal explanation delivery, such as combining text with visuals, voice assistants or chatbot interfaces. Previous literature, for example, Belanche et al., (2020) has emphasized the role of design in service recovery. Understanding how message format influences emotional response and perceived credibility would help optimize user experience.

Finally, future work could examine the long-term effects of recovery communication by tracking user behavior over time. Do people who receive empathetic + technical explanations come back? Do they rate the platform more favorably in future interactions? Such longitudinal data would enrich our understanding of how trust is repaired or lost, following algorithmic failures.

6. CONCLUSION

This study explored how user-friendly explanations influence customer responses to algorithmic failures in social media, particularly in emotionally sensitive identity-related contexts. It was motivated by growing concerns about offensive recommendations, such as weight-loss promotions targeting vulnerable users and how such incidents erode user trust. Building on service recovery theory and justice-based frameworks, this research examined whether different types of recovery messages (empathetic, technical, both or none) could improve user satisfaction and perceived fairness following such failures.

An online experiment was designed to test three hypotheses. H1 predicted that user-friendly explanations would improve customer satisfaction; H2 proposed that empathetic explanations would lead to higher perceived justice; and H3 hypothesized that perceived justice would positively influence satisfaction. Participants encountered a simulated offensive recommendation followed by one of four explanation conditions. Their responses were assessed using structured 7-point Likert scales measuring Customer Satisfaction, Interactional Justice, and Credibility.

The findings offer partial support for the proposed hypotheses. H1 was partially supported: the combined empathetic and technical explanation significantly improved satisfaction ($p = 0.047$), while technical-only explanations were marginally significant ($p = 0.056$) and empathetic-only messages were not ($p = 0.120$). H2 was not supported, as empathetic explanations did not significantly outperform others in increasing justice perception. However, H3 received strong support. Interactional justice significantly predicted satisfaction ($p = 0.0043$), highlighting fairness as a critical mediator in user experience after algorithmic errors.

These results provide important implications. For practitioners, the key takeaway is that recovery messages matter, and the best-performing messages combine emotional acknowledgment with concrete explanation. Businesses should avoid relying on vague apologies and instead design hybrid communications that reflect accountability and empathy. For academia, the study extends current literature on algorithmic failures, explainable AI, service recovery by empirically validating the role of justice perception and highlighting that logical transparency alone is insufficient without emotional resonance.

Nonetheless, the study has limitations. It was based on a simulated scenario, used a demographically narrow sample (mostly students and young adults) and tested only one type of identity-sensitive failure. Additionally, scale reliability for some constructs was slightly below ideal thresholds and long-term behavioral effects were not measured. These factors limit generalizability.

Future research should broaden the context by testing different types of offensive failures, engaging more diverse samples, conducting field experiments in live platforms and measuring behavioral outcomes over time. Such efforts will help refine guidelines for designing ethical and emotionally intelligent communication in AI-mediated environments.

Overall, this research reframes recovery from algorithmic failures as not only a technical challenge, but also a communicative one. In other words, trust is rebuilt not just through system accuracy, but through human-centered, transparent and empathetic messaging.

7. BIBLIOGRAPHY:

Gelbrich, K., Hagel, J., & Orsingher, C. (2021). Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence. *International Journal of Research in Marketing, 38*(1), 176-193. http://doi.org/10.1016/j.ijresmar.2020.06.004

Rosenbaum, M. S., Walters, G., Lear-Edwards, K. E., & Gonzalez-Arcos, C. (2022). The unintended consequences of digital service technologies. *Journal of Services Marketing, 36*(2), 97–109. http://doi.org/10.13140/RG.2.2.18746.26561

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81*, 1-15. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Malhotra, N. K. (2010). *Marketing research: An applied orientation* (6th ed.). Pearson Education. https://dokumen.pub/download/marketing-research-an-applied-orientation-6thnbsped-978-0-13-608543-0.html

Kuphanga, D. (2024). *Questionnaires in research: Their role, advantages, and main aspects* [Preprint]. ResearchGate. http://doi.org/10.13140/RG.2.2.15334.64325

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. http://doi.org/10.1109/ACCESS.2018.2870052

Abosag, I., Ramadan, Z. B., Baker, T., & Jin, Z. (2020). Customers' need for uniqueness theory versus brand congruence theory: The impact on satisfaction with social network sites. *Journal of Business Research, 117*, 862–872. http://doi.org/10.1016/j.jbusres.2019.03.016

Agnihotri, A., & Bhattacharya, S. (2024). Chatbots’ effectiveness in service recovery. *International Journal of Information Management, 76*, 102679. https://doi.org/10.1016/j.ijinfomgt.2023.102679

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Service robot implementation: A theoretical framework and research agenda. *The Service Industries Journal, 40*(3–4), 203–225. http://doi.org/10.1080/02642069.2019.1672666

Burgoon, J. K. (1978). A communication model of personal space violations: Explication and an initial test. *Human Communication Research, 4*(2), 129–142. http://doi.org/10.1111/j.1468-2958.1978.tb00603.x

Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *Journal of Marketing, 65*(2), 81–93. https://doi.org/10.1509/jmkg.65.2.81.18255

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G.B., & Rafaeli, A. (2023). Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. *Computers in Human Behavior, 144*, 107714. http://doi.org/10.1016/j.chb.2023.107714

Grace, D., & O'Cass, A. (2005). Examining the effects of service brand communications on brand evaluation. *Journal of Product & Brand Management, 14*(2), 106–116. http://doi.org/10.1108/10610420510592581

Liao, Q. V., Gruen, D., & Miller, S. (2021). *Questioning the AI: Informing design practices for explainable AI user experiences* [Preprint]. arXiv.https://doi.org/10.48550/arXiv.2001.02478

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*, 102551. http://doi.org/10.1016/j.ijhcs.2020.102551

Sparks, B. A., & McColl-Kennedy, J. R. (2001). Justice strategy options for increased customer satisfaction in a services recovery setting. *Journal of Business Research, 54*(3), 209–218. http://doi.org/10.1016/S0148-2963(00)00120-X

Srinivasan, R., & Abi, G. S. (2021). When algorithms fail: Consumers' responses to brand harm crises caused by algorithm errors. *Journal of Marketing, 85*(5), 74–91. http://doi.org/10.1177/0022242921997082

Tax, S. S., Brown, S. W., & Chandrashekaran, M. (1998). Customer evaluations of service complaint experiences. *Journal of Marketing, 62*(2), 60–76. http://doi.org/10.2307/1252161

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing, 60*(2), 31–46. http://doi.org/10.2307/1251929

Tiggemann, M. (2004). Body image across the adult life span: Stability and change. *Body Image, 1*(1), 29–41. http://doi.org/10.1016/S1740-1445(03)00002-0

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE Publications. https://sadbhavnapublications.org/research-enrichment-material/2-Statistical-Books/Discovering-Statistics-Using-IBM-SPSS-Statistics-4th-c2013-Andy-Field.pdf

Jones-Jang, S. M., & Park, Y. J. (2023). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication, 28*(1), 1-8. https://doi.org/10.1093/jcmc/zmac029

8. APPENDIX:

8.1. Questionnaire Used in the Study

Participant Consent

Dear respondent,

Thank you for participating in this survey! My name is B Samvav and I am a Master of Management in Data Science student at Hasselt University. As part of my master's thesis, I am conducting a research on how do user-friendly explanations influence customer reactions to algorithmic failures ?

There are no right or wrong answers. Your personal and honest opinion is what counts. Completing this questionnaire takes about 3 minutes.

With this form we ask for your permission to process your personal data in the context of an educational project.

Information concerning the processing of your data: -

We wish to process your data for research analysis, teaching and potential scientific publication, including the publication of my master thesis.

We will collect the following data:

In this survey, you will be shown an advertisement image, after which you will be asked for your opinion. At the end, you will answer demographic questions, including age and level of study.

Data processing responsibility and contacts:-

Your data will be processed under my responsibility (B Samvav, b.samvav@student.uhasselt.be).

For questions about your rights and all other matters concerning the processing of your personal data, you can also always contact our Data Protection Officer, he/she can be reached via dpo@uhasselt.be. He/She independently monitors compliance with regulations regarding the processing of personal data at UHasselt.

Consent to participate:-

By completing this questionnaire, you agree to participate in this research. Participation is voluntary and you can stop at any time without consequences. Your answers will be processed anonymously and used exclusively for research purposes.

If you have any questions about this research, you can contact the researcher via the email address: b.samvav@student.uhasselt.

Consent :

By entering this research, you are indicating that you have read the consent form and that you participate voluntarily. You give your consent to the UHasselt to process your personal data in the manner and according to the modalities described in this form.

- I give consent to participate in this research project. I have received, read and understood the above as a condition for my consent to the processing of my personal data for the research project.
- I do not consent and I do not wish to participate in the study.

Scenario

Please imagine the following situation:

You find yourself overweight and have been struggling with bad body image for years. Even though your Body Mass Index (BMI) is in healthy zone, you are having a hard time having a good image of yourself. While scrolling on social media, you are shown this advertisement:

"Struggling with belly fat? Start shedding it today – no excuses."

You find this suggestion intrusive and uncomfortable. You decide to report the advertisement.

Customer Service Response Conditions

You receive a reply from customer service. Four explanation conditions were used:

- Empathy Explanation "We're truly sorry for the recommendation you received. We understand that this may have been frustrating or upsetting. We are actively working to ensure that our recommendation systems treat every customer with the care, dignity, and respect they deserve."
- Technical Explanation "This recommendation occurred due to a data-related issue in our system. Our algorithm mistakenly prioritized outdated patterns in the data. We are actively working to fix this."
- Empathy + Technical Explanation "We're truly sorry for the recommendation you received. This occurred due to outdated patterns in our data system. We understand that this may have been frustrating or upsetting. We are actively working to fix this and ensure that our recommendation systems treat every customer with the care, dignity, and respect they deserve."
- No Explanation "We have received your complaint on our recommendations. We are working to improve them."

Customer Satisfaction (7-point Likert scale)

Please indicate your agreement with the following statements:

- I am satisfied with the service provided by the organization.
- I am happy with the service provided by the organization.
- I think the service provided by the organization is satisfactory.
- (Attention check) Please select "I don't agree at all."

Interactional Justice (7-point Likert scale)

Please indicate your agreement with the following statements:

- The customer service agent seemed very concerned about my problem.

- I felt I was treated rudely.
- The customer service agent did not appear to be telling me the truth.
- I was given a reasonable account as to why the original problem occurred.
- The customer service agent put a lot of positive energy into handling my problem.

Credibility (7-point Likert scale)

Please indicate your agreement with the following statements:

- I think there are employees like this in real life.
- I think there are customers like this in real life.
- I think there are service situations like this in real life.
- I was able to adopt the role of the customer.

Demographic Questions

Age

- Under 18
- 18–24
- 25–34
- 35–44
- 45–54
- 55–64
- 65+

Gender

- Male
- Female
- Non-binary / third gender
- Prefer to self-describe
- Prefer not to say

Education

- Some primary school
- Completed primary
- Some secondary school
- Completed secondary school
- Vocational or similar
- Some university, no degree
- Bachelor’s degree
- Graduate or professional degree (MA, MS, MBA, PhD, etc.)

- Prefer not to say

Final Feedback (Optional)

Do you have any additional comments or feedback about this study?
[Free text box]

End of Survey

We thank you for your time spent taking this survey. Your response has been recorded.

8.2 Data File

Data File: Data_Excel.xlsx – https://docs.google.com/spreadsheets/d/1PdzzILNNQdrDQPGIo-QzyQIGvelNRLbg/edit?usp=share_link&ouid=102629044697791109366&rtpof=true&sd=true
The above link takes you to the excel sheet.