Evaluating the Impact of Assignment Group and Category Classification Prediction of Incoming Service Requests on the Perceived Service Quality

A Quasiexperimental Study in the Enterprise Software Industry

Van Gassel, Jeroen; Janssen, Marijn

DOI
10.1109/TEM.2024.3401545

Publication date
2024

Document Version
Final published version

Published in
IEEE Transactions on Engineering Management

Citation (APA)
Van Gassel, J., & Janssen, M. (2024). Evaluating the Impact of Assignment Group and Category Classification Prediction of Incoming Service Requests on the Perceived Service Quality: A Quasiexperimental Study in the Enterprise Software Industry. *IEEE Transactions on Engineering Management*, *71*, 10740-10750. https://doi.org/10.1109/TEM.2024.3401545

Evaluating the Impact of Assignment Group and Category Classification Prediction of Incoming Service Requests on the Perceived Service Quality: A Quasiexperimental Study in the Enterprise Software Industry

Jeroen van Gassel and Marijn Janssen

***Abstract*—Machine learning (ML) is reshaping customer service, tackling the growing complexity and volume of customer requests. This article investigates the effects of ML on perceived service quality (PSQ) across various customer service measures within an organizational context. Utilizing a quasiexperimental design, we analyzed 131 978 service requests submitted to the service desk of a large enterprise software organization. Over a 2-year period, these requests were made by 1252 organizations and were associated with 85 654 predictions and 19 720 returned PSQ postservice questionnaires. Our regularized logistic regression model aimed to categorize service requests into 56 categories and dispatch them into one of nine assignment groups to enhance resolution efficiency. Contrary to expectations, the overall PSQ did not significantly improve, while five specific metrics, such as time to resolve and first-time resolution, improved. This may be attributed to the increase in the first personal response time. The article highlights the complexities of implementing ML-based classification and underscores the importance of organizational structure. We found that expert groups prioritizing accurate problem-solving over quick responses led to an increase in the response time for incoming service requests. Theoretical contributions include an understanding of how classification in customer service affects PSQ, offers practical tactics to counteract negative impressions, and sets the groundwork for future article on ML in customer service management, despite the limitations, such as the potential influence of external factors and the study's generalizability.**

***Index Terms*—Assignment group and category classification, customer service, machine learning, organizational aspects, perceived service quality, service desk, service requests.**

I. Introduction

CUSTOMER service is a mission-critical process aimed at creating customer value and consequently facilitating the execution of organizational strategies [1]. Providing excellent customer service is crucial due to its direct impact on overall customer satisfaction and indirect organizational profitability [2]. However, due to the explosive growth of customer service systems, communication technologies, and complexity, handling customer service requests has become more challenging [3]. The variety and volume of customer service requests increase continuously, and customers have different expectations and demands to which customer service must respond [4]. Also, communication channels, such as email and self-service portals, are increasingly used to interact with customer service [5], [6]. However, contrary to the phone channel, self-service channels, such as email and portals, do not facilitate interactive communication to respond to and triage incoming service requests, which can lead to less satisfied customers [7].

Consequently, more and more service desk agents are needed to process all incoming service requests. Furthermore, email and portal-created service requests usually have an unstructured text format, making triaging service requests more challenging [8]. Hence, it has become crucial to establish an effective customer service system to support and anticipate the rising and diverse customer demands and meet their expectations in an increasingly complex and demanding service environment. [4]. Furthermore, recent studies across various industries have established that perceived service quality (PSQ) directly impacts customer satisfaction, leading to improved purchase intention [9], [10].

Existing literature suggests that PSQ depends on the performance of human agents but also on the design and performance of the information systems (IS) [11], [12]. Many organizations have started using machine learning (ML) to improve their customer service experience [13], [14]. ML refers to several methods that can help to enhance the execution of specific tasks through training experience [15]. ML applications are developed by training a model using historical data to improve performance when executing tasks [15]. ML can potentially transform the customer service experience: examples include the automatic dispatching of work tasks to employees [16], informing customers on predicted delayed service times [17], automatically handling routine tasks, recommending the next step in a business

Manuscript received 26 January 2024; revised 1 April 2024; accepted 7 May 2024. Date of publication 15 May 2024; date of current version 5 June 2024. Review of this manuscript was arranged by Department Editor G. Marzi.
(Corresponding author: Jeroen van Gassel.)

Jeroen van Gassel is with the Faculty of Technology, Policy & Management, Delft University of Technology, 2728BX Delft, Zuid Holland, The Netherlands (e-mail: j.j.g.vangassel@tudelft.nl).

Marijn Janssen is with the School of Technology, Policy and Management, Delft University of Technology, NL-2600 GA Delft, The Netherlands, and also with the Faculty of Technology, Policy & Management, Delft University of Technology, 2728BX Delft, Zuid Holland, The Netherlands (e-mail: Marijnj@tbm.tudelft.nl, m.f.w.h.a.janssen@tudelft.nl).
Digital Object Identifier 10.1109/TEM.2024.3401545

process, and optimized agent staffing based on case call center arrival's forecasting [18].

Although many organizations and industries recognize that creating exceptional customer experience has become a priority, empirical evidence on using ML to improve PSQ is scarce [9]. Only a few studies have investigated the application of ML within specific areas of customer service [14]. These include service level agreement violation prediction within IT-service management and an ML intelligent knowledge-based chatbot for customer service [19], [5]. Others investigated specific ML use cases, such as improving customer complaint management velocity by automatic email classification [20]. These studies examine ML in a context outside the organization, do not assess its impact on the organization, and have a limited emphasis on the role of human decision-making. To our knowledge, Li et al. [3] were among the first to investigate the relationship between ML and its impact on service quality, shedding light on the pivotal role of voice AI systems in customer service and demonstrating its influence on consumer behavior outcomes. However, the impact and the organizational implications of ML on the customer service PSQ are hardly understood.

The current article investigates the effects of ML classification on perceived customer service quality in a real-life setting. Thus, formulates the following two research questions (RQs):

RQ1: What is the effect of ML-based assignment group and category classification of incoming service requests on the perceived customer service quality?

RQ2: What are the organizational implications and operational factors influencing the extent to wherein the ML improved the perceived customer service quality?

The aim of this article is to investigate whether ML classification can improve customer service PSQ in organizations. Specifically, we investigated the effect of automating assignment group classification and category classification using ML. To analyze the impact of ML application on PSQ measures, data were collected for 2 years—1 year before and 1 year after ML was applied. The analysis compared differences in the overall PSQ and six specific PSQ measures between both periods. In summary, we contribute to filling the knowledge gap by investigating the effect and magnitude of ML classification on specific and overall customer service PSQ measures. We also provide insights into the effect of organizational structures, agents, and changes in technology on the adoption of ML in customer service. Although it might limit generalizability, it gives insight into the effects in real-life settings and can reveal new influencing variables.

This rest of this article is organized as follows. Section II presents the literature background about ML and PSQ, followed by our research approach in Section III. In Section IV, we discuss research findings based on an actual customer service process. Section V provides the recommendations for further article. Finally, Section VI concludes this article.

II. Background

Recent studies investigated ML applications in automating tasks, optimizing staffing [18], and identifying service improvement areas [14]. However, there is limited empirical evidence regarding the effects of ML system implementation on customer service performance [21]. Therefore, we first review existing work on models for measuring PSQ for customer service to understand specific customer service process-related measures. The derived measures are then used as input to review existing work on classification applied within customer service processes and the identified customer service process-related measures. Consequently, we use the result of the literature background as input for conducting interviews to determine the specific customer service PSQ measures. Finally, the measures are used to describe the scope of the quasiexperiment and determine where our research complements existing work.

A. *Perceived Service Quality and Measures*

For organizations, it has become crucial to build an appropriate customer service support system to anticipate increasing customer demands and meet the expectations of service requestors. Often, models like SERVQUAL are used to understand and improve the PSQ. PSQ can be evaluated by a postconsumption evaluation method that expresses the customers' discrepancy between the expected and PSQ [22]. In these studies, the customer is asked to assess different aspects of the PSQ on a specific scale. The lowest score represents a "very dissatisfied" customer, whereas the highest score represents a "very satisfied" customer [22]. Usually, one assessment aspect represents the overall PSQ [23]. If the PSQ is lower than the expected PSQ, the quality is disconfirmed, and the overall PSQ will be lower. Conversely, when the PSQ confirms or exceeds the expected PSQ, the overall PSQ will be higher [23].

There are a variety of models for measuring the PSQ. SERVQUAL can be considered the most used model to calculate the PSQ, primarily in physical service transactions. SERVQUAL contains five constructs (tangibles, responsiveness, reliability, assurance, and empathy); twenty-two measures were generated to represent the five constructs [24]. With the development of IS, the SERVQUAL model faced more challenges in addressing context-specific PSQ measures [25]. The impact of the tangible construct decreased mainly as most measures are less applicable to IS. For example, the item "Physical facilities are visually appealing" is less relevant within IS as most interactions today occur online due to the widespread use of digital channels [8]. Therefore, Kettinger and Lee [26] revised SERVQUAL multiple times to address specific PSQ applications [27]. Examples of commonly used PSQ models derived from SERVQUAL are IS-SERVQUAL, Software as a Service (SaaS-Qual), and a quality model for cloud services [28], [29].

In conclusion, there are a high number of models for measuring PSQ. Existing models, such as SERVQUAL, IS-SERVQUAL, and SaaS-Qual measure, to some extent, the PSQ for customer service processes. Consequently, based on a literature review, nine high-level measures (service response reliability, service delivery reliability, service delivery quality, service responsiveness, service delivery expectations, service response expectations, user experience, service customization, and service contract reliability) were derived from these models and used to identify existing literature on how classification can improve certain aspects of PSQ. The nine high-level measures were also used as input for selecting the relevant measures in the quasiexperiment.

B. Classification in Customer Service

Classifying service requests, such as distinguishing complaints from general queries, aids in identifying urgent cases. Extensive research has been conducted to classify incoming service requests, especially on detecting emotions in human–computer dialogs, to help pinpoint customers needing immediate attention [3], [31], [32], [33]. Studies like Gupta et al. [31] extract email sentiments and Karthik and Ponnusamy [33] categorizing emotions for better response prioritization. However, Ilk et al. [34] found that classifying live chats accurately can reduce mismatches between customers and agents by 14%, enhancing efficiency. Also, in more recent research, ML-based classification of incoming IT tickets has demonstrated that it can have higher accuracy than manual classification. However, while this improvement in classification accuracy is noteworthy, its direct impact on the overall PSQ remains assumed [30]. These studies show that classification is a promising ML technique; however, its implications for PSQ are hardly understood.

C. Machine Learning and Perceived Service Quality

The complexity of ML technology makes it challenging to assess its real-world effectiveness beyond training data. Deployed ML systems must adapt to unforeseen situations, such as handling exceptions or organizational changes, which can impact their performance [12]. Therefore, Wang et al. [21] argue that understanding how ML implementation affects service performance is essential for further empirical research.

To our knowledge, no studies investigated the effect of classification on the overall customer service PSQ score and, consequently, on individual PSQ factors derived from models for measuring the PSQ.

Coussement and Poel [20], Khowongprasoed and Titijaroonroj [30], and Nenkova and Bagga [8] argued that automatically classifying incoming requests into a small group of classifiers (2, 3, and 5, respectively) helps agents to identify service requests that require an immediate and correct response. However, classifying incoming service requests into a more significant number of categories and dispatching incoming service requests to more assignment groups is a critical challenge for agents [35]. Therefore, we investigated whether ML would help agents improve the classification accuracy of 56 service request categories and nine assignment groups. ML is often used to automate decision-making fully; however, human-supported decision-making is given less attention.

In contrast to earlier studies (Coussement and Poel [20] Khowongprasoed and Titijaroonroj [30], and Nenkova and Bagga [8], in this article), ML will not automatically classify incoming service requests. Instead, ML predicts the category and assignment group within a second after each service request has been created and presents them as values when the agent starts working on the service request. The human agent could apply the recommended classification values or overrule ML with another category or assignment group. Suggesting service request classification values complements Reis et al. [36] findings that ML can make sense of complex tasks that humans may miss. However, it allows agents to correct the prediction if it is considered incorrect. This hybrid model should improve PSQ measures and potentially improve the overall PSQ. Also, whereas previous articles focused on one simultaneously applied classification capability, we focus on the impact of two simultaneously applied classification applications. A model does not yet exist to measure the customer service PSQ specifically. This article focuses on one overall PSQ and six specific PSQ measures derived from the nine high-level PSQ measures considered in existing literature as a critical aspect of the overall PSQ.

III. Research Approach

We aim to investigate ML classification's effect on the PSQ. For this, we conducted a quasiexperiment (Fig. 1) by introducing ML classification in a customer service process. A quasiexperiment investigates causal relationships without fully randomized participant assignment but still involves manipulating the independent variable and comparing groups [37]. One type of quasiexperiment is studying interventions by comparing before and after an intervention using testable propositions. This is particularly suitable for complex situations where a controlled experiment might not reveal all influencing variables. In a quasiexperiment conducted in practice, the relevant variables influence the outcomes. This enables the investigation of factors from practices that might not be visible in advance. Not only is looking at the predefined dependent and independent factors, but given the complexity, other variables might also play an influential role. We used triangulation on two levels to ensure the study's overall reliability and intrinsic and extrinsic validity. First, we used a set of generic PSQ measures derived from the literature background as input for semistructured interviews with seven experts from different backgrounds to identify relevant customer service PSQ measures. The rationale for selecting the PSQ measures stems from experts' expectations that these measures influence the overall PSQ through the introduction of ML-based assignment group and category classification. Only measures that can be quantified using the service request and survey performance data were included.

Second, to understand which unexpected operational factors might have influenced each quasiexperiment PSQ measure and, consequently, the PSQ score, we interviewed three agents, each having more than 5 years of tenure in customer service within the enterprise. To minimize the potential influence of researcher bias on the quasiexperimental results, the researchers were not directly involved in the execution of the article. This way, they could observe and analyze the organizational impact qualitatively and quantitatively. Furthermore, we used this to evaluate construct validity by observing if the measurement reflected the intended constructs.

The quasiexperiment starting situation was without ML, whereas ML was introduced without having any other changes that might affect the effect. The situation without and with ML was compared using PSQ measures. We evaluated seven PSQ measures whereby the overall PSQ was expected to be influenced positively by assignment group and category classification [4]. Two measures were expected to improve the service timelines

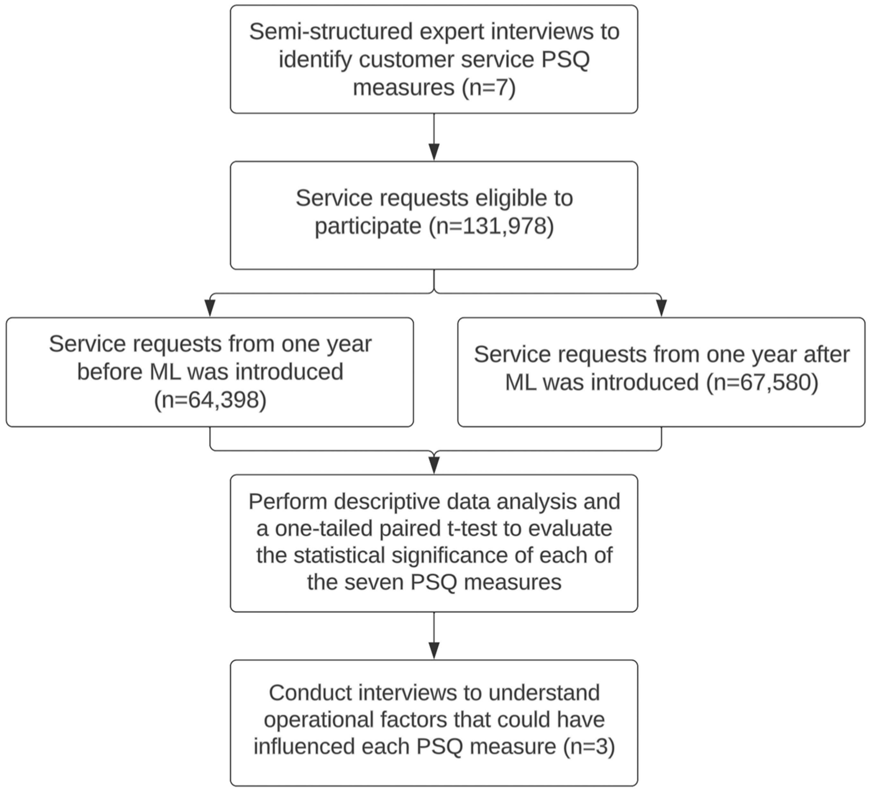

Fig. 1. Flowchart of the quasiexperimental study method.

("time to resolve"—representing the time between the creation and resolution a service request related to a customer need, and "first personal response time"—meaning the time between service request creation and personal agent receipt notification), two measures were expected to improve the service quality ["% first time right" (% FTR)—representing the percentage of service requests where the first assigned group was able to handle the request until resolution, and "average reassignment count"—meaning the average number of assignment groups that were assigned to service request until resolution], and two measures were expected to improve the service expectations ["variance in responsiveness"—representing the performance variation of time to resolve and the first personal response time calculated as the average standard deviation (STD), and "variance in reliability"—meaning the performance variation calculated as the STD of the average reassignment count] [38].

Data were collected from the customer service system of a large enterprise, a business-to-business software organization providing a subscription-based cloud platform solution to structure and automate various business processes. The large enterprise serves a diverse clientele, including small, medium, and large organizations across the globe, representing all industries. Customers can submit service requests via an online service portal. Included service request types are generic questions, unexpected product results, implementation issues, outages, requests, and performance issues. Service requests classified under "other" were omitted, as their pertinence to actual service needs could not be verified. Additionally, we excluded service requests initiated by automated systems, such as integration or performance monitoring solutions, due to their irrelevance to our analysis. Furthermore, we noted that enterprise employees made some service requests on behalf of customers; these were excluded because they do not represent the initial point of contact between the service requestor and the agent, rendering the PSQ measures invalid for our analysis. Our dataset was also cleansed of requests missing an assignment group, those with resolution times recorded before their opening times, instances where open and close times were nearly identical, and cases where the reaction time occurred within 1 s of the service request's creation.

After resolving a service request, the enterprise sends an email containing a postservice questionnaire to each requestor to gather feedback. A survey consisted of six questions, and we only used the scores of the "overall support experience" for this service request' evaluation criteria. The score ranges from 0, representing a "very dissatisfied" customer, to 10, representing a "very satisfied" customer. A dataset containing 131 978 service requests submitted by 1252 customer organizations was analyzed. In total, 64 398 data points were created the year before ML was introduced, and 67 580 were collected the year after ML was introduced (Fig. 1). The motivation for using regularized logistic regression is its adeptness with high-dimensional text data, such as our unlabeled service request description. This method identifies complex patterns for accurate classification and provides a probability model, which is valuable when offering agent prediction values, including probability [39]. Regularization is essential to guard against overfitting, ensuring the model remains generalizable and reliable [40]. Our quasiexperiment includes

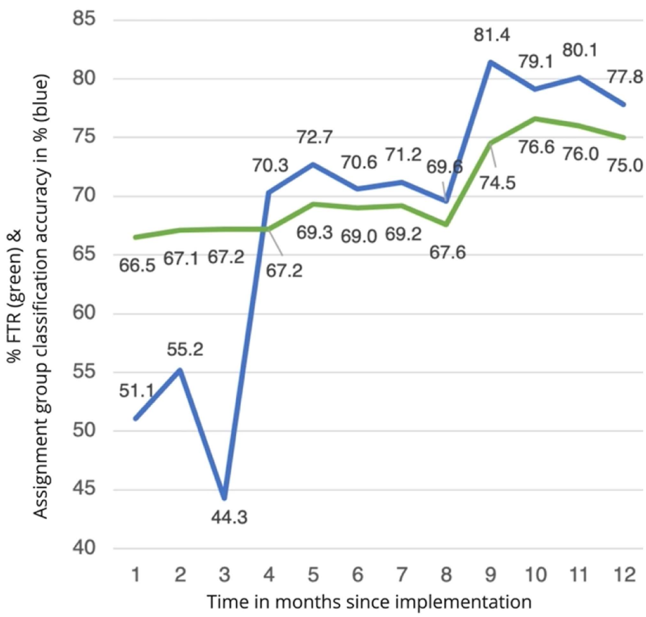

Fig. 2. Overview of % FTR and assignment group classification accuracy during a 1-year period after ML implementation. Blue = % assignment group classification accuracy. Green = % of service requests resolved without assigning to another assignment group—% FTR.

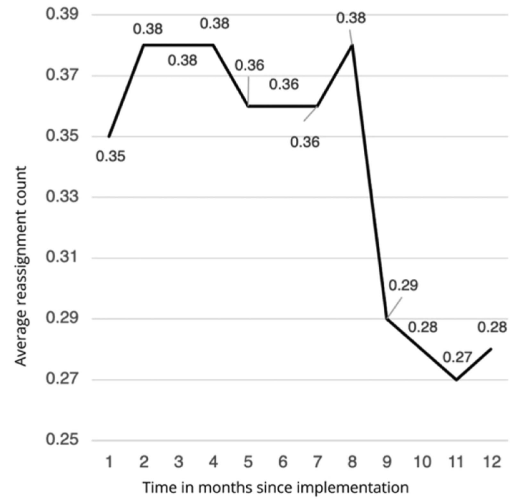

Fig. 3. Overview of average reassignment count over 1 year after ML implementation (number of service request reassignments from one to another assignment group).

service requests related to interruption of service, unexpected results of their product, usage requests, product performance issues, overall complaints, inquiries, and questions. To enable comparison before and after the experiment, only customer organizations that raised at least three service requests and completed at least two surveys 1 year before and 1 year after ML had been deployed. All service requests were excluded for situations without direct interaction between the service requestor and human agent. Hence, the following types of service requests were excluded: those created by a service monitoring system, service requests not created directly by a customer, those where resolution time was equal or before creation time, and all service requests that have a reaction time of less than a second after creation time.

IV. Findings

A. Interview

One overall and six individual measures were identified as relevant quasiexperiment PSQ measures. All experts mentioned time to resolve, first personal response time, and FTR by six experts, variance in reliability by three experts, number of reassignments, and variance in responsiveness by two experts. One expert said: "Time to resolve is the number one thing." Other PSQ measures, such as the friendliness of the agent and the price paid for the service, were omitted, and the experts mentioned that introducing ML-based assignment groups and category classification does not directly influence these measures.

B. Descriptive Data Analysis

Fig. 2 shows the development of the measures % FTR depicted in green in relation to the assignment group prediction accuracy (blue line). Fig. 3 shows the development of the measure average reassignment count. Both figures depict the development of the measures until 1 year after the implementation of assignment group classification. Both % FTR and average reassignment count show a progressive, long-term improvement after assignment group classification was deployed, reflecting the strong positive trend in classification accuracy. From the eighth to the ninth month after implementation, the classification accuracy increased from 69.6% to 81.4% (Fig. 2, blue line). Simultaneously, the PQS measures average reassignment count, and % FTR improved significantly: % FTR with 9.2% (Fig. 2, green line) and average reassignment count with 18.4% (Fig. 3, black line).

C. Classification Accuracy Development

In terms of prediction, after implementing ML classification for service requests, the following classifications were made: 26 740 (39.5%) requests had both the assignment group and category classified, 31 422 (46.6%) requests had only the category classified, 752 (1.1%) requests had only the assignment group classified, and 8666 (12.8%) requests had no classification applied. From the 27 492 assignment group classifications (40.7% coverage), 19 220 were classified correctly (69.9% accuracy). For the 58 162 category classifications (86.1% coverage), 26 798 were classified correctly (46.1% accuracy).

Fig. 4 (blue line) shows the 12-month trend reflecting the increasing assignment group classification accuracy from 51.5% to 78.3% 1 year after ML was implemented (52.0% accuracy increase). Also, the category classification accuracy (orange line) increased from 43.5% to 53.1% a year later (a 22.1% accuracy increase).

The initial classification accuracy and the classification accuracy development during the year were consistently lower for category classification than for assignment group classification. This is because there are 56 categories versus 9 assignment

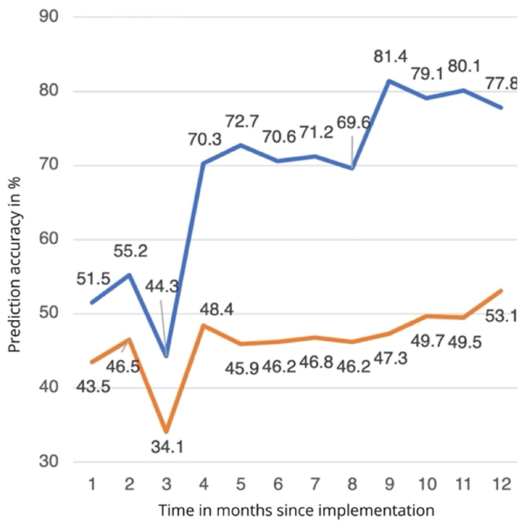

Fig. 4. Overview of assignment group (blue line) and category (orange line) classification accuracy after ML was applied.

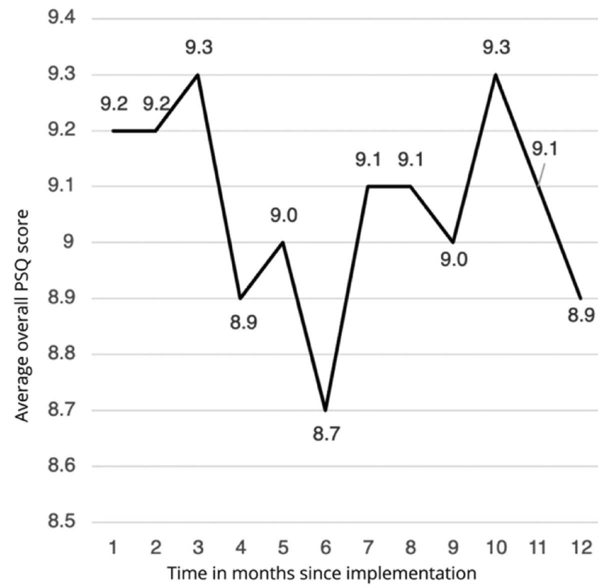

Fig. 5. Average overall PSQ score during a 1-year period after ML go-live.

groups, making it harder to get a high accuracy for category classification than for assignment group classification: category has, on average, fewer data per category item to get trained accurately. In addition, the category classification has coverage of 86.1% vs. 40.7% for assignment classification, whereas the higher coverage (% of service requests where prediction has been applied) results in lower accuracy. Also, the classification accuracy shows a progressive, long-term improvement due to the algorithm learning effect: wrongly predicted service requests are corrected by human agents and consequently will be picked up by the ML algorithm in the following training round.

D. Evaluation of the PSQ Measure Before and After ML has been Deployed

A one-tailed paired t-test was used to determine whether the difference in PSQ score and the six measures before and after ML were statistically significant. Table I shows the results of the descriptive data analysis and the one-tailed t-test for seven PSQ measures. The mean of the overall PSQ score was 8.901 (1.24) before ML was implemented and 8.903 (1.190) after ML was implemented. This difference (0.002) was not statistically significant ($p = 0.4764$).

Regarding the separate PSQ measures, five improved after implementation of ML (first time right +3.8%; time to resolve −4 days; variance in responsiveness −48.9%, variance in reliability −0.03, and average reassignment count −0.06), whereas one PSQ measure deteriorated (first personal response time +1 h and 47 min). Variance in responsiveness is measured as the average STD of the two responsiveness measures, time to resolve and the first personal response time. These differences were statistically significant (all p-values < 0.0001).

Fig. 5 shows the development of the overall PSQ score calculated as an organizational average over 1 year, starting after the assignment group and category classification go-live date. The after-go-live date shows two major decreases in overall PSQ: between the third and sixth months and the tenth and twelfth months. Interviewees mentioned that this is most likely a seasonal trend where the following factors play a role.

1) A relatively high number of problems were administratively closed between the third and sixth months after implementation. Consequently, all related long-running service requests were closed, resulting in a peak in the mean and STD of the resolution time and decreased PSQ scores.
2) Between the third and sixth months after implementation, the period had fewer working days and decreased capacity, resulting in longer resolution times.
3) An increase in backlog leads to a higher response.

Finally, in addition to time to resolve, % FTR, and average reassignment count, two measures show significant improvements: the variance in reliability and the variance in responsiveness [reflected by the STD and interquartile range (IQR)]. Both measures show an improvement after ML has been deployed across all measures in the scope of this quasiexperiment: Resolution time (STD: −24.9%, IQR: −26.0%), first personal response time (STD: −72.3%, IQR: +72, 7%), reassignment count (STD: −5.5%, IQR: −18.4%), and PSQ score (STD: −2.4%). Therefore, assignment group classification and category classification mechanisms can profoundly affect service delivery.

V. Discussion

An explanatory study was conducted using a quasiexperiment. The article covered 131 978 service requests submitted by 1252 organizations, 85 654 applied predictions, and 19 720 PSQ-returned surveys. ML has been deployed to match incoming service requests into 56 existing categories and to dispatch incoming service requests into one of the nine assignment groups that will most likely resolve the service request. The classified values for the incoming service requests were suggested options for agents to select while creating service requests. The agent could apply the recommended classification values or

TABLE I
RESULT OF THE DESCRIPTIVE DATA ANALYSIS AND THE ONE-TAILED t-TEST ON SEVEN PSQ MEASURES

Measure	Measure	Pre-ML period (1 year) n=1 252	ML period (1 year) n=1 252	Delta (absolute/relative)	p value *= statistically significant
Time to resolve (HRS:MIN:SEC)	Mean	27 days 16:00:05	23 days 02:52:22	–4 days 13:07:43/ –16.4%	$p < 0.0001$ *
	Median	24 days 02:04:22	20 days 22:00:22	–3 days 04:04:00/ –13.2%	-
First personal response time	Mean	06:06:59	07:54:37	01:47:38/ 29.3%	$p < 0.0001$ *
	Median	05:10:12	07:22:12	02:12:00/42.6%	-
Average reassignment count	Mean	0.408	0.348	-0.06 /–14.9%	$p < 0.0001$ *
	Median	0.381	0.317	–0.064 /–16.8%	-
First time right	Percent	66.8%	70.6%	3.8 / 5.8%	-
Variance in responsiveness	STD of time to resolve	18 days 05:39:38	13 days 16:36:07	–48.9%, calculated as an average of both underlying measures:	$p < 0.0001$ *
	STD of first personal response time	15:45:44 HRS:MIN:SEC	04:22:16 HRS:MIN:SEC	–24.9% STD time to resolve –72.8% STD time to personal respond	$p = 0.0433$
Variance in reliability	STD of average reassignment count	0.206	0.195	–0.03 /–5.5%	$p < 0.0001$ *
Overall PSQ	Mean	8.901	8.903	0.002 /0.0%	$p = 0.4764$
	(STD)	(1.24)	(1.190)	(–0.03 /–2.4%)	

overrule ML with another category or assignment group. The article used data covering 2 years, thus spanning a relatively long period: this diminished day and monthly variations due to seasonal influences. Surprisingly, our investigation into the effect of ML-based assignment group and category classification of incoming service requests on the customer service PSQ revealed that both capabilities did not improve the overall PSQ measure in this article (Table I), although some specific measures were improved. In our quasiexperiment, which investigated the organizational implications and operational factors influencing the extent to which ML improved perceived customer service quality, we explored the influence of organizational structures on the PSQ. We found that expert groups prioritizing accurate problem-solving over quick responses led to a decrease in the response time for incoming service requests.

A. *Theoretical Implications*

The current article answers Wang et al.'s [21] call for more empirical research on how the implementation of ML affects service performance. Also, our findings highlight the organizational complexity of improving service quality through automated classification. This aligns with existing literature emphasizing the nuanced elements, such as the subjective valuation of being human, essential for the perfect substitution between ML and human-delivered services [41]. We also complement existing literature on understanding the relationship between assignment group classification and category classification on customer service PSQ. Saberi et al. [4] suggested a strong correlation between the classification of incoming service requests for directing them to the right operator and an improved PSQ with empirical data on the impact of assignment group classification and category classification on one overall PSQ and six specific PSQ measures for customer service. Khowongprasoed and Titijaroonroj [30] focused more on comparing human versus ML classification and an assumed positive impact on the overall PSQ. A wide variety of studies solely researched how to optimize classification algorithm accuracy for classifying incoming service requests [42], [43].

This article supplements the current literature, where ML classification is used to help agents classify incoming service requests more quickly and accurately within three areas. First, we found that classification can improve the performance of five specific customer service PSQ measures (time to resolve, first time right, reassignment count, the variance in reliability, and the variance in responsiveness). Second, we found that classification can be applied to customer service processes containing a magnitude of possible predicted values successfully [this article contained 65 values (9 assignment groups and 56 categories)] by suggesting agents a category and assignment group to select, rather than automatically applying: Coussement and Poel [20], Khowongprasoed and Titijaroonroj [30], and Nenkova and Bagga [8] used 2, 3, and 5 classifiers, respectively, whereby the predicted value was automatically applied. Third, the current article complements Wang et al. [21] on how the implementation of ML affects service performance and is the first to investigate the implications of assignment group classification and category classification on operational aspects of an organization and the implications of organizational changes on the impact and accuracy of assignment group classification and category classification.

The article showed that one PSQ measure changed unfavorably after ML implementation, whereas five PSQ measures improved. A large unfavorable increase in the first personal response time was observed after introducing ML (the mean and median increased by 29.3% and 42.6%, respectively). A possible explanation for this is that automatically dispatching incoming service requests to expert groups will result in an extended first personal response time. In contrast to help-desk agents, agents in expert groups focus on solving complex problems correctly rather than replying as quickly as possible to incoming service requests. The other five individual PSQ measures (time to resolve, % FTR, average reassignment count, the variance in reliability, and the variance in responsiveness) improved after ML implementation, some substantively (mean −16.4%, +5.8%, −14.9%, −5.8%, and −48.9%, respectively). The overall PSQ score did not change significantly (0.001% mean increase). Several potential explanations exist for this lack of overall PSQ improvement after ML implementation. First, it may be that unfavorable changes in some individual PSQ measures compensated for favorable changes in other individual PSQ measures, resulting in a null change in the overall PSQ score. However, this seems unlikely as only one individual PSQ measure changed unfavorably, whereas five individual PSQ measures changed favorably. These changes were similar in effect size compared with the unfavorable change. Second, operational developments within the enterprise that occurred during the period of the quasiexperiment might have influenced individual PSQ measures that were not included in this article but might have contributed to the overall PSQ score. All three quasiexperiment-related interviewees mentioned introducing a new support group as the cause of operational confusion. One interviewee said, "An organizational realignment whereby an additional assignment group was introduced caused a lot of manual re-assigning service requests to assignment groups." The original algorithm was calibrated to the pre-existing structure. With the new group added, the parameters and data patterns the algorithm relied on have shifted, causing a mismatch between service request attributes and group responsibilities. Consequently, the system's confidence in making accurate predictions has waned, necessitating a higher degree of manual intervention to ensure service requests are assigned correctly.

After updating and finetuning the algorithm, the prediction accuracy improved subsequently. The interviewees also argued that the seasonality of new products directly impacts the PSQ: After a product is released, existing problems get closed as "fixed" or "not fixed." The number of incidents related to these problems is higher in this period, and they are prone to low PSQ scores as they are long open. The interviewees also mentioned the following operational changes: the complexity of the underlying emerging product and the conception of customer urgency that influenced PSQ measures. These factors (and possibly other factors not mentioned by the interviewees) might have influenced one or more PSQ measures or the overall PSQ score to an unknown extent.

Both metrics, average reassignment count and accuracy of first-time assignments, improved after implementing ML-based classification. However, there was a shift in manual intervention for misclassified assignments, as agents could no longer return service requests to the original assignment group but had to identify the correct one. This affected the time-to-resolve metric, which could negatively impact overall PSQ.

B. Practical Implications

The findings suggest that the impact of ML capabilities might have a more complex relationship than thought. Although the overall PSQ score did change significantly, the findings show that the use of ML in service management processes has the following advantages: the improved time to resolve, % FTR, average reassignment count, the variance in reliability, and the variance in responsiveness. However, there are also disadvantages, e.g., expert assignment groups' longer first personal response time. These insights can help managers avoid the disadvantages when implementing ML or implement countermeasures, such as an intelligent chatbot for customer service, to improve response times [5]. Models conceptualizing and evaluating ML should consider the different types of capabilities addressed. However, operational factors, such as the complexity of the underlying emerging product, backlog clean-up action, or reorganizing the customer service department, can also play a role.

When organizations aim to improve the PSQ within customer service with ML, this article provides important insights into five implementation success factors, including strategies to mitigate the negative impacts of ML on PSQ measures, such as first personal response time and misclassifications. First, organizations should be mindful when applying ML to new or realigned parts of their business because of the potential lack of representative data or relevant training features to train the model to maximize relevance, classifier accuracy, recall, and precision [44]. To ensure the model maintains classification accuracy despite new situations, organizations can opt to train the new model in a subproduction environment and replace the existing production model only when the classifier accuracy, recall, and precision meet the minimum acceptable performance levels. Second, organizations should anticipate who in an expert group is responsible for correcting faulty predictions in advance. Another mitigation strategy is that faulty predicted records should be automatically routed to first-level support, which typically handles triaging and classifying service request categories, thereby ensuring work is assigned to the appropriate group. This prevents expert groups from wasting time locating the correct group and avoids the issue of service requests being shuffled around the organization, as experts often lack reassignment knowledge, ultimately leading to longer resolution times and, consequently, lower overall PSQ.

Third, the PSQ measures can be continuously monitored, including a threshold for minimum performance, and the ML model can be used to identify performance drops as they occur and investigate any causal relationship between PSQ and ML model performance deterioration.

Fourth, process participants should anticipate the operational implications through better knowledge of the ML-related decisions and operational impact on the PSQ measures [45], e.g., expert groups need to react faster to incoming new service requests to ensure a quick first personal response time. This can be achieved simply by assigning agents a dispatch role responsible for picking up incoming cases and assigning them to the right agent within that team. Also, ML can be used to predict response time or resolution time to inform users about the situation and that their expectations might not be relevant anymore. This will reduce the delta between the expected and actual PSQ and will most likely result in a move positive overall PSQ [24]. Additionally, a potential migration strategy is to implement a queue prioritization system to avoid long personal response times. This ensures fair handling of all service requests and reduces overall queue time by addressing more straightforward requests first [46].

Fifth, if a prediction's accuracy is too low, the top three predicted values can be presented as options rather than automatically implementing ML-predicted values. Users are empowered to override the predicted values with their preferred choices. Any manually overridden value should be incorporated as training data to enhance the precision of the ML model.

Our focus was on two types of capabilities, and the effect of the implementation and adoption of those two capabilities were found to be different. This suggests that not all ML capabilities have the same effect. More research into the effects of the other capabilities is needed. Also, understanding which capabilities will result in which types of improvement are needed. Nevertheless, the findings of this article can be used as a guideline for implementing and adopting ML.

C. Limitations and Future Article

This article was conducted within a single organization with a large dataset covering the interaction with 1252 customer organizations with 19 720 PSQ-returned surveys. We compared the old situation with the new one and used the old situation as the control group. The disadvantage is that we have a control group without the same factors, such as the increased complexity of the emerging product, a backlog cleanup action resulting in the closure of many long-running service requests, and the introduction of a new support group. The absence of the control group impacts the article's internal validity, making it more challenging to attribute observed changes in PSQ directly to the interventions. Other factors, such as the length of time an organization has been a customer of the large enterprise software organization, could inherently change their perceptions and how they respond to service quality surveys [47]. Finally, internal factors, such as the consistency of interaction between the requester and a particular service agent, might also influence the overall PSQ. This consistency can lead to survey response biases within a single service instance or over several interactions. Requesters who frequently interact with the same agent might report higher satisfaction due to familiarity or higher or lower expectations based on previous service experiences [48].

In further article, we have four main recommendations to validate the current article's findings. First, this article provides the foundations and the findings can be generalized statistically by employing a replication strategy for quasiexperiments on customer service processes to identify and assess external factors affecting PSQ. This facilitates easy comparison of both replicated studies, allowing for a deep dive into discrepancies, enhancing result validity, and enabling targeted service improvements. [49]. Second, the article can be expanded to different ML capabilities. More specifically, we recommend investigating the impact of service request throughput time prediction on PSQ, as customer service expectations are considered a critical aspect of the overall PSQ [6], [50]. The third recommendation is to expand and examine the relationship between the dimensionality of the PSQ measures. The first step would be to list all measures that might influence the overall customer service PSQ. As a second step, conduct a factor analysis to identify measures having a high impact on the PSQ that potentially can be improved by ML. Fourth, we recommend including a control group in future quasiexperiments where feasible. This addition will help determine if changes in PSQ are due to ML or external factors. A control group remains unaffected by the ML intervention, providing a clear baseline for comparison.

VI. Conclusion

This article was structured around two pivotal questions: what is the effect of ML-based assignment group and category classification of incoming service requests on the perceived customer service quality, and what are the organizational implications and operational factors influencing the extent to wherein the ML improved the perceived customer service quality? We have found that the effects of ML classification on PSQ are not addressed within organizational settings. This article revealed that ML resulted in both positive and negative effects. The article shows that the assignment group and category classification capabilities positively impacted the % FTR and average reassignment count measures. It also significantly influenced the PSQ measures time to resolve, the variance in reliability (−16.4%, −5.5%, respectively), and it improved the responsiveness variance to some extent (−48.9%). However, assignment group and category classification had no significant positive effect on the overall PSQ score. The negative effect included a large and significant increase in the first personal response time. This is because agents in expert groups focus on solving complex problems correctly rather than replying as quickly as possible to incoming service requests, resulting in longer lead times for the first personal response. Hence, the introduction of ML might have unexpected organizational consequences. Other operational factors that impacted the PSQ include the complexity of the emerging product, a backlog cleanup action resulting in the closure of many long-running service requests, and the introduction of a new support group. This suggests that the use of ML for improving customer service processes might be more complex than is often assumed in the literature, and organizational changes are needed. The type of capability and context should be considered when evaluating the impact of ML.

REFERENCES

[1] V. Arvidsson, J. Holmström, and K. Lyytinen, "Information systems use as strategy practice: A multi-dimensional view of strategic information system implementation and use," *J. Strategic Inf. Syst.*, vol. 23, no. 1, pp. 45–61, 2014, doi: 10.1016/j.jsis.2014.01.004.

[2] D. Golmohammadi, M. Parast, and N. Sanders, "The impact of service failures on firm profitability: Integrating machine learning and statistical modeling," *IEEE Trans. Eng. Manage.*, vol. 69, no. 6, pp. 3038–3052, Dec. 2022, doi: 10.1109/TEM.2020.

[3] B. Li, L. Liu, W. Mao, Y. Qu, and Y. Chen, "Voice artificial intelligence service failure and customer complaint behavior: The mediation effect of customer emotion," *Electron. Commerce Res. Appl.*, vol. 59, May/Jun. 2023, Art. no. 101261, doi: 10.1016/j.elerap.2023.101261.

[4] M. Saberi, O. Khadeer Hussain, and E. Chang, "Past, present and future of contact centers: A literature review," *Bus. Process Manage. J.*, vol. 23, no. 3, pp. 574–597, 2017, doi: 10.1108/BPMJ-02-2015-0018.

[5] E. W. T. Ngai, M. C. M. Lee, M. Luo, P. S. L. Chan, and T. Liang, "An intelligent knowledge-based chatbot for customer service," *Electron. Commerce Res. Appl.*, vol. 50, Nov./Dec. 2021, Art. no. 101098, doi: 10.1016/j.elerap.2021.101098.

[6] P. F. Hsu, T. K. Nguyen, and J. Y. Huang, "Value co-creation and co-destruction in self-service technology: A customer's perspective," *Electron. Commerce Res. Appl.*, vol. 46, Mar./Apr. 2021, Art. no. 101029, doi: 10.1016/j.elerap.2021.101029.

[7] R. W. Buell, D. Campbell, and F. X. Frei, "Are self-service customers satisfied or stuck?," *Prod. Oper. Manage.*, vol. 19, no. 6, pp. 679–697, 2010, doi: 10.3401/poms.1080.01151.

[8] A. Nenkova and A. Bagga, "Email classification for contact centers," in *Proc. ACM Symp. Appl. Comput.*, 2003, pp. 789–792, doi: 10.1145/952532.952689.

[9] S. H. Liao, D. C. Hu, and H. L. Chou, "Consumer perceived service quality and purchase intention: Two moderated mediation models investigation," *Sage Open*, vol. 12, no. 4, pp. 1–15, Oct. 2022, doi: 10.1177/21582440221139469.

[10] M. Qin, W. Zhu, S. Zhao, and Y. Zhao, "Is artificial intelligence better than manpower? The effects of different types of online customer services on customer purchase intentions," *Sustainability*, vol. 14, no. 7, Apr. 2022, Art. no. 3974, doi: 10.3390/su14073974.

[11] A. C. Chen, "Linking service quality to perceived operator-computer contribution in customer services," in *Proc. 40th Int. Conf. Comput. Ind. Eng.: Soft Comput. Techn. Adv. Manuf. Service Syst.*, 2010, pp. 1–4, doi: 10.1109/ICCIE.2010.5668242.

[12] W. J. Orlikowski, "The duality of technology: Rethinking the concept of technology in organizations," *Org. Sci.*, vol. 3, no. 3, pp. 398–427, Aug. 1992, doi: 10.1287/orsc.3.3.398.

[13] M. Zdravković, H. Panetto, and G. Weichhart, "AI-enabled enterprise information systems for manufacturing," *Enterprise Inf. Syst.*, vol. 16, no. 4, pp. 668–720, 2021, doi: 10.1080/17517575.2021.1941275.

[14] X. X. Liu and Z. Y. Chen, "Service quality evaluation and service improvement using online reviews: A framework combining deep learning with a hierarchical service quality model," *Electron. Commerce Res. Appl.*, vol. 54, Jul./Aug. 2022, Art. no. 101174, doi: 10.1016/j.elerap.2022.101174.

[15] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015, doi: 10.1126/science.aaa8415.

[16] K. Kirkpatrick, "AI in contact centers," *Commun. ACM*, vol. 60, no. 8, pp. 18–19, 2017, doi: 10.1145/3105442.

[17] E. Chocron, I. Cohen, and P. Feigin, "Delay prediction for managing multiclass service systems: An investigation of queueing theory and machine learning approaches," *IEEE Trans. Eng. Manage.*, vol. 71, pp. 4469–4479, 2024, doi: 10.1109/TEM.2022.3222094.

[18] T. Albrecht, T. M. Rausch, and N. D. Derra, "Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting," *J. Bus. Res.*, vol. 123, pp. 267–278, 2021, doi: 10.1016/j.jbusres.2020.09.033.

[19] S. M. Practice, T. Vafeiadis, and K. Diamantaras, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, 2015, doi: 10.1016/j.simpat.2015.03.003.

[20] K. Coussement and D. Van Den Poel, "Improving customer complaint management by automatic email classification using linguistic style features as predictors," *Decis. Support Syst.*, vol. 44, no. 4, pp. 870–882, 2008, doi: 10.1016/j.dss.2007.10.010.

[21] L. Wang, N. Huang, Y. Hong, L. Liu, X. Guo, and G. Chen, "Voice-based AI in call center customer service: A natural field experiment," *Prod. Oper. Manage.*, vol. 32, no. 4, pp. 1002–1018, Apr. 2023, doi: 10.1111/poms.13953.

[22] D. K. Tse and P. C. Wilton, "Models of consumer satisfaction formation: An extension," *J. Marketing Res.*, vol. 25, pp. 204–212, 1988, doi: 10.2307/3172652.

[23] G. A. Churchill and C. Surprenant, "An investigation into the determinants of customer satisfaction," *J. Marketing Res.*, vol. 19, no. 4, pp. 491–504, 1982, doi: 10.2307/3151722.

[24] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *J. Retailing*, vol. 64, no. 1, pp. 12–40, 1988, doi: 10.1016/S0148-2963(99)00084-3.

[25] E. Babakus and G. W. Boller, "An empirical assessment of the SERVQUAL scale," *J. Bus. Res.*, vol. 24, no. 3, pp. 253–268, 1992.

[26] W. J. Kettinger and C. C. Lee, "Pragmatic perspectives on the measurement of information systems service quality," *MIS Quart.*, vol. 21, no. 2, pp. 223–239, 1997, doi: 10.2307/249421.

[27] W. J. Kettinger and C. C. Lee, "Zones of tolerance: Alternative scales for measuring information system service quality," *MIS Quart.*, vol. 29, no. 4, pp. 607–623, 2005.

[28] Y. Hwang and D. J. Kim, "Customer self-service systems : The effects of perceived web quality with service contents on enjoyment, anxiety, and e-trust," *Decis. Support Syst.*, vol. 43, pp. 746–760, 2007, doi: 10.1016/j.dss.2006.12.008.

[29] X. Zheng, P. Martin, K. Brohman, and L. D. Xu, "Cloudqual: A quality model for cloud services," *IEEE Trans. Ind. Inform.*, vol. 10, no. 2, pp. 1527–1536, May 2014, doi: 10.1109/TII.2014.2306329.

[30] K. Khowongprasoed and T. Titijaroonroj, "Automatic thai ticket classification by using machine learning for IT infrastructure company," in *Proc. 19th Int. Joint Conf. Comput. Sci. Softw. Eng.*, 2022, pp. 1–6, doi: 10.1109/JCSSE54890.2022.9836250.

[31] N. Gupta, M. Gilbert, and G. Di Fabbrizio, "Emotion detection in email customer care," *Comput. Intell.*, vol. 29, no. 3, pp. 489–505, 2013.

[32] A. H. Chiang, S. Trimi, and Y. J. Lo, "Emotion and service quality of anthropomorphic robots," *Technol. Forecasting Social Change*, vol. 177, Apr. 2022, Art. no. 121550, doi: 10.1016/j.techfore.2022.121550.

[33] K. Karthik and R. Ponnusamy, "Adaptive machine learning approach for emotional email classification," in *Proc. 14th Int. Conf. Towards Mobile Intell. Interact. Environ.* , 2011, pp. 552–558, doi: 10.1007/978-3-642-21616-9.

[34] N. Ilk, G. Shang, and P. Goes, "Improving customer routing in contact centers: An automated triage design based on text analytics," *J. Oper. Manage.*, vol. 66, no. 5, pp. 553–577, Jul. 2020, doi: 10.1002/joom.1084.

[35] M. Jäntti, "Towards an improved IT service desk system and processes: A case study," *Int. J. Adv. Syst. Meas.*, vol. 5, no. 3, pp. 203–215, 2012.

[36] C. Reis, P. Ruivo, T. Oliveira, and P. Faroleiro, "Assessing the drivers of machine learning business value," *J. Bus. Res.*, vol. 117, pp. 232–243, 2020, doi: 10.1016/j.jbusres.2020.05.053.

[37] D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. USA: Ravenio Books, 2015.

[38] R. Day, "Modeling choices among alternative responses to dissatisfaction," *Adv. Consum. Res.*, vol. 11, no. 1, pp. 496–500, 1984.

[39] T. Zhang and F. J. Oles, "Text categorization based on regularized linear classification methods," *Inf. Retrieval*, vol. 4, no. 1, pp. 5–31, 2001, doi: 10.1023/A:1011441423217.

[40] B. M. Hsu, "Comparison of supervised classification models on textual data," *Mathematics*, vol. 8, no. 5, pp. 7077–7091, May 2020, Art. no. 851, doi: 10.3390/MATH8050851.

[41] A. Tubadji and H. Huang, "Emotion, cultural valuation of being human and AI services," *IEEE Trans. Eng. Manage.*, vol. 71, pp. 7257–7269, 2024, doi: 10.1109/TEM.2023.3246930.

[42] A. Revina, K. Buza, and V. G. Meister, "Designing explainable text classification pipelines: Insights from IT ticket complexity prediction case study," in *Interpretable Artificial Intelligence: A Perspective of Granular Computing*, W. Pedrycz and S.-M. Chen, Eds. Berlin, Germany: Springer, 2021, pp. 293–332, doi: 10.1007/978-3-030-64949-4_10.

[43] S. P. Paramesh, C. Ramya, and K. S. Shreedhara, "Classifying the unstructured IT service desk tickets using ensemble of classifiers," in *Proc. IEEE 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solutions*, 2018, pp. 221–227, doi: 10.1109/CSITSS.2018.8768734.

[44] M. Roopak, G. Y. Tian, and J. Chambers, "Multi-objective-based feature selection for DDoS attack detection in IoT networks," *IET Netw.*, vol. 9, no. 3, pp. 120–127, May 2020, doi: 10.1049/iet-net.2018.5206.

[45] P. Bedué and A. Fritzsche, "Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption," *J. Enterprise Inf. Manage.*, vol. 35, pp. 530–549, 2022, doi: 10.1108/JEIM-06-2020-0233.

[46] B. Avi-Itzhak, H. Levy, and D. Raz, "Quantifying fairness in queuing systems," *Probability Eng. Inf. Sci.*, vol. 22, no. 4, pp. 495–517, 2008, doi: 10.1017/S0269964808000302.

[47] D. T. Campbell and J. C. Stanley, "Experimental and quasi-experimental designs for research," 2015.

[48] S. Reig et al., "Not some random agent: Multi-person interaction with a personalizing service robot," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact.*, 2020, pp. 289–297, doi: 10.1145/3319502.3374795.

[49] W. Firestone and R. E. Herriott, "Multisite qualitative policy research: Optimizing description and generalizability," *Educ. Res.*, vol. 12, no. 2, pp. 14–19, 1983.

[50] A. Calabrese and F. Scoglio, "Reframing the past: A new approach in service quality assessment," *Total Qual. Manage. Bus. Excellence*, vol. 23, no. 11–12, pp. 1329–1343, 2012, doi: 10.1080/14783363.2012.733259.

Jeroen van Gassel received the MBA degree in business and IT from the Neyenrode Business University, Breukelen, The Netherlands, in 2017. He is currently working toward the doctoral degree in improving perceived customer service with the Delft University of Technology, Delft, The Netherlands.

He is currently a Product Manager of Process Mining, and he is exploring the integration of machine learning with perceived service quality enhancement with the Faculty of Policy and Management, Delft University of Technology. His research titled "Improving Perceived Service Quality with Machine Learning and the Operational Implications" investigates the transformative potential of machine learning in enhancing service quality and examines the operational shifts necessary to accommodate this technology. This research is critical in delineating the practical impacts of artificial intelligence technologies on service management, offering pivotal insights for industrial practices.

Marijn Janssen is currently a Full Professor with the ICT & Governance in the Technology, Policy and Management Faculty, Delft University of Technology, Delft, The Netherlands, the Head of the Engineering Systems & Servies Department, and a (honorary) Visiting Professor with the Bradford University, Bradford, U.K., the KU Leuven, Leuven, Belgium, and the Universiti Teknologi Mara, Shah Alam, Malaysia. He has authored/coauthored more than 600 refereed publications, his Googlescholar H-score is 92, having more than 36K citations. His research interests include governance of smart and open digital government, open data and emerging technologies, like AI, and quantum computing and blockchain, which fundamentally change the organizational landscape and influence the governance. He is particularly interested in situations in which multiple public and private organizations want to collaborate, in which ICT plays an enabling role, sociotechnical solutions are constrained by organizational realities and political wishes, and there are various ways to proceed.

Prof. Janssen is a Co-Editor-in-Chief of Government Information Quarterly, past-chair of the IFIP WG8.5 in ICT and public administration, conference chair of the IFIP EGOV–CeDEM–ePart series, and past-president of the Digital Government Society.