This is a peer-reviewed, author's accepted manuscript of the following conference output: (Eds.) (in press).
System and user strategies to repair conversational breakdowns of spoken dialogue systems: a scoping review.
In ACM Conversational User Interfaces 2024 (CUI '24) Association for Computing Machinery. Advance online publication.
https://doi.org/10.1145/3640794.3665558

System and User Strategies to Repair Conversational Breakdowns of Spoken Dialogue Systems: A Scoping Review

Essam Alghamdi
University of Strathclyde
Glasgow, UK

Martin Halvey
University of Strathclyde
Glasgow, UK

Emma Nicol
University of Strathclyde
Glasgow, UK

ABSTRACT

Spoken Dialogue Systems (SDSs) are critical in facilitating natural and efficient human-machine interaction through speech. SDSs frequently encounter challenges in managing complex dialogues, resulting in communication breakdowns, which include misunderstandings— where the system misunderstands user input— and non-understandings— where the system fails to interpret the input at all. Strategies to repair these breakdowns have been investigated across multiple disciplines; despite this interest, the findings from these studies are inconsistent and hinder comparative analysis due to the use of diverse methodologies and terminologies. To address this gap, this scoping review systematically examines SDS and user repair strategies within a broad spectrum of literature. Based on 36 papers out of 818 found, we provide two comprehensive frameworks: one categorising SDS system-repair strategies into six distinct categories and the other user-repair strategies into five categories. Our analysis reveals a disparity in the literature's focus on repair strategies, highlighting, in particular, the lack of research on less explored strategies, such as Information and Disclosure repair strategies, providing potential avenues for future research directions in this area.

CCS CONCEPTS

• **Human-centered computing** → *Natural language interfaces.*

KEYWORDS

Spoken Dialogue System, Conversational Agents, Conversational Breakdowns, Repair Strategies, Recovery Strategies

ACM Reference Format:
Essam Alghamdi, Martin Halvey, and Emma Nicol. 2024. System and User Strategies to Repair Conversational Breakdowns of Spoken Dialogue Systems: A Scoping Review. In *ACM Conversational User Interfaces 2024 (CUI '24), July 08–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3640794.3665558

1 INTRODUCTION

Spoken dialogue systems (SDSs) are a type of interface designed to enable humans to interact with machines naturally through spoken language [80]. In SDSs, voice is the preferred mode of communication with machines because it offers intuitive and hands-free interactions [12, 13, 42]. SDSs are now a fundamental part of our interactions with technology, profoundly impacting domains ranging from customer service to personal assistance. The emergence of SDSs in the form of digital homes, such as Amazon Echo, and personal assistants, such as Siri, has markedly elevated expectations regarding voice-based interfaces in everyday technology use. According to a 2021 report by Statista, the number of digital voice assistants is projected to exceed the global population by 2024, reaching 8.4 billion units [70].

Nevertheless, the widespread adoption of SDSs and their consistent and successful long-term use remain challenging [5]. Despite notable advancements, SDSs frequently encounter difficulties with complex dialogues and varied speech patterns; thus, expecting them to fully grasp every human utterance remains impractical. These challenges often result in conversational breakdowns and a variety of negative effects. Not only do these breakdowns significantly diminish user satisfaction and damage trust, they also decrease the likelihood of continued use [18, 44]. Furthermore, they often prevent users from completing tasks, leading to task abandonment [5], negatively impacting user experience and damaging the overall effectiveness of SDS performance [5, 18, 41, 44]. In SDSs, communication breakdowns, primarily misunderstandings and non-understandings, pose significant challenges [9]. Misunderstandings occur when the system incorrectly interprets user input, and are particularly challenging to detect and repair since the system believes it has understood the user correctly. Non-understandings, on the other hand, are recognised immediately by the system, which lacks any viable interpretation of the user's input [9] (See Figure 1). Given these challenges, particularly the prevalence of conversational breakdowns, some research has shifted towards developing effective repair strategies, which are crucial for enhancing SDS performance and user satisfaction [15].

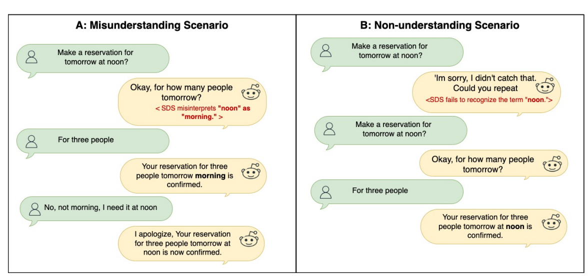

Figure 1: Comparison of Misunderstanding and Non-understanding Scenarios in SDS

CUI '24, July 08–10, 2024, Luxembourg, Luxembourg

ACM ISBN 979-8-4007-0511-3/24/07
https://doi.org/10.1145/3640794.3665558

Centred on the main dialogue actor in the breakdown-repairing process, research in repair strategies has been broadly divided into two areas: investigations into system-repair strategies and studies on user-repair strategies [10, 16]. System-repair strategies research explores how SDSs can be designed to effectively manage conversations, including mechanisms for grounding, interruptions, and corrections, and assesses their impact on user experience and SDS efficiency [17, 60]. In this scoping review, system-repair strategies refer to those strategies related to user interaction and experience rather than delving into the technical approaches of how these systems execute repairs. Purely technical repairs, which involve back-end algorithmic adjustments, data processing techniques, or system performance optimisations that occur without direct user interaction, fall outside the scope of this review. In contrast, user-repair strategies research aims to comprehend how users interact with and repair conversational breakdowns in SDSs across various domains and contexts, often highlighting adaptations in communication style to improve interactions [34, 45, 53, 54]. In this review, our focus is on strategies adopted by both SDS's main dialogue actors (system and user) to correct conversational breakdowns in spoken SDS.

The development and effectiveness of these repair strategies have been investigated across multiple disciplines, such as conversational analysis and business. A consequence of this can be inconsistent findings, which hinder comparative analysis because of the diverse methodologies and terminologies used. Importantly, the comparison and combination of different repair strategies have gained attention as researchers have discovered that employing multiple repair strategies is more effective than relying on a single repair strategy [37, 77]. Such diversity not only complicates the aggregation of findings but also prevents interdisciplinary collaboration, which negatively impacts the development of more effective repair strategies. Our objective is to present the first comprehensive scoping review of repair strategies in SDSs, with the aim of consolidating and analysing the various strategies employed by these systems and their users. We also aim to analyse the methodologies that are used for studying repair strategies in the SDS domain and defining the characteristics of SDSs. To achieve this, we have identified and categorised the diverse range of repair strategies developed and proposed for use in SDSs, drawing from a broad spectrum of literature. This includes both system-repair and user-repair strategies. This review contributes to Human-Computer Interaction (HCI) by providing frameworks for both system and user repair strategies, offering a standardised terminology and conceptual foundation that researchers can adopt to harmonise terminology across the literature. This study is guided by three research questions:

- RQ1: What types of user-facing strategies are employed by SDSs to repair breakdowns?
- RQ2: What strategies or tactics do users apply to repair breakdowns in SDSs?
- RQ3: What research methods have been utilised to investigate and evaluate repair strategies in SDSs, and in this context, what are the defining characteristics of SDSs?

The most relevant studies to our review were conducted by Feng [23] and Benner et al. [7]. Feng developed a taxonomy to categorise system-repair strategies in conversational agents (CAs), focusing on both theoretical and practical aspects. Their taxonomy includes five dimensions: acknowledgement, reasons, explanation, repair, and delegation to humans. Benner et al. [7] conducted a systematic review of strategies for handling conversational breakdowns in both text- and voice-based agents. They categorised system-repair strategies into six categories: confirmation, information, disclosure, social, solve, and ask. While these studies significantly contribute to the field, our scoping review adopts a broader approach by systematically examining both system-repair and user-repair strategies in SDSs and providing a comprehensive overview. In contrast to the scope of Feng and Benner, which includes text- and voice-based interactions, our review focuses on voice-based interactions. Additionally, we provide an overview of research methods and key characteristics of SDSs employed across relevant literature, aspects not addressed by previous studies. It is important to note that a detailed evaluation of the effectiveness of these strategies on user experience or SDS performance is outside the scope of our study.

2 METHOD

A scoping review enables us to gather a broad spectrum of literature, offer a comprehensive overview, and address the research questions. As [59] and [51] have suggested, this approach provides a broader perspective compared to delving into a detailed research agenda. Scoping reviews have become an increasingly popular and valid approach for synthesising and mapping research evidence [3, 20, 51]. This review follows both the PRISMA-ScR guidelines and the methodological framework proposed by [3] for scoping reviews. This framework has six iterative stages: 1) identifying the research question; 2) finding relevant studies; 3) selecting studies; 4) charting the data; 5) collecting, summarising, and reporting the results; and 6) an optional consultation exercise. Stage 6, regarding consultation work, is optional and outside the scope of this review.

2.1 Identifying relevant studies

2.1.1 Databases. Our cross-disciplinary search included four electronic databases: ACM Digital Library, IEEE Xplore, Scopus, and Web of Science. The ACM Digital Library and IEEE Xplore were chosen for their comprehensive coverage of computing and technology papers, making them essential for research in the HCI field [16, 69]. To capture a broader range of scientific domains, we incorporated Scopus and Web of Science into our search strategy. These were selected due to their extensive inclusion of multidisciplinary research and their recognised use as sources for literature reviews in the HCI community [4, 52, 71]. From Web of Science and Scopus, we retrieved papers spanning multiple domains beyond computing, including psychology (e.g., [42, 46]), linguistics (e.g., [38]), marketing and hospitality management (e.g., [29, 72]), as well as gerontology and cognitive science (e.g., [47, 74]). This enriched our review with diverse perspectives on the study of breakdowns and repair strategies in HCI.

2.1.2 Search query. The next step was a keyword search within the selected databases to identify papers. The literature uses numerous terms inconsistently to describe dialogue systems operating on different devices, such as 'personal digital assistant', 'conversational agent', 'voice chatbot' and 'conversational user interface' [65]. For the current review, the list of terms was derived directly from similar papers that performed systematic reviews on speech-based and conversational interfaces [7, 16, 52]. Two main concepts needed to be elucidated in this scoping review: SDSs and repair strategies (see Table 1)

Table 1: General Query Structure for All Searches

Concepts	Terms
SDSs	["spoken dialog system" OR "speech interface" OR "voice user interface" OR "voice system" OR "speech-based" OR "voice-based" OR "speech-mediated" OR "voice-mediated" OR "human computer dialog" OR "human machine dialog" OR "natural language dialog system" OR "natural language interface" OR "conversational interface" OR "conversational agent" OR "conversational system" OR "conversational dialog system" OR "automated dialog system" OR "interactive voice response system" OR "spoken human machine" OR "intelligent personal assistant"] AND
Repair strategies	["repair strateg*" OR "mitigation strateg*" OR "error-handling strateg*" OR "dialogue confirmation strateg*" OR "handling breakdown*" OR "conversational breakdown*" OR "handling error*" OR "mitigating error*" OR "recognition error*" OR "system error*" OR "error recovery" OR "error handling" OR "Correction type*" OR "failure*" OR "correction*" OR "error detection" OR "error recovery" OR "miscommunication" OR "erroneous interpretation" OR "erroneous situation*" OR "correction mechanism*" OR "recovery mechanism*"]

Wildcards were used to simplify the search string tailored to each database's specific features. Employing these keywords, a comprehensive search across four databases was conducted, targeting all peer-reviewed scholarly articles and full conference papers published up to August 2023. To capture the entire spectrum of relevant research, no lower time limit was set, allowing for a thorough inclusion of foundational and contemporary studies. Keywords were used to search titles and abstracts, yielding 818 papers. These were imported into Endnote 20.6 as a single library, including basic metadata e.g. title, author, year, and conference/journal details. Duplicate papers were removed, resulting in 505 papers from the initial database search.

2.2 Study selection

In systematic review methods, inclusion and exclusion criteria based on a specific research question are developed to ensure consistency in decision-making [3]. For this scoping review, a series of inclusion and exclusion criteria were developed through multiple rounds of discussions between authors.

Due to the wide variety and modalities of dialogue systems, the scope of this review was narrowed to voice interactions and excluded papers dealing with modalities such as text, gesture, eye-tracking and multi-modal interaction (where voice was not a primary modality). As this review focused on the user-interaction perspective, studies purely concerned with technical strategies for handling errors were also excluded. The finalised formal criteria are as follows:

2.2.1 *Inclusion criteria.*

(1) User studies examining strategies to repair or mitigate breakdowns in SDSs.
(2) Research involving solely or primarily voice-based user interactions (e.g., voice assistants, smart speakers).
(3) Studies involving at least one user and one SDS, focusing on user interaction rather than system performance or machine-machine interaction.
(4) Original research in peer-reviewed journals and full-length conference papers.

2.2.2 *Exclusion criteria.*

(1) Papers not specifically investigating repair strategies in SDSs.
(2) Studies on multi-modal interactions where voice is not the main modality.
(3) Dissertations, review articles, conference abstracts.
(4) Publications not in the English language

To ensure the relevance of studies to our research questions, we applied inclusion and exclusion criteria in a two-step process. Initially, the titles and abstracts of 505 papers obtained from a database search were screened using these criteria. This led to the identification of 128 papers meeting the inclusion criteria. In cases where a paper's relevance was unclear from its title and abstract, we conducted a full-text review to further assess eligibility. Ambiguities about paper eligibility were resolved through discussions among authors, resulting in 17 papers considered eligible from the database search. Additionally, we employed manual searches to capture relevant studies that might not have included key search terms in their titles or abstracts [3, 7]. This involved both backward chaining (examining the bibliographies of identified studies) and forward chaining (using Google Scholar to check citations), together yielding an additional 19 papers. This combined method of electronic and manual searches identified a total of 36 papers (see Figure2).

3 RESULTS

The findings are organised along the lines of the research questions into three sections.

3.1 The characteristics of included studies and SDSs

3.1.1 *Publication Overview.* The publication dates of the reviewed papers ranged from 2002 [66] to 2023 [2, 17, 29, 75]. We found no studies in the period 2010-2017. Approximately one or two studies were published per year before 2011, and this number increased to five in 2023, indicating an increase in awareness among researchers about the importance of handling breakdowns in SDSs (See Figure 3 for the number of papers by publication year). Whereas the initial contributions in this field were more exploratory in nature, covering

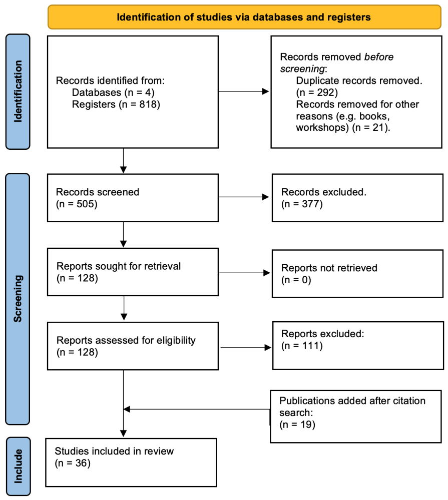

Figure 2: PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flowchart.

a range of spoken dialogue breakdowns, more recent papers have tended to be more specific, focusing on particular system-repair and user-repair strategies within more defined contexts and settings. For example, around 75% (8 out of 11) of the studies published before 2011 primarily focused on analysing audio corpora from real-world environments in order to explore system-repair and user-repair strategies in communication breakdowns in SDSs (e.g. [66]). Conversely, recent papers (e.g.[17]) delved into specific topics, such as investigating the impacts of four distinct humour styles (employed as system-repair strategies by voice assistants) on user frustration during communication breakdowns. Table 2 lists 32 unique publication venues identified from the reviewed studies.

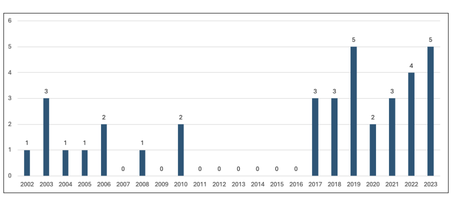

Figure 3: Distribution of Relevant Literature According to Year

The studies reviewed were conducted in various countries and languages. Some studies explicitly mentioned both country and language (N=12), while others specified only country (N=7) or only language (N=8). There were also instances in which neither the country nor language was mentioned (N=9). English was the most prevalent language (N=16), followed by German (N=4). The United States was the most common location (N=16), followed by Germany (N=4). A few studies were carried out in other languages and regions, including China [29, 75], Sweden [22, 67], UK [60], Brazil [50], Denmark [17], South Korea [40] and Taiwan [42]. Almost all studies included in the review were conducted in a single language, except one [38], which was a multi-language study encompassing nine languages.

3.1.2 Methodological Overview. Given the diverse and multifaceted nature of the included studies, we organised our findings by broad methodological approach, then by specific technique. This allowed us to illustrate not only the variety of methodological approaches but also how these interrelate and contribute to the field of SDSs. The papers were divided into three broader approaches: qualitative, quantitative and mixed-method. To ensure accurate categorisation, we initially searched for explicit labels—qualitative, quantitative, or mixed-method—within each paper. Where these labels were not specified, we analysed the research designs and methods described to determine the appropriate category. Studies employing in-depth interviews or observational analyses were classified as qualitative, those using statistical tests or numerical data analysis were classified as quantitative, and studies indicating substantial use of both approaches were classified as mixed-method. Within each approach, we discuss specific research designs and methods of each study. This includes the data source of corpora (lab- and field-collected corpora), experimental designs (between- and within-subjects experiments) and specific research methods (e.g. interviews, questionnaires and Wizard of Oz method (WoZ)).

The majority of the studies in our review adopted a mixed-method approach (N=16), followed by qualitative (N=10) and quantitative (N=10) approaches. Among these studies, over half conducted analyses of audio corpora, which were categorised into two main types: field- and lab-collected corpora. Field-collected corpora (N=11) are derived from real-world sources, while lab-collected corpora (N=10) are generated and analysed by the authors of the study itself. Most studies utilising field-collected corpora were published prior to 2008 and predominantly focused on telephone platforms, such as the DARPA Communicator Dialogue Travel Planning System [11, 27, 32, 66], the RoomLine conference reservation system [63], the Pizza Corpus ordering system [15], the Let's Go bus schedule information system [63] and the TOOT train information system [43]. Apart from corpus analysis, some of the reviewed studies employed experimental research designs, including between-subject experiments (N=10) and within-subject experiments (N=10).

A range of research techniques was employed for data collection in both qualitative and quantitative studies. These methods predominantly include interviews, questionnaires and Wizard of Oz techniques [26, 34, 42, 47, 67, 78]. A number of studies (N=7) favoured semi-structured interviews [19, 34, 42, 50, 53, 77, 78] while others chose structured interviews [2] or did not specify interview format [6, 67]. Questionnaires were also popular, with 13 studies employing them, predominantly using a 7-point Likert Scale [2, 17, 19, 26, 29, 67, 72, 79], though some used a 5-point scale [33, 34, 40, 78] and others varied in the type of scale used [22, 42, 46]. To visualise the prevalence of the most common techniques across the three broad methodological approaches, a heat map is provided in Figure 4. This heat map reveals that corpus analysis is the predominant technique in qualitative studies, while a combination of questionnaires and interviews is typically used in experimental mixed-method approaches. This is followed by questionnaires as the typical technique for experimental quantitative studies.

Table 2: Publication Venues

Publication Type	Names	Frequency
Academic Journals	International Journal of Human-Computer Interaction	3
	Speech Communication	2
	In INTERSPEECH	2
	International Journal of Hospitality and Tourism Administration	1
	The Service Industries Journal	1
	Universal Access in the Information Society	1
	Recent trends in discourse and dialogue	1
	Frontiers in Psychology	1
	Association in Computer Science	1
	Elektronische Sprachsignalverarbeitung	1
	In Proceedings of SST	1
	Computational linguistics	1
	International Ergonomics Association	1
Conference Proceedings	CHI conference on human factors in computing systems	4
	Proceedings of the ACM on Human-Computer Interaction	1
	ACM conference on interaction design and children	1
	HCI International Conference	1
	Conversational Interruptions in Human-Agent Interactions	1
	ISCA tutorial and research workshop on error handling in spoken dialogue systems	1
	Conversational User Interfaces	1
	International Conference on Automotive User Interfaces and Interactive Vehicular Applications	1
	International Conference on Human-Agent Interaction	1
	Conference on Human Information Interaction and Retrieval	1
	IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)	1
	Conference of the International Speech Communication Association	1
	International Conference on Spoken Language Processing	1
	Conversational User Interfaces	1
	Conference on Conversational User Interfaces	1
	IEEE Workshop on Automatic Speech Recognition and Understanding	1

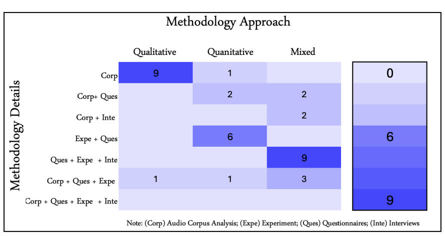

Figure 4: Heat Map of Methodology Details for the Selected Studies

Moreover, the contextual and environmental settings of the reviewed studies varied, with a majority conducted in real-world environments (N=12) and laboratory settings (N=10), followed by controlled field settings (N=5), simulated environments (N=4) and online or crowdsourced settings (N=4). Approximately half of the SDSs (N=16) employed in the reviewed studies were mock-ups in which participants interacted with the SDSs through Wizard of Oz (WoZ) setups(e.g. [17, 77]). These were mostly used in mixed-methods studies (N=10), followed by quantitative studies (N=5) and one qualitative study (N=1).

Sample size ranged from as few as five participants in a qualitative study in a real-world setting [60] to as many as 850 participants in an online multi-language quantitative study [38]. Although most study participants were drawn from the general population, four studies specifically included participants with computer science backgrounds [17, 22, 53, 79]. In studies reporting gender information (N=30), there was a notable gender disparity, with males outnumbering females by approximately 20%. Finally, participants' ages varied considerably, from as young as three years old [14] to as old as 92 years old [57].

3.1.3 SDS Characteristics Overview. In this sub-section, we review the key characteristics of SDSs, to provide a comprehensive overview of their features and interactions. Our literature review revealed a heterogeneous landscape of domains and platforms for SDSs. We categorised SDS into six distinct domains and seven platforms based on their application contexts, functionalities, and task types in the literature. We identified recurring themes and interaction types across studies, organising them into categories reflecting their primary functions and user interactions. Each domain corresponds to settings with specific interaction types and tasks. For example, the personal assistance domain includes SDS applications that aid tasks such as scheduling and reminders, featuring personalised daily interactions. We refined these categories iteratively to accurately reflect the diverse aspects of SDS applications observed in the included studies.

Each study in our review falls into one of the following domains: customer service, personal assistance, public assistance, gaming, automotive and smart home assistance. The platforms identified, including computer speakers, smart speakers, mobile devices, smart home devices, in-vehicle voice user interfaces (VUIs), telephones, and robots, were determined via a detailed analysis of systems utilised across the studies reviewed. Most studies were related to customer service (N=10) and computer speakers (N=10). Certain domains are aligned with particular platform types, indicating specialised applications e.g. the smart home assistance domain was associated with devices such as Amazon Alexa [6, 17, 19, 33, 46, 60], the INSPIRE system [74], and devices not specified [38]. Similarly, the automotive domain was uniquely paired with in-vehicle VUIs in simulated environments [2, 34], and field settings [75].

Some studies involved more complex combinations of platforms presented in other domains, indicating a broader range of functionalities and user interactions. Studies within the public assistance domain utilised smart speakers [36] or robots [42] in their SDSs. By contrast, studies within the customer service domain utilised telephones [9, 11, 15, 27, 32, 66] or computer speakers [29, 38]. Further emphasising the diversity of platforms, studies within each domain of the personal assistance and gaming categories feature the integration of at least three distinct types of platform. Specifically, in the personal assistance domain, individual studies focused on utilising either computer speakers [53, 54, 57, 67, 79], mobile devices [50] or smart speakers [26, 40]. In the gaming domain, studies employed either computer speakers [35, 78], mobile devices [14]or robots [22] in their SDS designs.

Other dimensions providing further insight into the context and interaction of SDSs included interactional sessions, language, interaction types, interface personification, motivation, and collaboration goals. The majority of studies (N=30) examined interactions within a single session in which participants engaged in one conversation with an SDS. A smaller number (N=5) explored multi-session interactions ranging from 1–5 weeks, predominantly in real-world environments with families, (e.g. [14, 46]). One study [77], which did not involve direct interaction with SDSs, investigated user classification of error messages and preferences for error responses using categorisation and sorting surveys. All but one study employed single-language SDSs, the exception being [38], which implemented multilingual SDSs across various regions to study differences in user responses to system breakdowns in different linguistic contexts. Most studies focused on single-individual interactions with SDS (N=32), while four studies explored interactions involving multiple people with SDSs [6, 14, 46, 60]. Regarding interface personification, which refers to the degree to which SDSs exhibit visual or physical human-like features [31], most studies (N=29) used disembodied SDSs. A few studies used embodied SDSs, employing animated systems [14, 78], avatars [35, 57, 72], and robots [22], mainly in the gaming domain, with one study each in the personal assistant and customer service domains.

Regarding collaboration goals, SDSs were either goal-oriented (N=27), aiding users in achieving specific objectives, or non-goal-oriented (N=9), focusing on broader engagement without a defined end goal. Participant motivation for engaging with SDSs was broadly classified into two categories, depending on the task at hand: assistance/facilitation (N=25) and entertainment/engagement (N=11). Roles of SDSs were primarily as facilitators (N=32), with only three studies [22, 35, 78], examining SDSs as peers, all within the gaming domain. Details of methodologies and repair strategies are presented in the literature matrix (See Table 3).

3.2 System-Repair Strategies

To categroise repair strategies in SDSs, we began by thoroughly examining system and user responses to SDS breakdowns reported in the included studies to identify potential categories. Where applicable, we adapted existing categories from frameworks presented in previous research(e.g, [7, 9, 50, 53]) to maintain consistency and address any gaps observed. Importantly, the criteria for our categorisation were based on the function and objective of each response within SDS interactions. This approach led to the merging of categories where similar objectives were pursued by the repair strategies. For example, under the category of "Modulation strategies" which are strategies that aim to resolve communication breakdowns by altering vocal characteristics, we merged three primary user-repair strategies: adjusting speech rate, enunciating clearly, and modifying speech volume. Each strategy shares the common goal of enhancing speech intelligibility and effectiveness in response to a breakdown. A consensus among authors was reached through discussion, ensuring a balanced and objective approach to the classification.

Following this methodological framework, to specifically classify system-repair strategies for SDSs, we adopted a framework derived from the systematic analysis by [7], which analysed recovery strategies in conversational agents. This framework is particularly relevant for our review as it offers well-defined categories for systematically categorising repair strategies in both text and voice interactions, which is crucial for our focus on voice interactions in SDSs. We adopted this framework to categorise system-repair strategies into six distinct categories: *confirmation, information, solve, social, ask, and disclosure.*

3.2.1 Confirmation. In this review, confirmation strategies in SDSs are not designed to completely repair a failed conversation but instead are intended to acknowledge or ignore errors. Regarding error acknowledgement, SDSs utilize both explicit and implicit strategies for verification or rejection. Explicit error acknowledgement involves the SDS admitting its misunderstanding with phrases like 'I don't understand,' which prompts users to guide the recovery process. In contrast, implicit error acknowledgement involves the SDS either staying silent or responding in ways that reflect the miscommunication. Notably, explicit acknowledgements often serve as precursors to further repair strategies, such as requesting a rephrasing or clarification of the user's utterance, details of which are explored later in this section.

Table 3: A Literature Matrix of Repair Strategies in SDSs

Authors/Year	Repair Strategies											Research Methodologies									
	SDSs						Users					Approach			Corpus		Experiments		Methods		
	Conf.	Info.	Social	Solve	Ask	Disc.	Clar.	Info.	Modu.	Syst.	Adap.	Qual.	Quan.	Mixed	Filed	Lap	Between	Within	Ques.	Interview	WoZ
Huang and Sénécal [2023][29]			•										•				•		•		
Clausen et al. [2023][17]			•										•				•		•		•
Xu et al. [2023][75]		•	•										•					•	•		•
Meck et al. [2023][47]	•				•									•				•	•	•	•
Wang et al. [2023][72]			•										•				•		•		
Mavrina et al. [2022][46]							•		•				•			•			•		
Kisser and Siegert [2022][36]	•	•					•		•			•			•						
Zargham et al. [2022][78]	•				•									•			•		•	•	•
Motta and Quaresma [2021][50]				•			•	•		•	•			•				•		•	
Cuadra et al. [2021][19]	•													•			•		•	•	
Lin et al. [2021][42]	•			•	•									•			•		•	•	•
Myers et al. [2021][54]							•	•		•		•				•					
Yuan et al. [2020][77]	•		•											•			•		•	•	
Kim et al. [2020][35]							•	•						•		•		•	•		•
Ge et al. [2019][26]			•											•				•	•	•	•
Beneteau et al. [2019][6]	•						•		•					•		•				•	
Lee et al. [2019][40]		•			•								•					•	•		•
Kim et al. [2019][34]				•	•		•	•	•					•				•	•	•	•
Kiesel et al. [2019][33]	•		•		•								•			•			•		
Porcheronet al. [2018][60]							•	•			•	•				•					
Cheng et al. [2018][14]							•				•	•				•					
Myers et al. [2018][53]							•	•	•	•	•			•		•				•	
Engelhardt et al. [2017][22]	•													•	•				•		•
Opfermann and Pitsch [2017][57]							•	•				•			•			•			•
Kraljevski and Hirschfeld [2017][38]									•	•			•			•		•			•
Wolters et al. [2010][74]				•										•		•			•		•
Zgorzelski et al. [2010][79]	•			•	•								•				•		•		•
Bohus and Rudnicky [2008][9]	•			•	•									•	•		•		•		•
Litman et al. [2006][43]							•	•	•			•			•						
Raux et al. [2006][63]	•			•	•								•		•						
Bulyko et al. [2005][11]			•		•		•			•				•	•			•	•		
Choularto et al., (2004)[15]							•	•				•			•						
Gabriel Skantze [2003][67]	•				•									•			•		•	•	•
Goldberg et al. [2003][27]	•		•		•		•			•	•	•			•						
Kazemza et al. [2003][32]							•	•		•	•	•			•						
Shin et al. [2002][66]	•			•	•		•	•		•	•	•			•						
Sum (N=36)	15	3	9	8	13	0	18	11	8	8	7	10	10	16	11	10	10	10	21	11	16

Note:(Conf) Confirmation; (Info) Information; (Disc)Disclosure; (Clar) Clarification; (Modu) Modulation; (Syst) System-centric; (Adap) User Adaptation; (Qual) Qualitative; (Quan) Quantitative; (Ques) Questionnaire

The error ignoration strategy, on the other hand, sees SDSs overlooking errors and continuing along a predefined conversational path. This strategy may involve posing new questions [2, 9, 67, 79] or taking steps towards goal completion [22, 33, 36, 67, 78], even in the face of misunderstandings. For example, [22] found that employing the ignore repair strategy, where a robot overlooks its failure and proceeds with a simple 'OK', resulted in enhanced perceptions of perceived intelligence and animacy (as in an animated system) compared to alternative strategies such as apologising and problem-solving.

3.2.2 *Information.*

The information strategy transcends basic confirmation, as SDSs strive not only to confirm but also to elucidate the situation by providing potentially useful messages or feedback about the error. Such a recovery strategy appeared in only three of the reviewed studies [36, 40, 75]. For example,[75] designed a multiple-recovery strategy consisting of an explanation of an automated driving system failure that aimed to mitigate the adverse effects of erroneous situations. Similarly, [40] employed a repair strategy known as 'elaborated feedback' to develop common ground (e.g. 'I don't understand what you said. It is noisy around here.'), which positively influenced user acceptance and usability of VUIs.

3.2.3 *Social.*

The social recovery strategy in SDSs integrates human-like behaviours to address and repair breakdowns in communication. Grounded in the Computers Are Social Actors (CASA) paradigm, these strategies embody a range of human-like qualities through various responses, including apologies and compensatory strategies [55]. For instance,[26] explored user preferences between two different strategies—apology ('I'm sorry') and humour ('My IQ is still recharging') —across distinct failure contexts. The results indicated a pronounced preference for smart speakers that offered apologies in both instances. On the other hand, humour as a repair strategy proved counterproductive, especially when smart speakers misunderstood user requests.

3.2.4 Solve. The solve strategy aims to actively resolve breakdowns by offering concrete solutions. This strategy is goal-oriented and distinct from that of the information strategy, as it surpasses the mere elucidation of breakdown causes. Instead, it encompasses the delivery of supplementary information and guidance tailored to help users effectively resolve breakdowns. SDSs can augment their support by providing users with a range of potential solutions in the form of options. Additionally, SDSs can introduce pre-defined speech structures or templates designed to assist users in their recovery efforts. For example, in instances where users use language or expressions beyond the SDS's processing capabilities, the SDS presents a set of understandable utterances to ensure mutual understanding between user and system.

Based on a study of conference room reservations, [9] assessed the effectiveness of ten strategies in addressing errors of non-understanding. Among these, 'Full Help' and 'TerseYouCanSay,' strategies where the system guides the user on what to say next, emerged as the most effective. When the system offers help, which includes sample responses related to the conversation (e.g. SDS: 'Have you had lunch yet? You can answer whether or not you have eaten.'), users can discover more effective ways (from the system's viewpoint) to express their intent and explore further dialogue options [42].

3.2.5 Ask. The ask strategy in SDSs transfers the repair responsibility to the user via three main techniques. Initially, the SDS may repeat the inquiry to give the user another chance to articulate their request [9, 11, 27, 34, 40, 42, 63, 79]. Alternatively, the SDS could ask the user to rephrase their utterance [9, 11, 27, 63]. Lastly, if the first attempt was unclear or incomplete, the SDS might request additional input to formulate an adequate response [2, 9, 22, 34, 63, 67, 78]. For instance, [11] examined the NIST 2000 Communication Evaluation, which included nine mixed-initiative telephone dialogue systems. They found that employing repetitions as a system-repair strategy resulted in significantly higher rates of frustration compared to rephrasing.

3.2.6 Disclosure. Similar to the information strategy, the disclosure strategy is designed to educate users on how to navigate potential communication issues. The key distinction is that disclosure aims not to directly address or repair breakdowns but to set realistic user expectations by requiring the SDS to identify itself as a computational entity and to openly state its capabilities and limitations. This approach helps manage user expectations by clarifying what the system can and cannot do, thereby potentially preventing some misunderstandings before they occur. Our dataset reveals an absence of studies focusing on the broader implications of such disclosure strategies in SDSs, echoing the findings of previous reviews and emphasising the need for more comprehensive research in this domain [7]. For a depiction of system repair strategies following a dialogue breakdown, see Figure 5.

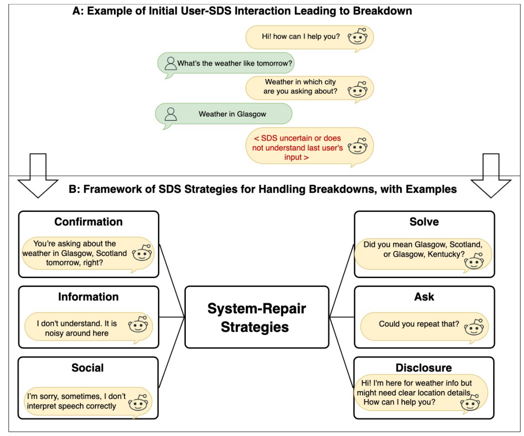

Figure 5: This is a flowchart depicting an initial interaction between a user and SDS that results in a communication breakdown (A). It is followed by a comprehensive framework of system-repair strategies (B). The framework includes examples of each strategy in action, demonstrating how the SDS can repair communication breakdowns.

3.3 User-Repair Strategies

Based on an aggregation of evidence from the included studies, we propose a framework of user-repair strategies when users interact with SDSs. We developed our framework by synthesising and adapting categories from the frameworks presented in the studies we reviewed, to address the need for a more comprehensive framework. Because of the complexities and unproductiveness of user reactions to SDS breakdowns, user-repair strategies can be categorised into five broad strategies: *clarification, information adjustment, modulation, system-centric and user adaptation.*

3.3.1 Clarification. Clarification strategies play an important role in user-repair in SDSs, focusing on ensuring clarity and correctness of the intended message. These strategies are the most widely employed by users, with repetition and reformulation being the primary forms. Repetition is a direct and common strategy, especially when users feel that the SDS has misunderstood their input. Users often repeat their previous utterances, hoping for better system comprehension (e.g. [6, 11, 66]). For example, [6] investigated family interactions with Alexa and found that children frequently use repetition as their initial attempt to repair communication breakdowns.

Reformulation involves a range of techniques to enhance communication when initial attempts are unsuccessful. Central to this strategy is rephrasing. Users alter the original utterance to make it more comprehensible to the SDS while maintaining their original intent [6]. This can include simplifying sentences, using synonyms or restructuring complex requests into simpler formats [6]. Users often rephrase by replacing verbs or nouns or by modifying sentence structure [35, 58]. Four studies we reviewed investigated user-repair strategies specifically focused on lexical adjustments (when users modify vocabulary), semantic adjustments (when users alter the meaning of the original statement) and syntactic adjustments (when users restructure sentences; e.g. [30, 46]).

3.3.2 Information Adjustment. Information adjustment strategies modify the quantity or precision of the information provided. This category encompasses three strategies: addition, subtraction and correction. Addition occurs when users provide additional details to repair a breakdown, particularly when the initial utterance lacks sufficient information for the SDS to comprehend the user's intent accurately [50, 54]. For example, a study on the sequence patterns of users interacting with VUIs [54] found that users often add more details to their utterances to clarify their intent when they encounter a breakdown with a new task. Subtraction strategies involve users simplifying their utterances by eliminating confusing or unnecessary details. Similarly, [50] examined how different types of voice assistants (e.g. Siri and Google Assistant) relate to user-repair strategies and found that users modify their utterances by adding or subtracting elements during breakdowns, indicating exploratory behaviours.

Correction strategies occur when users directly repair specific inaccuracies interpreted by the system. This user-repair strategy is particularly relevant when the SDS misinterprets part of the user input, leading to unsatisfactory responses (e.g. [9, 66]). For instance, if a user requests 'Play relaxing music', but the SDS plays rock music instead, the user may correct this by stating, 'No, play relaxing music.' Unlike addition, which provides extra context, or subtraction, which simplifies the input, correction specifically targets inaccuracies in the SDS's understanding. Some studies (e.g. [32, 66]) show that correction often manifests as a contradiction, wherein the user interrupts or 'barges in' to correct and guide the SDS's incorrect process. For example,[32] analysed the acoustic features of user-repair strategies and found that contradictions, as a form of correction user strategies, were associated with higher energy and frequency compared to other responses, such as repetition.

3.3.3 Modulation. Modulation strategies, which aim to resolve communication breakdowns by altering vocal characteristics, encompass three primary user-repair strategies: adjusting speech rate, enunciating clearly and modifying speech volume. Recent studies, particularly those produced in the past five years, have shown increased interest in exploring modulation strategies across various domains and settings, primarily through the analysis of lab-collected corpora (e.g. [35, 46, 53]). These studies found that in correction and non-correction dialogue acts, the prosodic elements of user utterances typically shift towards hyper-articulated speech patterns. For example, [35] investigated frequent breakdowns and user reactions to them in voice-based dialogue interfaces. This study revealed that attempts of many participants to amend errors by repeating phrases more slowly or with clearer enunciations often resulted in higher rates of error recognition by the SDS.

3.3.4 System-Centric. System-centric strategies are tactics employed by users that reflect their understanding of the capabilities and limitations of the SDS. In response to breakdowns, users often emphasise specific keywords that activate certain functions within the SDS. Utilising existing telephone dialogue systems, [11] explored how users react to re-prompting – when the SDS repeatedly asks for the same information. They found that commands such as 'Back up', 'Start over' and 'Scratch that' are processed more efficiently by the SDS than would be the case by simply repeating phrases. Another system-centric strategy is interface switching. When an SDS supports alternative input modes, users may opt to switch to typing [50] or using a touchscreen [53] to overcome communication breakdowns.

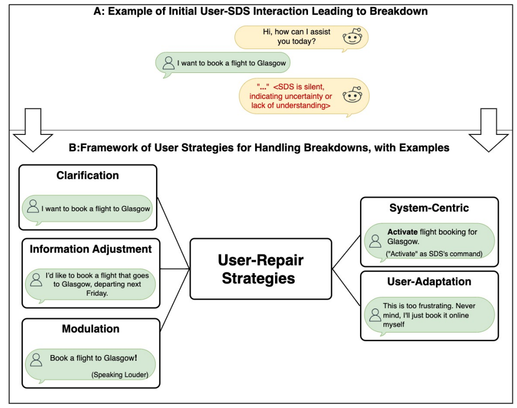

Figure 6: This flowchart highlights user-driven repair strategies in response to SDS communication breakdowns (A). It is followed by a comprehensive framework of use-repair strategies (B). The framework includes examples of each strategy in action, demonstrating how the user can repair communication breakdowns.

3.3.5 User Adaptation. User adaptation strategies encompass a broad range of repair strategies that extend beyond the four previously described categories and comprise a diverse and complex class of user behaviours. Within this category, three primary user-repair strategies for SDS breakdowns can be identified: contextual adjustment, emotional expression and decision to disengage. Contextual adjustment strategies involve alterations to the user's approach or surroundings to enhance interaction with the SDS and repair breakdowns. For example, some studies (e.g. [14, 60]) highlighted how users proactively adjust their environment, such as a child repositioning a tablet following parental advice or participants minimising ambient noise to aid voice recognition – actions that reflect an instinctive adaptation to improve communication with the system.

Emotional expression strategies encompass users' verbal and non-verbal reactions to SDS breakdowns, often reflecting frustration [53]. This can range from swearing, vocal interjections like 'ah' and laughter, to non-verbal cues such as eye-rolling [32, 53, 57]. Finally, decisions to disengage are characterised by users stepping back from interaction with SDSs, which may involve quitting, expressing a desire to stop or ending a call [30, 50, 53, 66]. These behaviours signal a user's intention to cease efforts to overcome breakdowns, which is frequently associated with confusion and frustration, leading to task abandonment. [53]. The literature underscores the importance of these strategies as they mark the moments when users choose to stop interacting with the task. For a depiction of user-repair strategies following dialogue breakdown, see Figure 6.

4 DISCUSSION AND FUTURE DIRECTIONS

4.1 RQ1 and RQ2: System and User Repair Strategies in SDSs

4.1.1 Frameworks of repair strategies . This review provided an overview of how SDSs and their users repair communication breakdowns. From the 36 papers included, we provide frameworks for both SDS and user repair strategies. For system-repair strategies, the framework categorises system-repair strategies into six categories: confirmation, information, social, solve, ask and disclosure. This comprehensive categorisation directly addresses RQ1 by identifying the specific types of strategies SDSs employ to repair conversation breakdowns. The majority of system-repair studies focused on confirmation, asking, and social strategies. However, the review identified only a limited number of studies (N=3) focusing on information repair strategies. Importantly, there is a complete absence of studies investigating the disclosure strategy as a means to repair communication breakdowns within SDS. The disclosure strategy, which involves the SDS openly acknowledging its limitations and capabilities to users, could be crucial for enhancing user interaction by setting realistic expectations. Evidence suggests that a conversational agent's self-disclosure as a computer artefact can efficiently set user expectations by acknowledging its non-human intelligence [64].

Similarly, our review answers RQ2 through the development of a framework for user-repair strategies, divided into five distinct categories: clarification, information, modulation, system-centric, and user adaptation. This framework highlights the active role users play in the repair process by employing various strategies to overcome communication breakdowns. In our review of studies on user-repair strategies, we observed a balanced distribution across all strategy types. However, while clarification and information strategies were more common, there was a notable gap in the exploration of modulation, system-centric, and user adaptation strategies. Expanding research to cover these under-explored strategies will provide a deeper understanding of strategies for addressing breakdowns in SDSs, providing a road map for future studies.

Furthermore, the literature in the review used inconsistent terminologies for repair strategies. For example, the 'error ignoration' system-repair strategy, classified under the confirmation category, involves an SDS ignoring a breakdown and continuing along a predetermined path. This strategy is named differently in the reviewed studies, though they refer to the same concept and mechanism of repair methods such as 'ignore' [22], 'task-related question' [67], and 'anticipatory error handling' [78]. Thus, for future work, adopting the frameworks proposed in this review will promote a standardised terminology and conceptual foundation within the research community and provide uniform references to repair strategies.

4.1.2 Interaction of repair strategies . The reviewed studies have demonstrated that researchers often employ a combination of system-repair strategies to repair complex communication breakdowns and improve user experience. Research shows the effectiveness of system-repair strategies is stronger when combined with other strategies as a multiple-repair strategy [37, 77]. For instance, [75] employed a multi-repair strategy combining social and information repair strategies. The system used an integrated prompt: "I am very sorry [apology]. There is a fault and a take-over is needed due to a possible out-of-memory error [explanation]. We will ensure this does not recur [promise]."(p. 5). Similarly, [27] found that an apology as a social system-repair strategy led to lower word error rates and lower frustration when combined with rephrasing as a confirmation strategy. This combination of system-repair strategies highlights the complexity of their interplay. Moreover, the integration of Large Language Models (LLMs) such as ChatGPT-voice and Alexa-LLM offers fertile ground for exploration. LLMs, known for their robust natural language processing capabilities, could significantly refine the adaptability and effectiveness of both system and user repair strategies [73, 76]. Future research could explore how these models can be customized to enhance interaction dynamics within SDSs, particularly in terms of real-time responsiveness and contextual understanding. This suggests the potential value of future studies considering the frameworks identified in our scoping review to identify and articulate these complex combinations more precisely.

Regarding user-repair strategies, certain strategies present more challenges to SDSs than others, with users frequently adopting clarification strategies, such as repetition [21, 66] and modulation strategies, such as changes in voice pitch or volume [24, 28, 46]. However, these strategies are often the least successful at repairing breakdowns and can negatively impact the SDS's ability to recognise voice commands accurately [21, 24, 27]. Importantly, the selection of user-repair strategies is not random but is significantly influenced by the system's repair strategies [43, 50, 60, 61]. This indicates a complex interaction between user behaviour and system feedback, where system prompts directly shape the user's choice of repair strategy. This dynamic interaction has been explored in only a few studies (N=8) within our review (e.g., [50]). For future research, it will be essential to investigate the development of SDSs that guide users away from less successful user-repair strategies and towards more effective ones.

In conducting this review, we aimed to map the landscape of strategies employed, rather than to evaluate their effectiveness on user experience or SDS performance. However, in the course of our review, we observed notable variations in how similar strategies were evaluated across studies, underscoring the diversity of research methods and contexts. For example, while some studies (e.g., [72, 75, 77]) suggest that the effectiveness of apologies employed as social repair strategies may be limited, others (e.g., [22, 26, 27]) report significant positive impact on users' perceptions of SDS. This divergence hints at the complex ways in which users interpret apologies from SDS, potentially as a standard response to failures [77]. Such observations reveal the nuanced and context-dependent effectiveness of repair strategies [75], suggesting fertile ground for future research. Further examination via systematic review or meta-analyses, could provide crucial insights into conditions influencing strategy effectiveness.

4.2 RQ3: Research Methods and Defining Characteristics of SDSs

Here, we provide an overview of research methods employed and key characteristics of SDSs in the papers reviewed, addressing RQ3. As the papers showcase a diversity of research methods, we classify methodologies into three broad approaches: qualitative, quantitative, and mixed-methods. Each employs distinct techniques and data collection methods. Our findings indicate that a significant number of qualitative studies (N=9) exclusively utilised corpus analysis, demonstrating a well-established foundation of research employing this method. However, it is important to note that the majority of these studies analysing field-data corpora from real-world settings were published prior to 2009. Their findings are grounded in older corpora, often focused on telephone-based platforms. For instance,[43] utilised the TOOT train information corpus from 1999. Given the significant advancement of SDSs, which are revolutionising the dynamics of user-system interaction, contemporary research may yield distinctly different results [56]. The evolving capabilities of these systems from real-world settings suggests the need for renewed investigations to better understand current repair strategies in SDS. Our review also reveals a trend towards combining research techniques, which underscores the complexity of SDS research. Nevertheless, apart from corpus analysis, our findings highlight the absence of other qualitatively-driven method designs such as focus groups, case studies, or ethnography. Such methods could enable an in-depth understanding of user interaction with SDSs in real-world contexts [49].

While some studies in our review did not specify geographical location, those that did were predominantly conducted in Western countries and were largely limited to the English language. This narrow focus highlights a gap in understanding the influence of cultural and linguistic diversity on acceptance and evaluation of SDSs. For instance, [62] illustrated that people prefer robots with communication styles familiar to their culture. Similarly, [38] found that speakers' responses to communication breakdowns varied significantly by language, with 55.3% of Italian speakers but only 9.6% of Turkish speakers using a special keyword as a system-centric repair strategy. Recently, [25] investigated the impact of interactional language, specifically examining user reactions to the use of "huh?" as a system-repair strategy. Findings revealed that English speakers generally found this approach somewhat acceptable whereas Spanish speakers found it unacceptable. Future research should thus be more inclusive of diverse cultural factors, offering a broader and more global perspective on the design and effectiveness of repair strategies in SDSs.

In our review, we observed that most studies do not explicitly ground their research in established theories despite the complexity of system-user interactions. These interactions often mirror human face-to-face interactions, as demonstrated by models based on Human-Human Interaction (HHI) theories [39]. Notable theories employed by such models include Common Ground Theory [47], focusing on shared knowledge and assumptions; Justice Theory [72], examining fairness in interactions; and Humour Theory [17], exploring the role of humour. Additionally, Politeness Theory provides a comprehensive framework for exploring and understanding interactions between humans and SDSs, offering insights into managing social repair strategies such as apologies, warmth, and humour to mitigate actions or statements that challenge an individual's dignity or self-esteem [11, 26, 29]. Exploring underutilised aspects of these theories, such as expressing solidarity, exerting power, or employing indirectness to soften messages [8, 48, 68], could reveal novel ways for SDSs to more effectively repair communication breakdowns.

There is substantial potential to advance research in SDSs by applying these theories in more innovative ways. For instance, while Common Ground Theory has traditionally emphasised the importance of confirmation and questioning [5, 47], future studies could investigate how variations in explicitness of confirmations influence user satisfaction and effectiveness across different cultural contexts. Politeness Theory could be extended to explore how cultural variations in perceptions of politeness directly shape the design and effectiveness of repair strategies in SDS responses. Such approaches could significantly enhance the interactivity and responsiveness of SDSs, leading to richer, more adaptive user experiences.

As a starting point for future research, we propose the following research questions:

- How does disclosing capabilities and limitations of an SDS as a strategy to manage user expectations, enhance recovery from conversational breakdowns and affect user experience?
- How do user characteristics, such as gender, cultural background, and language proficiency influence perceptions of SDS performance in error scenarios across various contexts?
- How can multi-theoretical principles such as those used in HHI impact the effectiveness of repair strategies in SDS and user experiences?
- How do modulation strategies in user-repair influence effectiveness of SDS in various communication contexts?
- What roles do system-centric and user adaptation strategies play in the dynamics of conversational repair in SDS?
- How can qualitatively-driven research methods, such as ethnography or focus groups, enhance understanding of user experiences with SDS in real-world settings?

5 LIMITATIONS

One limitation of this review is its reliance on predefined frameworks for analysing repair strategies in conversational agents, which may not fully account for the multifaceted nature of repair strategies in SDSs. Although the reviewed frameworks provide a detailed understanding, future research could adopt more comprehensive or integrated approaches. For example, leveraging the recent taxonomy of user-repair strategies [1] or employing a theory-building approach could yield frameworks that more accurately capture the complexities of SDS repair strategies. Additionally, the subjective nature of categorising repair strategies—largely based on the authors' interpretations and compounded by terminology inconsistencies—poses another limitation. Efforts were made to categorise strategies based on definitions, examples, and the objectives described in the studies reviewed. The inclusion of combined repair strategies further complicates this task, as it blurs the distinctions between categories. Moreover, the review's scope was limited to the four selected datasets, predefined search terms, and inclusion criteria, indicating a potential gap in capturing the global spectrum of repair strategy research in SDSs.

6 CONCLUSION

Our objective was to tackle the existing challenges associated with repair strategies in SDSs, specifically addressing the absence of comprehensive conceptual frameworks and standardised terminologies. To this end, we conducted an extensive scoping review, aiming to consolidate and analyse the repair strategies employed by both SDSs and their users. We have developed comprehensive frameworks for system and user repair strategies in the SDSs domain, which provide a foundational basis for future research to explore less investigated areas.

REFERENCES

[1] Francesca Alloatti, Francesca Grasso, Roger Ferrod, Giovanni Siragusa, Luigi Di Caro, and Federica Cena. 2024. A tag-based methodology for the detection of user repair strategies in task-oriented conversational agents. *Computer Speech & Language* (2024), 101603.

[2] Christoph Draxler Anna-Maria Meck and Thurid Vogt. 2023. Failing with Grace: Exploring the Role of Repair Costs in Conversational Breakdowns with in-Car Voice Assistants. *International Journal of Human,ÄìComputer Interaction* 0, 0 (2023), 1–19. https://doi.org/10.1080/10447318.2023.2266791 arXiv:https://doi.org/10.1080/10447318.2023.2266791

[3] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. *International journal of social research methodology* 8, 1 (2005), 19–32.

[4] Anneliese Arnold, Stephanie Kolody, Aidan Comeau, and Antonio Miguel Cruz. 2024. What does the literature say about the use of personal voice assistants in older adults? A scoping review. *Disability and Rehabilitation: Assistive Technology* 19, 1 (2024), 100–111.

[5] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[6] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[7] Dennis Benner, Edona Elshan, Sofia Schöbel, and Andreas Janson. 2021. What do you mean? A review on recovery strategies to overcome conversational breakdowns of conversational agents. In *International Conference on Information Systems (ICIS)*. 1–17.

[8] Timothy Bickmore and Justine Cassell. 2001. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 396–403.

[9] Dan Bohus and Alexander I Rudnicky. 2008. Sorry, I didn't catch that! An investigation of non-understanding errors and recovery strategies. *Recent trends in discourse and dialogue* (2008), 123–154.

[10] Marie-Luce Bourguet. 2006. Towards a taxonomy of error-handling strategies in recognition-based multi-modal human–computer interfaces. *Signal Processing* 86, 12 (2006), 3625–3643.

[11] Ivan Bulyko, Katrin Kirchhoff, Mari Ostendorf, and Julie Goldberg. 2005. Error-correction detection and response generation in a spoken dialogue system. *Speech Communication* 45, 3 (2005), 271–288.

[12] Megan Cackett, Laura Fulton, and David Pearl. 2022. Sound as an Interface, Methods to Evaluate Voice User Interface (VUI) Experiences in Various Contexts. In *Proceedings of the Future Technologies Conference*. Springer, 120–129.

[13] Chen Chen, Ella T Lifset, Yichen Han, Arkajyoti Roy, Michael Hogarth, Alison A Moore, Emilia Farcas, and Nadir Weibel. 2023. Screen or No Screen? Lessons Learnt from a Real-World Deployment Study of Using Voice Assistants With and Without Touchscreen for Older Adults. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–21.

[14] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? Voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM conference on interaction design and children*. 337–348.

[15] Stephen Choularton and Robert Dale. 2004. User responses to speech recognition errors: consistency of behaviour across domains. In *Proceedings of SST*, Vol. 4.

[16] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with computers* 31, 4 (2019), 349–371.

[17] Mikkel Clausen, Mikkel Peter Kyhn, Eleftherios Papachristos, and Timothy Merritt. 2023. Exploring Humor as a Repair Strategy During Communication Breakdowns with Voice Assistants. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1–9.

[18] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. " What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services*. 1–12.

[19] Andrea Cuadra, Shuran Li, Hansol Lee, Jason Cho, and Wendy Ju. 2021. My bad! repairing intelligent voice assistant errors improves interaction. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–24.

[20] Helena ML Daudt, Catherine van Mossel, and Samantha J Scott. 2013. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC medical research methodology* 13 (2013), 1–9.

[21] Doris Dippold. 2023. "Can I have the scan on Tuesday?" User repair in interaction with a task-oriented chatbot and the question of communication skills for AI. *Journal of pragmatics* 204 (2023), 21–32.

[22] Sara Engelhardt, Emmeli Hansson, and Iolanda Leite. 2017. Better Faulty than Sorry: Investigating Social Recovery Strategies to Minimize the Impact of Failure in Human-Robot Interaction.. In *WCIHAI@ IVA*. 19–27.

[23] Shengjia Feng. 2023. How to Convey Resilience: Towards A Taxonomy for Conversational Agent Breakdown Recovery Strategies. (2023).

[24] Matthew Frampton and Oliver Lemon. 2008. Using dialogue acts to learn better repair strategies for spoken dialogue systems. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 5045–5048.

[25] Matthew Galbraith. 2023. An Analysis of Dialogue Repair in Voice Assistants. *arXiv preprint arXiv:2311.03952* (2023).

[26] Xiang Ge, Dan Li, Daisong Guan, Shihui Xu, Yanyan Sun, and Moli Zhou. 2019. Do smart speakers respond to their errors properly? A study on human-computer dialogue strategy. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments: 8th International Conference, DUXU 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*. Springer, 440–455.

[27] Julie Goldberg, Mari Ostendorf, and Katrin Kirchhoff. 2003. The impact of response wording in error correction subdialogs. In *ISCA tutorial and research workshop on error handling in spoken dialogue systems*.

[28] Julia Hirschberg, Diane Litman, and Marc Swerts. 2001. Identifying user corrections automatically in spoken dialogue systems. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

[29] Bo Huang and Sylvain Sénécal. 2023. How should voice assistants be heard? The mitigating effect of verbal and vocal warmth in voice assistant service failure. *The Service Industries Journal* (2023), 1–21.

[30] Mohit Jain, Ramachandra Kota, Pratyush Kumar, and Shwetak N Patel. 2018. Convey: Exploring the use of a context view for chatbots. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–6.

[31] Antje Janssen, Jens Passlick, Davinia Rodríguez Cardona, and Michael H Breitner. 2020. Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. *Business & Information Systems Engineering* 62 (2020), 211–225.

[32] Abe Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan. 2003. Acoustic correlates of user response to error in human-computer dialogues. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No. 03EX721)*. IEEE, 215–220.

[33] Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. 2019. Clarifying false memories in voice-based search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 331–335.

[34] Jihyun Kim, Meuel Jeong, and Seul Chan Lee. 2019. " Why did this voice agent not understand me?" error recovery strategy for in-vehicle voice user interface. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*. 146–150.

[35] Sujeong Kim, David Salter, Luke DeLuccia, and Amir Tamrakar. 2020. Study on text-based and voice-based dialogue interfaces for human-computer interactions in a blocks world. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 227–229.

[36] Lea Kisser and Ingo Siegert. 2022. Erroneous reactions of voice assistants" in the wild"–first analyses. *Elektronische Sprachsignalverarbeitung* (2022), 113–120.

[37] Esther S Kox, José H Kerstholt, Tom F Hueting, and Peter W de Vries. 2021. Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous agents and multi-agent systems* 35, 2 (2021), 30.

[38] Ivan Kraljevski and Diane Hirschfeld. 2017. Hyperarticulation of Corrections in Multilingual Dialogue Systems.. In *INTERSPEECH*. 2531–2535.

[39] Nicole C Krämer, Astrid von der Pütten, and Sabrina Eimler. 2012. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. *Human-computer interaction: The agency perspective* (2012), 215–240.

[40] Dasom Lee, Young June Sah, and Sangwon Lee. 2019. Improving usability perception of error-prone AI speakers: Elaborated feedback mitigates negative consequences of errors. *International Journal of Human–Computer Interaction* 35, 17 (2019), 1645–1652.

[41] Minha Lee and Sangsu Lee. 2021. "I Don't Know Exactly but I Know a Little": Exploring Better Responses of Conversational Agents with Insufficient Information. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.

[42] Weijane Lin, Hong-Chun Chen, and Hsiu-Ping Yueh. 2021. Using different error handling strategies to facilitate older users' interaction with Chatbots in learning information and communication technologies. *Frontiers in Psychology* 12 (2021), 785815.

[43] Diane Litman, Julia Hirschberg, and Marc Swerts. 2006. Characterizing and predicting corrections in spoken dialogue systems. *Computational linguistics* 32, 3 (2006), 417–438.

[44] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.

[45] Amama Mahmood, Jeanie W Fung, Isabel Won, and Chien-Ming Huang. 2022. Owning mistakes sincerely: Strategies for mitigating AI errors. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–11.

[46] Lina Mavrina, Jessica Szczuka, Clara Strathmann, Lisa Michelle Bohnenkamp, Nicole Krämer, and Stefan Kopp. 2022. "Alexa, You're Really Stupid": A Longitudinal Field Study on Communication Breakdowns Between Family Members and a Voice Assistant. *Frontiers in Computer Science* 4 (2022), 791704.

[47] Anna-Maria Meck, Christoph Draxler, and Thurid Vogt. 2023. Failing with Grace: Exploring the Role of Repair Costs in Conversational Breakdowns with in-Car Voice Assistants. *International Journal of Human–Computer Interaction* (2023), 1–19.

[48] Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 2022. When to say what and how: Adapting the elaborateness and indirectness of spoken dialogue systems. *Dialogue & Discourse* 13, 1 (2022), 1–40.

[49] Janice M Morse. 2016. *Essentials of qualitatively-driven mixed-method designs*. Routledge.

[50] Isabela Motta and Manuela Quaresma. 2021. Users' error recovery strategies in the interaction with voice assistants (VAs). In *Congress of the International Ergonomics Association*. Springer, 658–666.

[51] Zachary Munn, Micah DJ Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology* 18 (2018), 1–7.

[52] Christine Murad, Heloisa Candello, and Cosmin Munteanu. 2023. What's The Talk on VUI Guidelines? A Meta-Analysis of Guidelines for Voice User Interface Design. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1–16.

[53] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–7.

[54] Chelsea M Myers, Luis Fernando Laris Pardo, Ana Acosta-Ruiz, Alessandro Canossa, and Jichen Zhu. 2021. "Try, Try, Try Again:" Sequence Analysis of User Interaction Data with a Voice User Interface. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–8.

[55] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[56] Quynh N Nguyen and Anna Sidorova. 2018. Understanding user interactions with a chatbot: A self-determination theory approach. (2018).

[57] Christiane Opfermann and Karola Pitsch. 2017. Reprompts as error handling strategy in human-agent-dialog? User responses to a system's display of non-understanding. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 310–316.

[58] Diana Pérez-Marín. 2021. A review of the practical applications of pedagogic conversational agents to be used in school and university classrooms. *Digital* 1, 1 (2021), 18–33.

[59] Jessica Peterson, Patricia F Pearce, Laurie Anne Ferguson, and Cynthia A Langford. 2017. Understanding scoping reviews: Definition, purpose, and process. *Journal of the American Association of Nurse Practitioners* 29, 1 (2017), 12–16.

[60] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[61] Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2017. " Do Animals Have Accents?" Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 207–219.

[62] PL Patrick Rau, Ye Li, and Dingjun Li. 2009. Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior* 25, 2 (2009), 587–595.

[63] Antoine Raux, Dan Bohus, Brian Langner, Alan W Black, and Maxine Eskenazi. 2006. Doing research on a deployed spoken dialogue system: One year of Let's Go! experience. In *Ninth International Conference on Spoken Language Processing*.

[64] Ryan M Schuetzler, G Mark Grimes, and Justin Scott Giboney. 2020. The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems* 37, 3 (2020), 875–900.

[65] Olga Seminck. 2023. Conversational AI: Dialogue systems, conversational agents, and Chatbots by Michael McTear. *Computational Linguistics* 49, 1 (2023), 257–259.

[66] Jongho Shin, Shrikanth S Narayanan, Laurie Gerber, Abe Kazemzadeh, Dani Byrd, et al. 2002. Analysis of user behavior under error conditions in spoken dialogs.. In *INTERSPEECH*. Denver, CO.

[67] Gabriel Skantze. 2003. Exploring human error handling strategies: Implications for spoken dialogue systems. In *ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*.

[68] Mengmeng Song, Huixian Zhang, Xinyu Xing, and Yucong Duan. 2023. Appreciation vs. apology: Research on the influence mechanism of chatbot service recovery based on politeness theory. *Journal of Retailing and Consumer Services* 73 (2023), 103323.

[69] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. 2022. Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[70] Lionel Sujay Vailshery. 2021. Number of digital voice assistants in use worldwide 2019-2024. *Statista Research Departement.[Online]. Available: https://www. statista. com/statistics/973815/worldwide-digital-voice-assistant-in-use* (2021).

[71] Abdul Wali, Saipunidzam Mahamad, and Suziah Sulaiman. 2023. Task Automation Intelligent Agents: A Review. *Future Internet* 15, 6 (2023), 196.

[72] Xingyu Wang, YooHee Hwang, and Priyanko Guchait. 2023. When robot (vs. human) employees say "sorry" following service failure. *International Journal of Hospitality & Tourism Administration* 24, 4 (2023), 540–562.

[73] Syed Mekael Wasti, Ken Q Pu, and Ali Neshati. 2024. Large Language User Interfaces: Voice Interactive User Interfaces powered by LLMs. *arXiv preprint arXiv:2402.07938* (2024).

[74] K Maria Wolters, Klaus-Peter Engelbrecht, Florian Gödde, Sebastian Möller, Anja Naumann, and Robert Schleicher. 2010. Making it easier for older people to talk to smart homes: The effect of early help prompts. *Universal Access in the Information Society* 9 (2010), 311–325.

[75] Zhigang Xu, Guanqun Wang, Siming Zhai, and Peng Liu. 2023. When Automation Fails: Examining the Effect of a Verbal Recovery Strategy on User Experience in Automated Driving. *International Journal of Human–Computer Interaction* (2023), 1–11.

[76] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin, and Xia Hu. 2023. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data* (2023).

[77] Sihan Yuan, Birgit Brüggemeier, Stefan Hillmann, and Thilo Michael. 2020. User preference and categories for error responses in conversational user interfaces. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–8.

[78] Nima Zargham, Johannes Pfau, Tobias Schnackenberg, and Rainer Malaka. 2022. "I Didn't Catch That, But I'll Try My Best": Anticipatory Error Handling in a Voice Controlled Game. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.

[79] Alexander Zgorzelski, Alexander Schmitt, Tobias Heinroth, and Wolfgang Minker. 2010. Repair Strategies on Trial: Which Error Recovery Do Users Like Best?. In *Eleventh Annual Conference of the International Speech Communication Association*.

[80] Victor Zue and Stephanie Seneff. 2008. Spoken dialogue systems. *Springer Handbook of Speech Processing* (2008), 705–722.