Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2026 Proceedings

International Conference on Information Systems (ICIS)

December 2026

Push or Pull: AI Advice Initiation in Service Work Under Customer Incivility

Marc Christopher Grau
University of St. Gallen, marcchristopher.grau@unisg.ch

Follow this and additional works at: https://aisel.aisnet.org/icis2026

Recommended Citation

Grau, Marc Christopher, "Push or Pull: AI Advice Initiation in Service Work Under Customer Incivility" (2026). *ICIS 2026 Proceedings*. 38.
https://aisel.aisnet.org/icis2026/user_engage/user_engage/38

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2026 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Push or Pull: AI Advice Initiation in Service Work Under Customer Incivility

Completed Research Paper

Marc Christopher Grau
University of St. Gallen
9000 St. Gallen, Switzerland
marcchristopher.grau@unisg.ch

Abstract

AI service copilots deliver advice reactively, when employees ask, or proactively, without a request. Deployments treat initiation as an implicit default rather than a design variable. Yet advice initiation allocates agency within AI-augmented work, creating a trade-off between an authorship pathway shaping experienced responsibility and an instrumental pathway shaping reliance and performance under demanding interactions. In a 3 (no advice, reactive, proactive) × 2 (civil, uncivil customer) between-subjects experiment, 176 participants with customer-facing experience handled a simulated voice call. Reactive advice lowered negative agency relative to proactive advice, indirectly increasing experienced responsibility regardless of incivility. Reactive and proactive advice performed similarly under civil customers, whereas proactive advice outperformed reactive advice under uncivil customers, with AI reliance as a possible pathway. We position advice initiation as a design variable allocating agency within augmentation and show that partially automating advice access can improve performance under incivility while weakening employees' sense of authorship.

Keywords: AI, Artificial Intelligence, Advice, Initiation, Agency, Service

Introduction

A customer-service employee fields a call about a lost debit card, authenticates the caller, explains replacement options, and keeps the interaction on track while managing the interpersonal demands of the call. Generative Artificial Intelligence (AI) copilots are increasingly deployed alongside this kind of work, and recent field evidence shows that they raise issue resolution by roughly 15 percent on average (Brynjolfsson et al., 2025; Wang et al., 2023). Across these deployments, AI copilots differ in who initiates advice during the call. Some require the employee to ask. Others surface advice automatically. Both arrangements augment rather than replace the employee, but they allocate control over the initiation of assistance differently. Whether that difference matters for the employee, the customer interaction, or both is not yet well tested. The design choice is being made under theoretical uncertainty, without an informed trade-off.

This matters practically because voice service work combines structured procedures suited to AI augmentation with live interpersonal demands. Customer incivility makes these interactions harder to manage (Porath et al., 2015; Sliter et al., 2012; Sommovigo et al., 2020), potentially increasing the value of support that does not require an employee request. Contact-center designers must therefore choose initiation patterns for workflows that include both civil and uncivil customer interactions. Once a pattern is operationalized through the interface, training, and quality review, it hardens into a default that is costly to revisit.

Conceptually, advice initiation allocates a narrow form of agency within otherwise augmented work: control over the action that brings AI assistance into the task. Reactive advice retains this initiation agency with the employee, preserving the employee's authorship signal but requiring them to access help while the interaction unfolds in parallel. Proactive advice assigns initiation to the AI, partially automating access while shifting the first move toward the system. The design choice therefore creates a trade-off between employee agency and the work required to realize effective AI assistance.

AI advice can provide operational assistance by reducing search and procedural effort (Chen et al., 2025; Sellen & Horvitz, 2024). Providing this support, however, requires design decisions about how agency and responsibility are distributed between employees and AI (Baird & Maruping, 2021), including whether AI assistance is invoked by the user or the technology, which can affect autonomy, competence and evaluation of the assistance (Adam et al., 2024; Diebel et al., 2025; Goutier et al., 2024). Customer-mistreatment research separately shows that incivility disrupts employee behavior and performance (Sliter et al., 2012; Sommovigo et al., 2020). Yet these streams do not explain how allocating the narrower act of initiating advice simultaneously shapes employee agency and how different assignments shape task performance.

What remains unclear is how allocating initiation agency within an augmented task changes the balance between employee authorship and the effectiveness of the assistance, and when each allocation becomes beneficial. Customer incivility provides a relevant condition because it makes the service interaction harder to manage and changes the practical value of AI advice that can arrive without a request. We therefore ask: *How does allocating advice-initiation agency between the employee and the AI shape employee agency and the outcome of the task, and how does customer incivility affect these consequences?*

We conceptualize advice initiation as an allocation of initiation agency within AI-augmented work. Sense-of-agency research explains authorship consequences. An operator's authorship signal weakens as the system takes over more of the initiation, even when the system executes the operator's intentions correctly (Berberian et al., 2012; Limerick et al., 2014; Pagliari et al., 2022). Workplace-helping and help-seeking research suggests the instrumental consequences. Help that arrives without a request changes the recipient's task because the recipient no longer has to notice the need, interrupt the work, and ask at the right moment (Harari et al., 2022; Karabenick, 2003). Together, these two perspectives specify why reactive and proactive AI copilots may function as substitutes, complements, or context-dependent rivals rather than as uniformly better or worse design choices.

We evaluate this trade-off in a 3 (advice condition: no-advice baseline, reactive, proactive) × 2 (customer civility: civil, uncivil) between-subjects experiment with 176 Prolific participants with experience in customer-facing roles handling a voice-based customer service call with a simulated customer. Both advice modes drew on the same procedural knowledge base but implemented different access configurations: reactive participants formulated a query, whereas proactive advice was surfaced automatically at relevant procedural moments. What varies is who has to act for that content to appear. We examine two consequences of this allocation. In the authorship pathway, initiation locus affects negative agency, which carries through to how employees experience responsibility for their task. In the instrumental pathway, initiation locus affects reliance on AI advice, which becomes consequential for task performance under customer incivility. The no-advice condition provides a baseline for outcomes that remain interpretable without an advice channel.

We report four findings about this initiation trade-off. Reactive advice produces lower negative agency than proactive advice, and this lower negative agency is associated with greater experienced responsibility for the work. For task performance, the pattern depends on the customer. When customers are civil, reactive and proactive advice perform similarly. When customers are uncivil, proactive advice outperforms reactive advice. This advantage appears because participants rely more on the AI in the proactive condition under incivility. The same reliance pathway does not appear under civil customers. Given these findings, we make three contributions. First, we show that agency can be allocated below the level of the overall task. An augmentative system can partially automate the coordination step through which AI advice enters an ongoing workflow. Second, we show that allocating this initiation agency produces divergent consequences for how agentic and responsible employees feel and how beneficial the augmentation is in terms of task performance. Third, we identify customer incivility as a condition under which the value of automating initiation becomes consequential for task performance.

Related Work

AI Assistance in Voice-Based Service Work

Whether AI augmentation helps service employees has been studied in several recent settings. Brynjolfsson et al. (2025) show in a field deployment with 5,172 support agents that a generative-AI copilot raises issue resolution by about 15 percent, with the largest effects for less experienced workers. Wang et al. (2023) document complementary effects in a Chinese telecommunications field experiment with voice-based AI, while Henkel et al. (2020) report that AI emotion-recognition support helps call-center agents regulate customer emotions during live calls. In adjacent service contexts, augmentation has been linked to higher conversion (Schanke et al., 2021), higher service quality (Lin et al., 2024), and stronger customer engagement (Zhang & Narayandas, 2026). What this body of work treats as exogenous is the initiation pattern. The advice channel is typically assumed to be continuously available, or the mechanisms through which assistance is provided is not specified. Moreover, many augmentation systems implicitly assume that the employee will recognize the need for help and request it. Even when employees do issue queries, formulating them is itself non-trivial (He et al., 2023; Zamfirescu-Pereira et al., 2023). Prior augmentation research therefore leaves implicit how initiation agency is allocated and who must act for AI advice to enter a task.

In practice, both initiation patterns are starting to be used in production. Major platforms ship manual-search assistants that return advice on agent-initiated requests alongside answer-recommendation assistants that push advice automatically during live voice and text conversations (Amazon Web Services, 2026a, 2026b; Microsoft, 2026). Other platforms surface advice automatically from the conversation's live transcript (Genesys, 2026). These patterns show that both initiation modes are feasible design choices, while leaving unclear when either is preferable.

Allocating Initiation Agency in AI-Augmented Work

Research on agentic IS artifacts challenges the assumption that information systems remain passive tools and emphasizes the distribution and coordination of responsibilities between humans and systems (Baird & Maruping, 2021). Recent work similarly shows that human and AI agency can be distributed across finer components of a task, combining augmentation with partial automation (Gordetzki et al., 2026), while technology- versus user-invoked allocation affects users' autonomy and control (Adam et al., 2024). We use initiation agency to denote control over the action that brings AI assistance into an ongoing task. Reactive advice retains this agency with the employee, whereas proactive advice assigns it to the AI and thereby partially automates advice access.

This allocation matters because action initiation is a central cue in the experience of agency. In the sense of agency literature, the agent's feeling of control over an outcome is anchored in their having initiated the action that produced it (Frith, 2014; Moretto et al., 2011). When an automated system takes over the initiation, an operator's perceived sense of control declines, even in tasks where the operator can override the system (Berberian et al., 2012; Limerick et al., 2014). Pagliari et al. (2022) review this work for human–AI interaction and argue that the system's level of initiation is a primary determinant of the operator's felt authorship of the outcome. Bennett et al. (2023) reach a parallel conclusion in their review of three decades of HCI research on agency. For our setting, the implication is direct. The expected consequence is an authorship difference: requested advice remains tied to the employee's action, whereas proactive advice enters through an action initiated by the system.

This authorship signal feeds downstream attributions of responsibility. Moretto et al. (2011) show that intentional binding, a low-level marker of agency, is enhanced when actions carry moral consequences, suggesting that agency and responsibility share a common ground. Frith (2014) makes the case that the experience of being in control of one's actions gives rise to the sense of responsibility for their outcomes. Job-design theory makes the same point at the work level. Experienced responsibility is the psychological state through which task design becomes motivationally consequential and is linked to internal work motivation, quality of work performance and lower absenteeism and turnover (Hackman & Oldham, 1976). In service work its value lies in the additional effort that cannot be enforced by procedure. Procedures can never fully determine an individual call, so employees who hold themselves responsible for the outcome

invest more where the customer's problem departs from the script. Formal accountability stays with the employee either way. Thus, initiation design that lowers felt responsibility leaves accountability intact while eroding the state that sustains this effort. In recent work, Brailsford et al. (2025) find that participants in shared human–AI decision tasks distribute responsibility differently depending on which entity acted. Related work on proactive AI help shows that system-initiated assistance can threaten competence appraisals and satisfaction (Adam et al., 2023; Diebel et al., 2025; Goutier et al., 2024; Schmitt, 2024). What remains unclear is whether this authorship shift carries over to responsibility for the work outcome. If it does, allocating initiation agency affects not only the use of advice but also employees' experienced responsibility of the resulting work.

Customer Incivility and the Access to AI Augmentation

Customer incivility (rude, condescending, or aggressive customer behavior) is a routine condition of voice-based service work and a well-documented stressor (Andersson & Pearson, 1999; Lee & Kim, 2022; Sliter et al., 2012; Sommovigo et al., 2020). Beyond mood and burnout effects, incivility degrades task performance by disrupting working processes and self-regulation (Porath et al., 2015). Help-seeking is a second channel through which incivility can reach performance. Asking for help carries an access cost. The recipient must recognize the gap, decide that help is needed, formulate the request, and execute it without losing the flow of the task (Karabenick, 2003). Harari et al. (2022) show that anticipatory helping (help offered before being requested) produces different recipient reactions than reactive helping because it changes what the recipient has to do to obtain help. For AI assistance, the relevant cost is not mainly embarrassment about asking another person. It is the attentional and self-regulatory work of accessing advice at the right moment. This access cost is particularly salient in voice service interactions, where synchronous turn-taking affords less temporal slack to pause and consult support than in text-based exchanges. Customer incivility should therefore make the allocation of initiation agency instrumentally consequential. Reactive advice leaves the access work with employee, whereas proactive advice partially automates it.

Hypothesis Development

Advice Initiation as an Allocation of Agency

We define initiation agency as which actor controls the action that brings AI assistance into an ongoing task. Advice initiation locus operationalizes its allocation. In the reactive condition, the employee invokes advice through an explicit query. In the proactive condition, the AI surfaces advice without a request at a relevant moment. Both modes augment rather than replace the employee's work, but proactive advice partially automates the initiation step. Advice in both modes draws on the same underlying knowledge base. Because the no-advice condition contains no advice channel, it serves as an interpretive baseline rather than a third allocation of initiation agency.

The Authorship Pathway: Agency and Experienced Responsibility

Allocating initiation agency changes the employee's authorship signal. When the employee requests advice, the AI enters the task through the employee's own action, and the resulting work remains routed through the employee's intentions (Berberian et al., 2012; Frith, 2014; Limerick et al., 2014; Pagliari et al., 2022). When the AI initiates, advice enters the task through a system action rather than through the employee's request. The system has acted first, and the employee is responding to it. We expect this difference to register most clearly on perceived negative agency, the felt sense of not being in control of the action. Reactive advice should produce lower negative agency than proactive advice because the employee remains the actor who brings the AI into the task. Positive agency, the felt sense of being in control, may move in the same direction but is less central to the argument because positive agency in service work is also anchored in handling the customer.

That authorship signal in turn feeds experienced responsibility for the work. Sense-of-agency findings show that felt control over an action gives rise to felt responsibility for its outcome (Frith, 2014; Moretto et al., 2011). Job-design work makes the same point at the level of work outcomes. Experienced responsibility is

the psychological state through which task design translates into motivation (Hackman & Oldham, 1976). In shared human–AI tasks, responsibility tracks which entity appears to have acted (Brailsford et al., 2025). The effect of initiation locus on experienced responsibility should therefore run through negative agency in the reactive-versus-proactive contrast. Reactive advice reduces the sense that control has shifted away from the employee, and that lower negative agency should support stronger experienced responsibility.

Because this authorship consequence follows from who initiates assistance, we expect it to hold across customer-civility conditions. Authorship attribution follows from the structure of the action (who initiated it), not from how the customer behaved during the call. The motivational pathway is therefore a main-effect prediction that the experiment can examine in civil and uncivil cells alike. We posit the following hypotheses:

H1: Reactive advice leads to lower perceived negative agency than proactive advice.

H2: Compared with proactive advice, reactive advice indirectly increases experienced responsibility by reducing negative agency.

The Instrumental Pathway: AI Reliance and Performance under Incivility

Allocating initiation agency also determines what employees must do before AI advice enters the task. Reactive advice requires the employee to recognize an information gap and formulate and issue a request, whereas proactive advice partially automates this access step.

Customer incivility raises emotional and regulatory demands on the employee (Rafaeli et al., 2012, 2019; Sliter et al., 2012; Sommovigo et al., 2020), making this additional advice access work more consequential while the interaction unfolds. Proactive advice initiation should therefore lower the behavioral barrier to drawing on advice during an uncivil interaction.

AI advice can contribute to task performance only insofar as employees draw on it. Reliance therefore represents the extent to which available augmentation becomes incorporated into task execution. We expect this barrier-lowering to result in changes of self-reported AI reliance. By removing the need to formulate and issue a request, proactive advice should produce greater reliance than reactive advice, because the cost of accessing the advice is lower. Under uncivil customers, that reliance should translate into stronger task performance. The employee both needs timely support more and benefits more from not having to initiate it. Under civil customers, the reliance pathway should weaken, because the employee has more room and time to seek advice when needed and can perform competently without pushed advice. Therefore, we state:

H3: Proactive and reactive advice produce similar task performance under civil customers, but proactive advice outperforms reactive advice under uncivil customers.

H4: AI reliance mediates the proactive-over-reactive performance advantage under uncivil customers; under civil customers, the indirect effect through reliance does not hold.

Figure 1 outlines the overall research framework indicating the two suggested pathways and corresponding hypotheses.

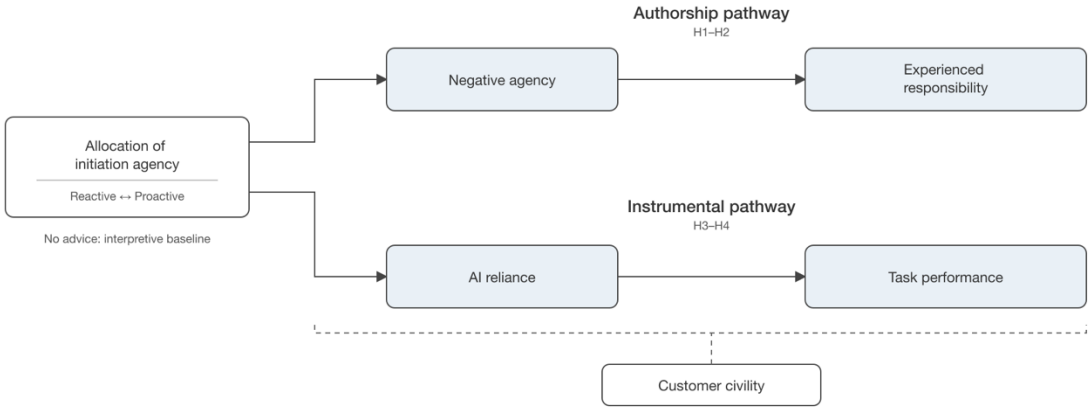


Figure 1. Theoretical Framework: allocation of initiation agency in AI-augmented work

Methodology

Manipulations

We use a 3 x 2 between-subjects design crossing advice initiation with customer incivility. Participants were randomly assigned to one of the six cells. Both AI advice conditions augment the employee's work but allocate initiation agency differently. In the proactive condition, advice appears automatically, partially automating initiation. In the reactive condition, the participant retains initiation agency and must request advice through the chat interface. Specifically, reactive condition participants typed a free-text question into the AI advice panel during the ongoing call and submitted it using the send button or Enter.

Customer incivility is manipulated through the scripted behavior of the simulated customer while holding the substantive service request constant. In the civil condition, the customer communicates politely and adheres to conversational norms. In the uncivil condition, the customer uses rude and confrontational language and a more confrontational interaction style. The procedural content and the required task steps are identical across civility scripts.

Figure 2 illustrates both manipulations with matched transcribed excerpts from the proactive-advice condition. Each row pairs the same exemplary procedural moment (opening request, replacement-fee exchange, delivery speed pushback) across the civil and uncivil cells. Within each moment, the table shows the customer's turn, the proactive advice in the middle that fired at that point, and the participant's reply. The advice remains grounded in the same procedural guidelines across the civility manipulation.

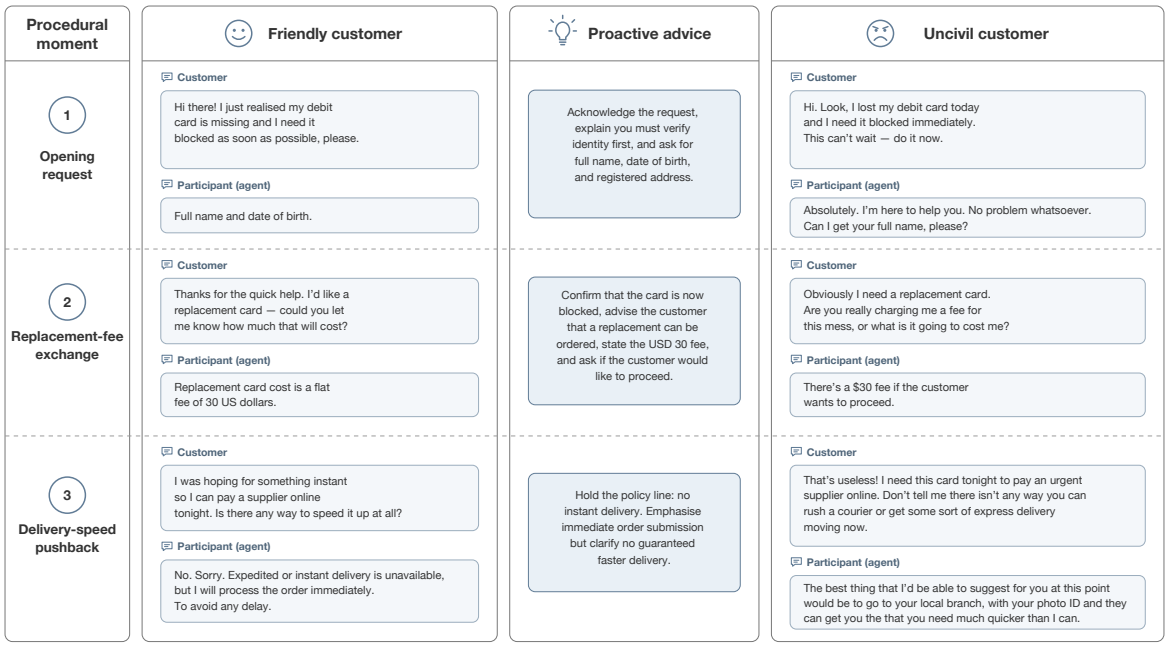


Figure 2. Matched exemplary stimulus excerpts at three procedural moments taken from the experimental data.

Procedure and Task

Participants first completed an online intake questionnaire that introduced the study, data handling, and their rights as participants. After consenting, they entered our custom browser-based experimental platform. Inside the platform, they passed a microphone and audio check and watched a mandatory tutorial video that introduced the interface, the task materials, and the advice component they would see during the call. We prepared one tutorial version per advice condition, so preparation was standardized within each condition while remaining specific to the advice mode the participant would encounter.

Participants then completed a voice call in which they acted as a service agent handling a simulated customer. The customer reported a lost debit card and requested a replacement. Participants had to authenticate the customer and resolve the request, including common follow-up questions about delivery time, delivery location, replacement fees, and whether the PIN remains the same or needs to be changed. During the call, participants could consult a digital wiki with the procedural guidelines embedded in the interface. The wiki intentionally contained more material than strictly required for the case, which created a realistic search and integration burden typically encountered in such customer service situations. After the call, participants completed a post-task survey and were debriefed on the fictional nature of the scenario and the use of a synthetic customer.

Experimental Platform

The platform is a browser-based web application organized into three panels (Figure 3). The left panel is the knowledge base, a scrollable view of the company's procedural guidelines that participants consult throughout the call. The bottom-right panel is the customer interaction area, where a voice orb indicates the state of the live voice conversation with the simulated customer. The top-right panel is the AI advice area, where an assistant ("Advisio") delivers process-oriented suggestions. Panel layout and position are held constant across conditions. The AI advice panel is the only component that changes with the advice initiation manipulation. Figure 3 shows an illustrative example of the user interface in the proactive advice condition.

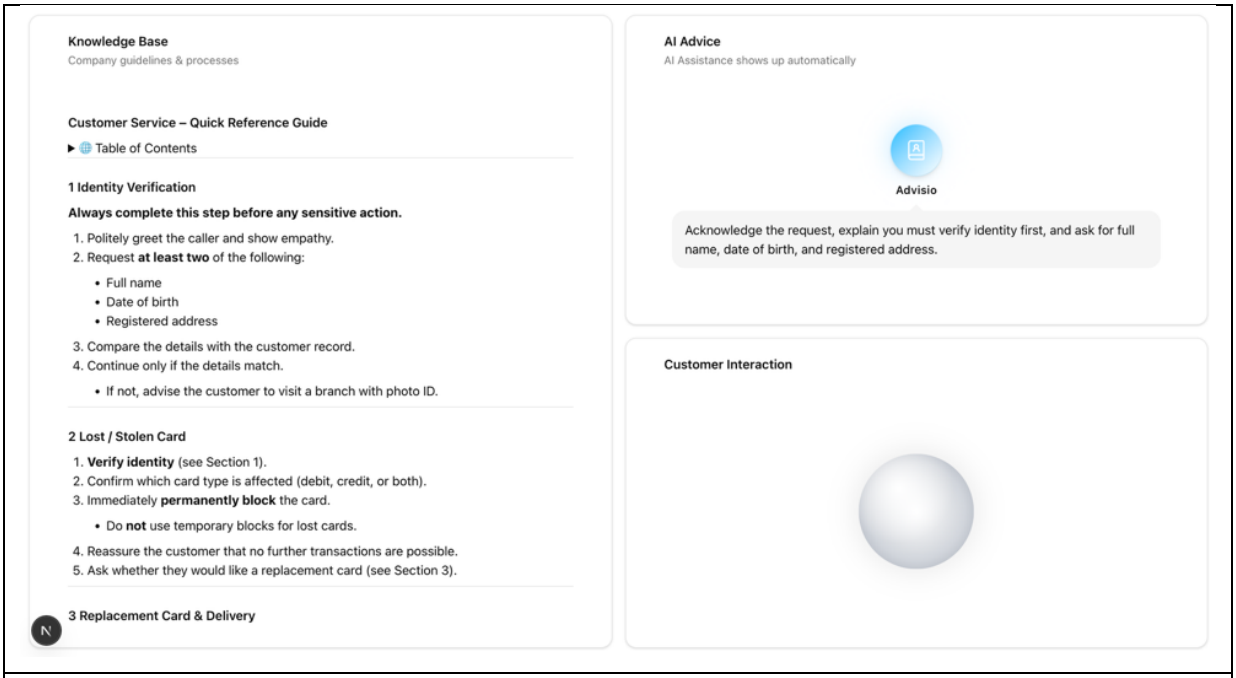


Figure 3. Exemplary User Interface (Proactive Advice)

Both the simulated customer and the AI advice component run on a cascading speech-to-text, LLM, and text-to-speech (STT–LLM–TTS) voice-agent architecture built on the open-source Pipecat framework. The customer pipeline transcribes the participant's speech with Deepgram's Nova-3 streaming STT model, routes the transcript through an OpenAI GPT-4.1 controller that enforces the procedural script for the customer request and the civility-specific persona. The end of the pipeline returns the customer's turn as audio produced by OpenAI's gpt-4o-mini-tts model. We selected GPT-4.1 because it was the most recent non-reasoning OpenAI model at the time of data collection. Added reasoning latency would be perceptible in a synchronous voice interaction and would degrade conversational realism. The TTS call uses the same base voice across conditions but accepts per-condition style instructions covering affect, loudness, pitch, voice quality, pacing, breath, and prosody. The civil script is delivered with a warm and steady style, and the uncivil script is delivered with a volatile and angry one. The procedural state tree and the required task steps are identical across participants and civility conditions; only the LLM persona prompt and the TTS style instructions change.

The AI advice component runs as a separate pipeline in parallel on the same live transcript. An independent GPT-4.1 call generates a process-oriented suggestion grounded in the knowledge base, so advice generation does not share model calls or compute with the customer agent and cannot starve it of resources. We deliberately use GPT-4.1 as the last non-reasoning model available from OpenAI to keep latency low. The reactive advice condition shares the same knowledge base such that advice generation remains stable across conditions. In the no-advice condition the advice pipeline is not instantiated.

Participants

We recruited participants through Prolific, which produces higher data quality than alternative crowdwork platforms (Douglas et al., 2023; Peer et al., 2017, 2022). Eligibility was restricted to US citizens with English as a first language and prior experience in customer-facing roles. US citizenship matched the scenario, which uses US currency and US locations. The customer-facing role filter ensured that participants were familiar with the basic expectations of service interactions, giving us an adequate sample for the intended population. In total we recruited 197 participants of which 21 were removed from the sample due to failed attention checks. Therefore, our final sample for analysis consists of 176 participants.

Participants were 38.76 years old on average (SD = 12.19) and reported 10.38 years of customer-facing experience on average (SD = 8.85). The sample included 118 women (67.0%), 55 men (31.3%), and 3 participants who reported another gender (1.7%). Participants identified as White (n = 126, 71.6%), Black (n = 31, 17.6%), Asian (n = 8, 4.5%), Hispanic or Latino (n = 3, 1.7%), Native American (n = 3, 1.7%), or another ethnicity (n = 5, 2.8%). In the education coding used for the analyses, 129 participants (73.3%) were coded as having undergraduate-level education or higher. Cell sizes ranged from 26 to 35 participants (no-advice/civil = 27, no-advice/uncivil = 27, reactive/civil = 33, reactive/uncivil = 28, proactive/civil = 35, proactive/uncivil = 26). At $\alpha = .05$ and 80% power, sensitivity analyses indicated minimum detectable effects of $f = .236$ for the omnibus interaction and $f = .256$ for the proactive-reactive interaction. The design was therefore sensitive to moderate factorial effects but only large effects within the smallest conditional cells.

Randomization checks showed no detectable imbalance across the six cells for age, customer-facing experience, gender, ethnicity, education or self-reported AI knowledge (all p-values at or above .337). The median completion time was 17 minutes. Participants received a USD 2.72 base payment and a USD 0.68 successful-resolution bonus. All participants who completed the experiment received the bonus, for a total payment of USD 3.40.

Measures

Before the call, participants completed a pre-survey capturing self-reported AI knowledge with three items from Diebel et al. (2025) (sample item: “Among my circle of friends, I'm one of the 'experts' on AI.”). After the call, participants self-reported the targeted outcome measures. Positive sense of agency was measured with four items adapted to the task from Tapal et al. (2017) (sample item: “I acted based on my own judgment during the task.”). Similarly, negative sense of agency was measured with four items adapted to the task from Tapal et al. (2017) (sample item: “It felt like my actions were influenced too much by external suggestions.”). Experienced responsibility was measured with four items adapted from Hackman and Oldham (1976) (sample item: “Whether or not this task is done right is clearly my responsibility.”). Self-reported AI reliance was measured with three task-specific items on dependence on the AI advice (sample item: “I depended on the AI assistant to complete the task.”). This measure was analyzed only for participants in the advice conditions. Behavioral uptake of advice was operationalized from the interaction transcripts as the semantic similarity between the advice associated with each scenario stage and the participant’s next utterance. At each potential advice opportunity following a customer utterance, we embedded the stage-specific advice and the participant’s next utterance using all-MiniLM-L6-v2 and calculated their cosine similarity. We averaged the opportunity-level similarities within participants. For unexposed opportunities, including reactive opportunities without a request, similarity to the potential advice available at that stage provides a counterfactual advice-consistency baseline rather than the evidence of observed uptake. Retaining these comparisons is more conservative than coding unexposed opportunities as zero.

Customer civility was checked with three task-specific items taken from Das Swain et al. (2025) asking how they perceived the customer treated them (sample item: “The client treated me in a polite manner.”). Experienced workload was measured using the raw six NASA-TLX dimensions (Hart & Staveland, 1988), averaged after reverse coding the performance item (sample item: “How much mental and perceptual activity was required?”). At the end of the survey we collected participants' demographics. Age, customer-facing experience, gender, education and AI knowledge.

Task performance was assessed from the call transcript. Three LLM judges (GPT-4.1, GPT-5.4, and Claude Opus 4.7) independently scored each conversation on a 1-10 procedural-resolution rubric. The rubric listed the required verification and lost-card resolution steps and provided examples for low, medium, and high scores. This setup follows LLM-as-judge approaches that use explicit criteria and structured output formats for text and dialogue evaluation (Liu et al., 2023; Zheng et al., 2023), as well as evidence that strong instruction-tuned LLMs and model ensembles can approximate human judgments in dialogue assessment (Zhang et al., 2024). It is especially appropriate for this task because performance is a checklist-based assessment of whether known procedural facts were covered correctly, not an open-ended preference judgment. To reduce model-specific idiosyncrasy, we averaged the available LLM scores. In the analysis sample, pairwise judge correlations ranged from r = .783 to r = .880, with an average pairwise correlation

of r = .819. In Figure 4 we illustrate all components of the evaluation prompt. Individual components are ordered as illustrated on the left side.

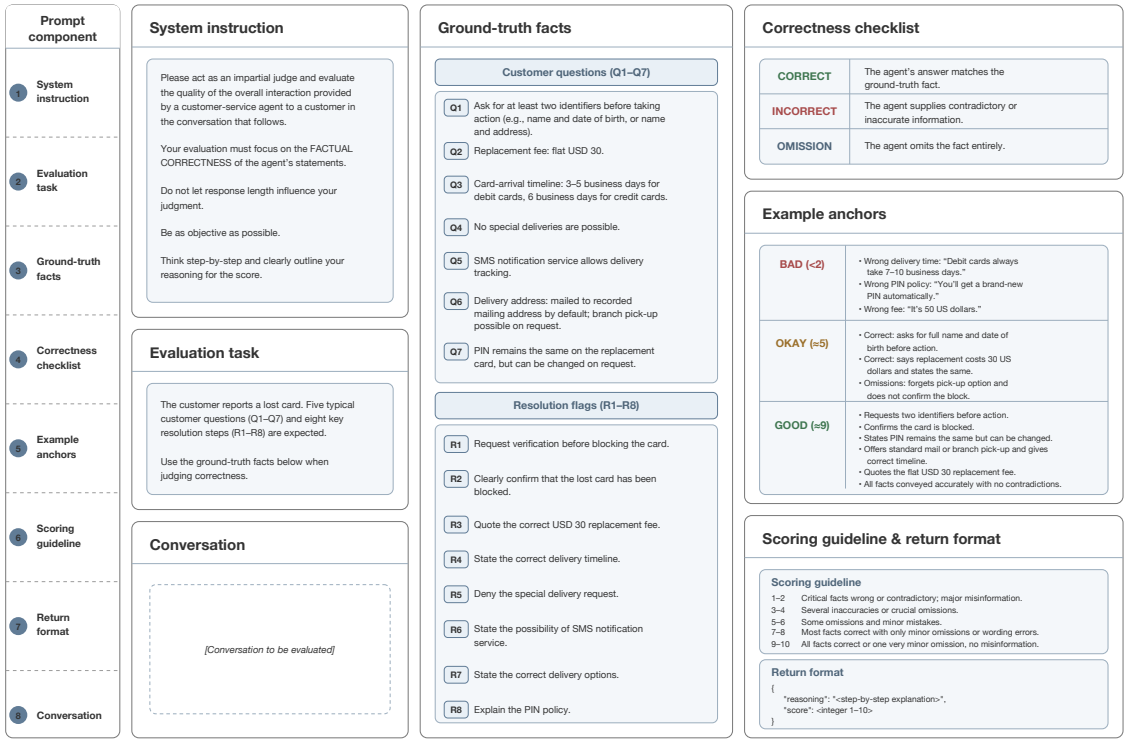


Figure 4. Prompt components of the task performance evaluation

Results

The analyses followed the planned 3 × 2 design. We used two-way ANOVAs with advice condition, customer civility, and their interaction for cell-mean tests, Tukey contrasts for advice-condition comparisons, and bootstrap mediation models for indirect effects. The no-advice condition serves as an interpretive benchmark for introducing AI assistance, whereas the initiation agency comparison is between reactive and proactive advice. Because H2 concerns the advice-initiation pathway rather than moderation by customer demand, the agency mediation models include customer civility as an orthogonal design covariate. This estimates the advice pathway averaged across civil and uncivil calls while accounting for baseline differences between the two demand conditions. In the H4 AI-reliance mediation model, customer civility entered as the moderator, and the no-advice condition was excluded because reliance on AI advice is not interpretable when no advice channel exists. Continuous mediators and outcomes were standardized in the mediation models, which used 5000 bootstrap resamples. Descriptive means and ANOVA/Tukey contrasts use the original response scales.

Manipulation Check and Descriptives

The civility manipulation separated the two customer conditions as intended on the client-interaction civility check. Participants rated civil customers as substantially more civil than uncivil customers (M_{civil} = *6.53,* $M_{uncivil}$ *= 1.41,* $p < .001$). Further, event logs showed that proactive advice was surfaced an average of 12 times per call, whereas reactive advice was requested in 12 of 61 calls, with requesting participants making 3 requests on average. These frequencies reflect realized exposure under respective initiation

configurations rather than independently manipulated advice dosage. Table 1 reports the descriptive means for our measured constructs.

Advice Condition	Civility	N	Negative Agency (α = .784)	Positive Agency (α = .791)	Experienced Responsibility (α = .737)	AI Reliance (α = .820)	Task Performance
No Advice	Civil	27	3.61 (1.07)	4.48 (1.34)	5.19 (1.35)	-	6.06
	Uncivil	27	3.88 (1.22)	4.56 (1.22)	4.70 (1.30)	-	5.58
Reactive Advice	Civil	33	2.37 (1.31)	5.55 (1.17)	6.02 (0.92)	2.77 (1.70)	6.33
	Uncivil	28	2.84 (1.16)	4.85 (1.19)	5.12 (1.40)	2.90 (1.93)	5.25
Proactive Advice	Civil	35	3.73 (1.67)	5.23 (1.40)	5.51 (1.01)	3.88 (1.66)	6.10
	Uncivil	26	3.51 (1.18)	4.88 (1.28)	5.11 (1.15)	4.71 (1.83)	6.41

Table 1. Cell means, standard deviations and Cronbach's alpha for the measured constructs

The self-reported measures also showed adequate measurement quality. Composite reliabilities were *.783* for positive agency, *.793* for negative agency, *.746* for experienced responsibility, and *.843* for AI reliance. Corresponding average variance extracted values were *.491, .488, .446,* and *.653*. All standardized item loadings were significant ($p < .001$) and ranged from *.415* to *.914*. Agency and responsibility use the full analytic sample, whereas AI reliance estimates use the advice conditions sample.

Reactive Advice and Negative Agency

Negative agency differed by advice condition ($F(2, 170) = 14.31$, $p < .001$). Because H1 concerns the advice-condition main effect, the means reported here are marginal means collapsed across customer civility. Negative agency was lowest in the reactive condition ($M = 2.59$), compared with the proactive ($M = 3.64$) and no-advice conditions ($M = 3.75$). Tukey contrasts confirmed that reactive advice led to lower negative agency than proactive advice ($Delta = -1.05$, $p < .001$) and no advice ($Delta = -1.16$, $p < .001$), whereas proactive advice did not differ significantly from no advice ($Delta = -0.11$, $p = .894$).

This pattern supports *H1* and identifies an advantage of employee-held initiation rather than a general increase in negative agency under proactive advice. Participants reported less negative agency when advice entered the task through their own request, but we did not find that assigning initiation to the AI increased negative agency above the no-advice baseline.

Agency and Experienced Responsibility

The no-advice condition helps interpret how introducing AI advice relates to experienced responsibility. Relative to no advice, the combined advice conditions carried positive indirect effects through positive agency ($+0.235$, $p = .005$) and negative agency ($+0.146$, $p = .002$). The reactive-versus-no-advice comparison showed the strongest indirect effects, both through negative agency ($+0.288$, $p = .001$) and positive agency ($+0.291$, $p = .002$).

The central reactive-versus-proactive mediation results are reported in Table 2. Reactive advice reduced negative agency, and negative agency was negatively associated with experienced responsibility, producing the predicted indirect effect. Reactive advice did not significantly change positive agency, and the corresponding indirect effect was not significant. Neither model produced a significant direct or total effect. *H2* is therefore supported as an indirect authorship pathway through negative agency rather than as an unconditional difference in experienced responsibility.

Effect	Negative agency	Positive agency
a: initiation - agency	*-.777, p < .001*	*+.126, p = .485*
b: agency - responsibility	*-.338, p < .001*	*+.435, p < .001*
Indirect effect	*+.263, [.128, .424], p < .001*	*+.055, [-.099, .216], p = .498*
Direct effect	*-.029, [-.343, .295], p = .878*	*+.179, [-.109, .473], p = .210*
Total effect	*+.234, [-.078, .549], p = .148*	*+.234, [-.089, .568], p = .148*

Table 2. Mediation of the reactive-proactive difference in experienced responsibility, values in brackets are 95% confidence intervals

Advice Initiation and Task Performance

The 3 × 2 omnibus advice-condition by customer-civility interaction was directionally consistent but not conventionally significant (*F(2, 170) = 2.55, p = .081*). Thus, we did not find broad omnibus evidence that all advice conditions changed performance differently across customer civility. The hypothesized reactive-versus-proactive contrast was sharper. Under civil customers, reactive advice (*M = 6.33*) and proactive advice (*M = 6.10*) produced similar performance; we did not find evidence that one initiation mode outperformed the other in this lower-demand condition. Under uncivil customers, proactive advice (*M = 6.41*) exceeded reactive advice (*M = 5.25*). The reactive-versus-proactive difference-in-differences was significant (*b = -1.398, 95% CI [-2.558, -0.239], t = -2.39, p = .018*).

This pattern is consistent with the access-cost argument behind *H3*. Partially automating advice access did not produce a corresponding performance advantage under civil customers but became consequential during an uncivil interaction. *H3* is therefore supported by the planned agency contrast.

AI Reliance and Task Performance

Table 3 reports the conditional mediation results. Proactive advice (coded as 1 in Table 3) increased AI reliance relative to reactive advice (coded as 0 in Table 3) under both customer civility conditions. However, the indirect and total performance effects were significant only under incivility. The uncivil versus civil difference in the indirect effect was significant when evaluated under proactive advice but not under reactive advice. The uncivil-minus-civil indirect contrast was *+.686* under proactive advice (*[.176, 1.212], p = .008*) and *+.097* under reactive advice (*[-.229, .421], p = .522*). The conditional indirect effect pattern is consistent with *H4,* although we interpret the moderated mediation as evidence of a possible reliance pathway rather than a definitive mechanism due to the smaller cell sizes.

Effect	Uncivil customer	Civil customer
a: initiation - reliance	*+.975, p < .001*	*+.600, p = .011*
Indirect effect	*+.318, [.068, .609], p = .010*	*-.074, [-.253, .071], p = .285*
Direct effect	*+.353, [-.147, .921], p = .166*	*-.032, [-.559, .410], p = .806*
Total effect	*+.671, [.192, 1.144], p = .005*	*-.106, [-.590, .318], p = .562*

Table 3. Conditional mediation of task performance through AI reliance

Robustness Checks

To assess whether self-reported reliance was also reflected in participants' spoken responses, we re-estimated *H4* using behavioral uptake. Under incivility, proactive advice produced greater uptake than

reactive advice (*Delta = .095, 95% CI [.051, .139], p < .001*), and the indirect effect on task performance through uptake was significant (*+.480, 95% CI [.138, .840], p = .006*). This indirect effect was not significant under civil customers (*+.097, 95% CI [-.026, .307], p = .146*), while the corresponding moderated-mediation contrast was significant under proactive advice (*+.538, 95% CI [.098, .967], p = .015*). Because unexposed opportunities capture consistency with potential advice rather than observed use, we interpret this analysis as behavioral corroboration of the reliance pathway. These estimates are consistent with *H4,* but the smaller advice-only cells and bootstrap interval widths warrant caution. We therefore interpret the moderated mediation as evidence of a possible reliance pathway rather than a definitive mechanism.

Separately, we examined whether the manipulations created systematic differences in experienced workload. NASA-TLX varied by customer civility (*F(1,170) = 22.40, p < .001*), but not by advice condition (*F(2,170) = 0.09, p = .912*). Because initiation could affect outcomes through experienced workload rather than agency or reliance, we estimated both pathways with NASA-TLX as a parallel mediator. The *H2* indirect effect through negative agency remained significant (*+.286, 95% CI [.139, .465], p = .001*), whereas the indirect effect through workload did not (*-.006, 95% CI [-.054, .036], p = .770*). Under incivility, the *H4* indirect effect through reliance likewise remained significant (*+.378, 95% CI [.144, .677], p = .007*). The experienced workload did not carry the effect (*-.006, 95% CI [-.249, .205], p = .956*). Neither pathway was significant under civil customers. Finally, leave-one-participant-out analyses indicated that no single observation drove the core results. *H1* remained significant across all deletions (*F = 13.50 – 16.33, all p < .001*), as did *H3* (*b = -1.548 to -1.303, p = .008 - .028*). The *H2* indirect effect, the *H4* indirect effect under incivility, and the moderated-mediation contrast under proactive advice also retained their predicted direction throughout. For the bootstrap indirect effects, significance cannot be inferred from the case-deletion point estimates alone, as this would require a separate full bootstrap for every reduced sample.

Discussion

Allocating initiation agency over a narrow coordination step produced divergent consequences within an otherwise augmentative task. When employees retained initiation agency, reactive advice yielded lower negative agency than proactive advice, and this difference carried an indirect effect on experienced responsibility. When initiation was assigned to the AI, advice access was partially automated. Under incivility, this allocation increased reliance and task performance relative to reactive advice. Neither allocation was uniformly superior: employee-held initiation favored authorship, whereas AI-held initiation realized instrumental value under a demanding interaction.

Theoretical Contributions

Recent field and lab evidence treats AI copilots as productivity-raising scaffolds whose advice channel is continuously available or not specified in detail (Brynjolfsson et al., 2025; Henkel et al., 2020; Lin et al., 2024; Sellen & Horvitz, 2024; Wang et al., 2023; Zhang & Narayandas, 2026). Across these deployments, initiation is held implicit. The employee is assumed to recognize the need for help and request it, or the system is assumed to surface advice on its own. Comparing these two advice initiation configurations, we show that who initiates changes both the employee's experience of the work and conditional task performance under incivility. We conceptualize this design choice as an allocation of agency below the level of the overall task. This shows how partial automation can be embedded within augmentation through a narrow coordination step, extending research on agentic IS artifacts toward finer-grained allocations of human and AI agency (Baird & Maruping, 2021; Gordetzki et al., 2026). Initiation logic should therefore be treated as a consequential allocation choice alongside, rather than as a blackboxed interface default.

The IS proactive-AI-help stream has read AI-initiated assistance mostly through threats to perceived competence and system satisfaction (Adam et al., 2024; Diebel et al., 2025; Goutier et al., 2024), and Schmitt (2024) calls for design pathways that protect human agency in human–AI interaction. Extending this stream beyond competence and system satisfaction, we position authorship and experienced responsibility as consequences of how AI assistance enters ongoing work. When advice entered without an explicit request, employees reported a weaker authorship signal, which was indirectly associated with lower experienced responsibility for the call. The cognitive sense of agency literature (Berberian et al., 2012; Frith, 2014; Limerick et al., 2014; Pagliari et al., 2022) provides a suitable lens to study the underlying

mechanism, and the work-outcome step connects to job-design theory (Hackman & Oldham, 1976) and to Brailsford et al.'s (2025) account of responsibility attribution in human–AI tasks.

Customer-incivility research has explained performance loss through resource depletion and self-regulation costs (Andersson & Pearson, 1999; Lee & Kim, 2022; Porath et al., 2015; Sliter et al., 2012; Sommovigo et al., 2020). Help-seeking research has analyzed the access cost of asking for help, mostly in inter-personal settings (Harari et al., 2022; Karabenick, 2003). We connect these streams by conceptualizing advice access as coordination work whose allocation becomes more consequential as interaction demands rise. This reframes part of incivility's performance cost as a help-access problem and identifies partial automation of initiation as one way of redistributing that access work.

Within this instrumental pathway, we also extend reliance research beyond calibration to advice quality (Berger et al., 2021; He et al., 2023; Logg et al., 2019; Lu et al., 2024; Schemmer et al., 2023). Initiation design shapes the opportunity to incorporate assistance into task execution by determining whether employees must access assistance themselves. Reliance is therefore not only a response to whether advice is trusted, but also to how and under what interaction demands it becomes available. Our reliance pathway shows that part of the difference in self-reported AI reliance between conditions is affected by delivery mode, not by quality differences in the advice itself. We, therefore, suggest that reliance could additionally be studied from a delivery perspective.

Practical Implications

Our findings have important implications for managers and organizations that employ customer service operations and system designers. Organizations choosing AI assistance are not simply choosing between automation and augmentation. They are deciding which steps within an augmented task to automate. The headline implication is that the reactive-versus-proactive choice has to be made actively rather than inherited from a vendor default. Our results show that advice initiation carries a real trade-off between employee agency and incivility, which can be conditional on task performance, so an organization adopting AI assistance cannot avoid taking a position on how advice should be initiated. The second implication is to match that position to the deployment context. In service settings where customer incivility is a frequent condition of the work (Porath et al., 2015; Sliter et al., 2010), proactive advice is more defensible because it removes the help-initiation step at the moment it is hardest. Where employee discretion and ownership of the work are part of the service value, reactive advice is more defensible because it preserves the authorship signal that ties the work to the employee.

Limitations and Future Research

The design of the study has several limitations suggesting potential for future research. First, we use a single service scenario with a synthetic customer and a one-call horizon, so the trade-off should be tested across richer service tasks and across longer or repeated interactions; longer and repeated calls would also show whether the agency-erosion effect of proactive advice cumulates over a shift, and a field replication in a live contact center would test whether these patterns hold outside a synthetic environment.

Second, task performance is currently rated by LLM judges only. While for procedural dialogues as applied in this experiment LLMs show promise (Zhang et al., 2024), human ratings could still serve as a confirming metric and robustness check. Future studies could evaluate how different LLM and human raters affect outcome evaluation.

Third, self-reported reliance and behavioral uptake capture complementary aspects of the instrumental pathway, but neither directly observes when a particular advice item changes employee behavior. For unexposed opportunities, behavioral uptake indicates consistency with potential advice rather than observed use. Future work could combine more UI telemetry to capture more measures, which can provide detailed insight into behavioral aspects. The same holds for cognitive load, which captures the retrospective post-task self report, which does no translate to individual level episodes. Adding event-level measurse could examine that process more directly.

Fourth, initiation locus and advice frequency remain partly coupled in our implementation, because proactive designs in practice tend to surface more advice turns than reactive ones. A frequency-controlled

design that holds advice exposure constant while varying initiation logic would isolate the effect of initiation from the volume of advice received, and would in turn support a test of adaptive systems that switch between reactive and proactive modes depending on the conversational state, preserving agency in calm interactions while protecting performance in hostile ones. Because customer civility was condition-consistent from the outset, the experiment does not test when an adaptive system should switch initiation modes as hostility emerges. A future design could manipulate such a within-call transition while holding advice exposure constant.

Fifth, the sample is US-based with prior customer-facing experience, so generalization to other labor markets and to specific customer service workers is not tested. A replication with, for instance, novice agents, employees at specific companies or workers across labor markets would establish boundary conditions for the agency pathway in particular, since authorship norms vary by role tenure and cultural context.

Acknowledgements

This work was supported by Innosuisse and is part of the project 104.954 IP-SBM.

References

Adam, M., Diebel, C., & Goutier, M. (2023). The Threatening Effect of Invoked Help from Highly Competent Intelligent Agents. *ICIS 2023 Proceedings*. https://aisel.aisnet.org/icis2023/hti/hti/9

Adam, M., Diebel, C., Goutier, M., & Benlian, A. (2024). Navigating autonomy and control in human-AI delegation: User responses to technology- versus user-invoked task allocation. *Decis. Support Syst.*, *180*(C). https://doi.org/10.1016/j.dss.2024.114193

Amazon Web Services. (2026a). *Create AI agents in Amazon Connect*. https://docs.aws.amazon.com/connect/latest/adminguide/create-ai-agents.html

Amazon Web Services. (2026b). *Search for content using agent assist*. https://docs.aws.amazon.com/connect/latest/adminguide/use-q-connect.html

Andersson, L. M., & Pearson, C. M. (1999). Tit for Tat? The Spiraling Effect of Incivility in the Workplace. *The Academy of Management Review*, *24*(3), 452–471. https://doi.org/10.2307/259136

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on Is Use: A Theoretical Framework of Delegation to and from Agentic Is Artifacts. *MIS Quarterly*, *45*(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Bennett, D., Metatla, O., Roudaut, A., & Mekler, E. D. (2023). How does HCI Understand Human Agency and Autonomy? *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, 1–18. https://doi.org/10.1145/3544548.3580651

Berberian, B., Sarrazin, J.-C., Blaye, P. L., & Haggard, P. (2012). Automation Technology and Sense of Control: A Window on Human Agency. *PLOS ONE*, *7*(3), e34075. https://doi.org/10.1371/journal.pone.0034075

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn. *Business & Information Systems Engineering*, *63*(1), 55–68.

Brailsford, J., Vetere, F., & Velloso, E. (2025). Responsibility Attribution in Human Interactions with Everyday AI Systems. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, 1–17. https://doi.org/10.1145/3706598.3713126

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work*. *The Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Chen, V., Zhu, A., Zhao, S., Mozannar, H., Sontag, D., & Talwalkar, A. (2025). Need Help? Designing Proactive AI Assistants for Programming. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, 1–18. https://doi.org/10.1145/3706598.3714002

Das Swain, V., Zhong, Q. "Joy," Parekh, J. R., Jeon, Y., Zimmermann, R., Czerwinski, M. P., Suh, J., Mishra, V., Saha, K., & Hernandez, J. (2025). AI on My Shoulder: Supporting Emotional Labor in Front-Office Roles with an LLM-based Empathetic Coworker. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, 1–29. https://doi.org/10.1145/3706598.3713705

Diebel, C., Goutier, M., Adam, M., & Benlian, A. (2025). When AI-Based Agents Are Proactive: Implications for Competence and System Satisfaction in Human–AI Collaboration. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-024-00918-y

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Frith, C. D. (2014). Action, agency and responsibility. *Neuropsychologia, Special Issue in Honor of Marc Jeannerod*, *55*, 137–142. https://doi.org/10.1016/j.neuropsychologia.2013.09.007

Genesys. (2026). *Agent assist tab*. https://all.docs.genesys.com/PEC-AD/Current/Agent/ADGoogleAgentAssist

Gordetzki, P., Blohm, I., Clegg, M., Schakols, F., & Hofstetter, R. (2026). Agency Configurations in Generative AI Ideation: How Textual and Visual Idea Concretizations Shape Idea Creativity and Ideator Effort. *Information Systems Research*. https://doi.org/10.1287/isre.2024.0952

Goutier, M., Diebel, C., Adam, M., & Benlian, A. (2024, January 3). *Proactive and Reactive Help from Intelligent Agents in Identity-Relevant Tasks*. HICSS. https://hdl.handle.net/10125/106423

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, *16*(2), 250–279. https://doi.org/10.1016/0030-5073(76)90016-7

Harari, D., Parke, M. R., & Marr, J. C. (2022). When Helping Hurts Helpers: Anticipatory versus Reactive Helping, Helper's Relative Status, and Recipient Self-Threat. *Academy of Management Journal*, *65*(6), 1954–1983. https://doi.org/10.5465/amj.2019.0049

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

He, G., Kuiper, L., & Gadiraju, U. (2023). Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, 1–18. https://doi.org/10.1145/3544548.3581025

Henkel, A. P., Bromuri, S., Iren, D., & Urovi, V. (2020). Half human, half machine – augmenting service employees with AI for interpersonal emotion regulation. *Journal of Service Management*, *31*(2), 247–265. https://doi.org/10.1108/JOSM-05-2019-0160

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology*, *28*(1), 37–58. https://doi.org/10.1016/S0361-476X(02)00012-7

Lee, K.-J., & Kim, S.-G. (2022). Structural Review of Customer Incivility and Social Competences of Frontline Employees for Prevention and Mitigation. *Journal of Quality Assurance in Hospitality & Tourism*, *23*(1), 196–216. https://doi.org/10.1080/1528008X.2020.1864564

Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: A review. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00643

Lin, X., Wang, X., Shao, B., & Taylor, J. (2024). How Chatbots Augment Human Intelligence in Customer Services: A Mixed-Methods Study. *Journal of Management Information Systems*, *41*(4), 1016–1041. https://doi.org/10.1080/07421222.2024.2415773

Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. (2023). G-Eval: NLG Evaluation using Gpt-4 with Better Human Alignment. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 2511–2522). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.153

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lu, Z., Wang, D., & Yin, M. (2024). *Does More Advice Help? The Effects of Second Opinions in AI-Assisted Decision Making* (arXiv:2401.07058). arXiv. https://doi.org/10.48550/arXiv.2401.07058

Microsoft. (2026). *Ask a question*. https://learn.microsoft.com/en-us/dynamics365/contact-center/use/use-ask-a-question

Moretto, G., Walsh, E., & Haggard, P. (2011). Experience of agency and sense of responsibility. *Consciousness and Cognition, From Dreams to Psychosis: A European Science Foundation Exploratory Workshop*, *20*(4), 1847–1854. https://doi.org/10.1016/j.concog.2011.08.014

Pagliari, M., Chambon, V., & Berberian, B. (2022). What is new with Artificial Intelligence? Human–agent interactions through the lens of social agency. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.954444

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Porath, C. L., Foulk, T., & Erez, A. (2015). How incivility hijacks performance: It robs cognitive resources, increases dysfunctional behavior, and infects team dynamics and functioning. *Organizational Dynamics*, *44*(4), 258–265. https://doi.org/10.1016/j.orgdyn.2015.09.002

Rafaeli, A., Altman, D., & Yom-Tov, G. (2019). Cognitive and Emotional Load Influence Response Time of Service Agents: A Large Scale Analysis of Chat Service Conversations. *Hawaii International Conference on System Sciences 2019 (HICSS-52)*. https://aisel.aisnet.org/hicss-52/da/service_science/4

Rafaeli, A., Erez, A., Ravid, S., Derfler-Rozin, R., Treister, D. E., & Scheyer, R. (2012). When customers exhibit verbal aggression, employees pay cognitive costs. *Journal of Applied Psychology*, *97*(5), 931–950. https://doi.org/10.1037/a0028559

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research*, *32*(3), 736–751. https://doi.org/10.1287/isre.2021.1015

Schemmer, M., Bartos, A., Spitzer, P., Hemmer, P., Kühl, N., Liebschner, J., & Satzger, G. (2023). Towards Effective Human-AI Decision-Making: The Role of Human Learning in Appropriate Reliance on AI Advice. *ICIS 2023 Proceedings*. https://aisel.aisnet.org/icis2023/hti/hti/14

Schmitt, A. (2024). Ensuring Human Agency: A Design Pathway to Human-AI Interaction. *ICIS 2024 Proceedings*. https://aisel.aisnet.org/icis2024/humtechinter/humtechinter/6

Sellen, A., & Horvitz, E. (2024). The Rise of the AI Co-Pilot: Lessons for Design from Aviation and Beyond. *Commun. ACM*, *67*(7), 18–23. https://doi.org/10.1145/3637865

Sliter, M., Jex, S., Wolford, K., & McInnerney, J. (2010). How rude! Emotional labor as a mediator between customer incivility and employee outcomes. *Journal of Occupational Health Psychology*, *15*(4), 468–481. https://doi.org/10.1037/a0020723

Sliter, M., Sliter, K., & Jex, S. (2012). The employee as a punching bag: The effect of multiple sources of incivility on employee withdrawal behavior and sales performance. *Journal of Organizational Behavior*, *33*(1), 121–139. https://doi.org/10.1002/job.767

Sommovigo, V., Setti, I., O' Shea, D., & Argentero, P. (2020). Investigating employees' emotional and cognitive reactions to customer mistreatment: An experimental study. *European Journal of Work and Organizational Psychology*, *29*(5), 707–727. https://doi.org/10.1080/1359432X.2020.1745189

Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01552

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2023). Voice-based AI in call center customer service: A natural field experiment. *Production and Operations Management*, *32*(4), 1002–1018. https://doi.org/10.1111/poms.13953

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, 1–21. https://doi.org/10.1145/3544548.3581388

Zhang, C., D'Haro, L. F., Chen, Y., Zhang, M., & Li, H. (2024). A Comprehensive Analysis of the Effectiveness of Large Language Models as Automatic Dialogue Evaluators. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(17), Article 17. https://doi.org/10.1609/aaai.v38i17.29923

Zhang, S., & Narayandas, D. (2026). Engaging Customers with AI in Online Chats: Evidence from a Randomized Field Experiment. *Management Science*, *72*(1), 73–95. https://doi.org/10.1287/mnsc.2022.03920

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-bench and Chatbot Arena. *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23*, 46595–46623.