University of St.Gallen
Institute of Information Systems
and Digital Business

SPITCH

Helping Agents Handle Challenging Calls in Contact Centers

A Field Study with AI Agent Assist

University of St.Gallen
Institute of Information Systems
and Digital Business

Helping Agents Handle Challenging Calls

An Independent Field Study with AI Agent Assist

University of St. Gallen

2026-06-05

Abstract

Contact center work is becoming increasingly complex, requiring agents to manage growing information loads, rising customer expectations, and emotionally demanding interactions while maintaining efficiency and service quality. AI-assisted support has emerged as one possible approach to reducing cognitive burden in these environments.

This field-study report presents findings from independent research conducted by the University of St.Gallen on proactive AI assistance during live customer interactions. The research evaluated Spitch's Agent Assist in operational customer-service settings, while research design, analysis, and interpretation remained the responsibility of the authors.

We tested the system inside the live workflow of twelve experienced agents at a bank and a health insurer, with a reflective interview after each session. Two controlled studies tested the same mechanism and quantified on what the field study surfaced.

Four findings stand out: First, on calm calls, agents respond more immediately instead of placing customers on hold to search for information. Second, on difficult calls, proactive AI assistance reduces cognitive load by taking over factual retrieval tasks, allowing agents to focus more effectively on the customer interaction itself. Agents reported feeling more composed and confident in their responses. Third, the benefits appear strongest for new hires, whose first difficult calls are especially challenging, whereas experienced colleagues often already possess the relevant knowledge. Fourth, effective assistance depends on company-specific knowledge bases and careful interface design. Generic suggestions were not considered sufficiently relevant for real customer interactions. Agents preferred concisely visualized, controllable and easily dismissible factual support integrated into their workflow.

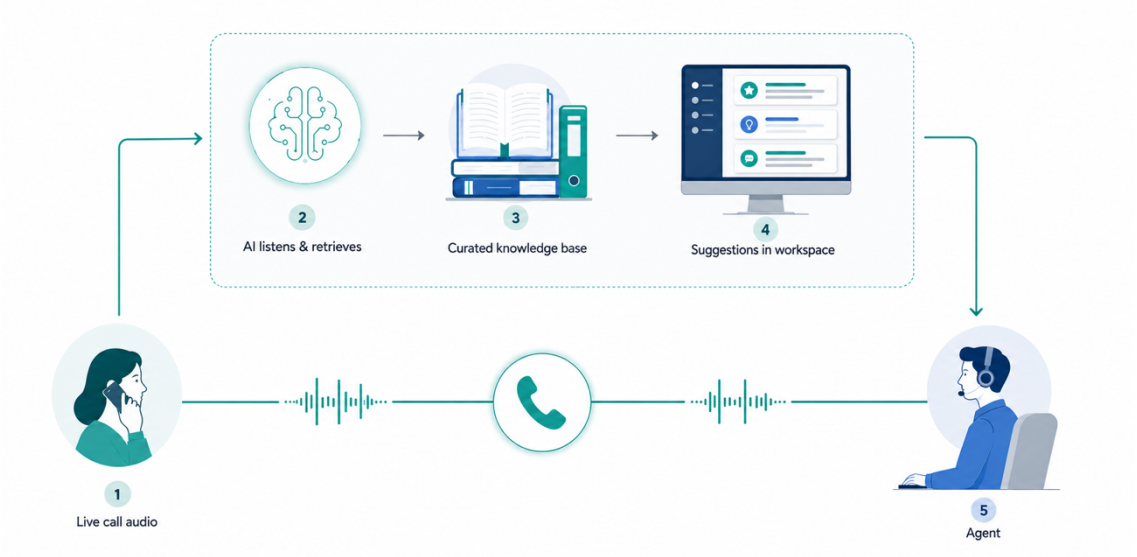
1. The compounding cost of customer incivility

Customer incivility is no longer an edge case in contact center work. Agents routinely face rudeness, dismissiveness, and outright aggression as part of their daily work. Research in service settings has documented these costs for more than two decades. Rude customers impair task performance, require additional emotion-regulation effort, and increase the risk of burnout and turnover over time (Porath and Erez 2007; Sliter et al. 2010; Grandey, Dickter, and Sin 2004). One of the agents we interviewed put the workplace side of this in her own words.

> "You notice that people are becoming cheekier. They know you're only the front line and that you can't change the situation, but you're the first point of contact, so of course you become the punching bag."
> — Agent 5

The cost of this shift is concentrated in a particular domain. When customers are rude, the agents' tasks become more difficult while they must engage in more emotional labor (Grandey, Dickter, and Sin 2004; Hochschild 2012). Reading the next step on screen, remaining composed on the line, and preventing the customer's tone from influencing their own speech all draw on the same cognitive resources. In contact center settings, this manifests as longer handling times on difficult calls and more frequent escalations, with exhaustion accumulating among agents who handle the largest share of such interactions.

Proactive AI assistance is one potential response to this problem. In this paper, the term refers to an assistant that listens to a live call, retrieves relevant information from a curated knowledge base, and surfaces it into the agent's interface in real time, based on what the customer has just said, without requiring the agent to stop and ask. The question we examine is a narrow one: When assistance of this kind of runs alongside a live call, where does it actually help, and what does that help feel like from the agent's perspective?



How AI Agent Assist works on a live call.

2. Three studies on the same question

We addressed this question through three studies. A field study evaluating Spitch's Agent Assist, run inside the live workflow of experienced agents at a bank and a health insurer, captured how the assistance was experienced by those who received it. An online study tested task performance at scale. A lab study measured the effects of the support on agents' cognitive state. Across all three studies, we varied customer tone (angry or friendly) and proactive AI assistance (present or absent), allowing the findings to build cumulatively.

The online study showed when the support helps. Proactive AI assistance improved task performance when the customer was angry. On calm calls it added little, because agents already had the required cognitive bandwidth. The positive effect appeared only when pressure pushed agents toward overload.

The lab study added a mechanism. Pupil dilation, a physiological marker of cognitive load (Sweller 1988; Paas, Gog, and Sweller 2010), decreased under AI assistance during difficult calls. The AI handles the factual load, which frees working memory for the emotion-regulation work the agent still has to do. We call this mechanism the cognitive buffer.

3. The field study

Twelve agents took part in a single session each, drawn from two partner companies. In each session, the agent took two scripted calls through the workspace they already use in production. One caller was an actor playing a friendly customer, the other an actor playing an angry one. We tailored the scripts to each company's sector and kept them clear of the routine cases agents could handle on autopilot, but the scripts still let each agent improvise inside their own style.

Spitch's AI Agent Assist solution integrated proactive AI assistance into the agent workspace. The assistant listened to the call by accessing the audio stream directly. As the customer spoke, Agent Assist retrieved relevant information from a curated knowledge base, built for each company so the content matched the sector, and surfaced the suggestions into the workspace as they became relevant. The agent did not need to query, type, or switch context to pull the information across. Agents could still use the standard guidelines and knowledge-base tools they would reach for on a normal day. The AI sat alongside the agent's existing tools as an augmentation.

Immediately after the second call, each agent sat for a 20- to 40-minute reflective interview where we discussed what happened, what helped, what did not, what they would change. Tenure in the sample ranged from about six months at the company to 28 years in customer-facing work.

We used actors on purpose. Scripting let us deliver the friendly and angry conditions reliably and on demand. Agents knew the caller was an actor, and their reflections account for that. The dynamics of a hard call still showed up. Agents described the pressure the customer put on them, the moments the conversation nearly slipped, and the effort it took to pull it back.

4. The agent side of the assistance

The twelve interviews yield five patterns. We present each along with the agent-side evidence supporting it. We translated the quotes from German and kept speakers anonymous.

4.1 Live answers and a handhold when the thread drops

The most concrete benefit agents named is saved time. When the AI surfaces the right information while the customer is still on the line, the agent answers right away, without putting the call on hold or jumping into another system.

> "[Right away] the answer arrives and I can give it straight to the customer. I don't have to open other systems or put the customer on hold."
> — Agent 1

The same effect appears in another form on harder calls. The agent loses the thread under load. Working memory drops, the customer has piled up several requests, and the agent has to recover. Agents described the AI as the cue that helped them grab the thread back.

> "Usually you just drop out, have a little blackout – and then as a support, I think, it's really quite helpful."
> – Agent 9

Both patterns share a mechanism. The information sits ready in the workspace and the agent reaches for it without breaking out of the call. On calm calls that means time saved. On hard calls it means a handhold for an agent who has briefly lost their place.

4.2 Task support creates room for emotional work without replacing it

The obvious assumption about an AI tool in an emotionally demanding job is that it should help with the emotions. Agents did not want it to. They pushed the assistant onto the factual side and kept the emotional work for themselves. The mechanism they described was indirect. If the tool handles the factual load, the agent has headroom to stay composed.

> "If I know the factual side is taken care of, that I can rely on what pops up, then I'm automatically more confident. And if that's already handled, the emotional side also gets a bit more regulated."
> — Agent 3

> "Information would be more helpful for me in that moment. If the customer is already getting nasty, you try to give them as much information as possible, which satisfies them, and if you have that ready in advance, you don't have the stress of only searching."
> — Agent 2

This shapes where the assistance fits. Several agents rejected an AI that interprets or responds to the customer's emotions (Grandey, Dickter, and Sin 2004). A minority were open to a companion-style emotional-support feature, so the preference was not unanimous. The view most agents supported was the narrower one. The assistant should help by taking factual work off the agent's plate, leaving the human free for the part of the job only a human can do.

4.3 Onboarding is where the assistance helps most, and where accuracy matters most

The sharpest positive signal came from the two newest agents in the sample. They described the assistant as most valuable for people just starting out. In the first months on the job, when every difficult call still feels unfamiliar, the assistant takes some of the weight off.

> "The first few nasty calls were worse than the 20th or the 100th."
> — Agent 12

> "At the start it's really helpful. Especially at the start, when you're constantly running into new things."
> — Agent 12

The complication came from a senior agent. Her warning was specific. A new hire, lacking the internal filter that comes from experience, will read imperfect AI output aloud to the customer.

> "Someone who's just starting types in questions and says whatever comes up. The same would be true if the information came in on its own. It really has to be 100 percent correct. That's the difficulty."
> — Agent 3

Both points hold. The group that benefits most is also the group least able to catch the assistance when it gets something wrong. Accuracy, source transparency, and clear uncertainty signals therefore become important requirements for a rollout. Other augmented-onboarding programs report the same tradeoff (Reinhard et al. 2024). The assistance helps new hires faster than anyone else, and in the same conditions it can make them wrong faster. This needs careful oversight. Governance and uncertainty signaling are particularly important for new hires.

4.4 The shape agents want is small by default, expandable on demand

Both companies' agents agreed on what the assistance should look like. Agents wanted a small, always-visible cue with the key information, that opens up on a click and only steps forward when the call gets hard. They did not always want full paragraphs on screen. They also did not want a reactive "ask and wait" assistant. Typing a good question while a customer is escalating only adds to the load.

> "Just one and a half bullet points. A sentence would already be too much."
> — Agent 3

Two concrete features came up across multiple agents. One was a way to skip or dismiss a suggestion ("I've seen this, don't keep showing it"). Repetition was one of the most-named pain points, and the fix sits on the design side. When the same suggestion keeps reappearing, agents stop reading the feed at all and waste attention checking whether something new has come in. The other was a clearer visual hierarchy. One of the most senior agents specifically called out color and layout cues for information as clearly useful. Neither feature requires a redesign of the assistant. Both depend on the assistant being properly integrated into the company's workflows based on close collaboration between agents and developers.

4.5 Presentation has to track the agent's load, including the assistance itself

Hard calls take up the agent's attention, and that makes long text on screen a problem rather than a help. Agents described this as a general feature of difficult conversations. Once a customer escalates, the agent can no longer read and listen in parallel. Reference documents, on-screen checklists, and paragraph-length guidance all hold the same content as before. The capacity to take it in is gone.

The same logic applies to proactive AI assistance. On calm calls, the same feed of suggestions was genuinely useful. On difficult ones, it became a second source of load.

> "With the first customer she pushed me hard and I could barely keep up with reading through the different bullets. In the second case, with the friendlier one, it was more pleasant, I could skim the bullets, and it was actually really good."
> — Agent 4

> "I had no time at all to read the text and at the same time hear what the customer is saying."
> — Agent 6

This is the cognitive-buffer mechanism in the agents' own words. When working memory is already taken up by managing tone and content at once, there is nothing left for reading, and putting more text on screen at that moment makes the situation worse. Agent 10 summed it up this way: "You read and at the same time you listen to the customer, and then it's too much at once."

The design consequence is direct. How much text the agent can read depends on the moment more than on the content. The assistance has to adjust to the agent's current load, showing a short cue when the call is hard and expanding to detail when the call is calm. Getting that contrast right is what keeps the tool useful in the moments where it most needs to help. How the suggestions look on the screen has to be worked out together with the agents who will use them.

5. What online and lab studies showed

The field study captured where the assistance helped from the agents' side. The two controlled studies put the same mechanism to the test and attached numbers to it. Both used the same scripted service case and varied whether proactive AI suggestions appeared during the call, which let the effect show up clearly under controlled conditions, without the noise of a live workflow.

What we measured

Two controlled studies, same scripted case:
online ($N = 103$) and lab ($N = 91$, physiological). Directions are for difficult (angry) calls.

Measure	How measured	Study	Under AI
Task performance	Scored task outcome	Online, Lab	Higher ↑
Cognitive load	NASA-TLX (self-report)	Online	Lower ↓
Cognitive load	Pupil dilation (physiological)	Lab	Lower ↓
Momentary anxiety	Self-report	Online	Lower ↓
Autonomic arousal	Skin conductance / EDA	Lab	Lower ↓
Call duration	Elapsed call time	Lab	Shorter ↓

On friendly calls, AI assistance made no reliable difference on any measure. The drop in anxiety ran entirely through lower cognitive load, not through a separate emotional channel.

The online study put 103 participants through a browser-based version of the task, crossing AI assistance (on, off) with customer behavior (friendly, angry). AI assistance significantly improved task performance when the customer was rude, and made no reliable difference when the customer was civil. Cognitive load moved the same way, significantly lower under AI assistance on rude calls only. The two findings linked up. The drop in cognitive load carried part of the performance lift under rudeness, and the drop in momentary anxiety ran entirely through that same lower-load channel. The emotional benefit came as an effect of easier factual work. The tool had no separate emotional effect of its own.

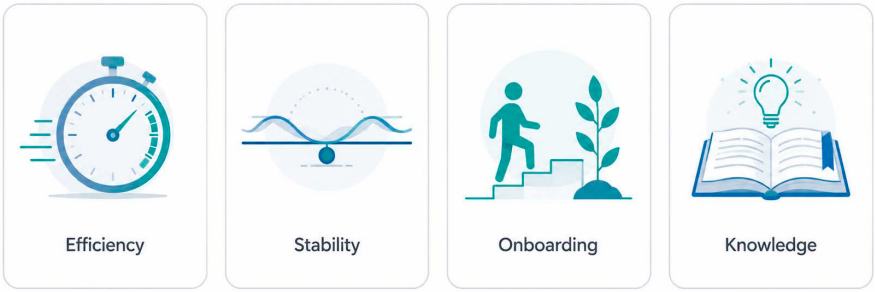
The lab study added physiological measurement. Ninety-one participants ran the same case under the "angry" script only, with or without proactive AI suggestions. The pupil widens when the brain is working harder, so pupil dilation tracks cognitive load from moment to moment (Beatty and Lucero-Wagoner 2000; Wel and Steenbergen 2018). Skin conductance picks up tiny shifts in electrical conductivity driven by sweat-gland activity, which rise under stress, and it tracks autonomic arousal (Boucsein 2012). Pupil dilation was significantly lower under AI assistance, matching the field reports with an objective signal. Task performance was significantly

higher and calls ran shorter. The same pattern showed up in the autonomic data. Cognitive load and arousal both fell.

Taken together, the two studies put the patterns the agents described on a measured, physiological base. The value sits where pressure and support meet, and the mechanism shows up in a signal that does not depend on what people say about themselves.

6. Key takeaways

The five field-study patterns, set alongside the online and lab results, distill into four core benefits for agents.



Four benefits Agent Assist delivers.

Efficiency on routine calls

On calm calls, the biggest improvement is a step that does not have to happen. When the right information appears in the agent's workspace while the customer is on the line, the agent answers directly. Agents described it as "more effective" than working without the support, though they stopped short of calling it a necessity. Lab and online studies confirm that in calm cases, assistance can be useful but remains optional.

Stability in difficult calls

When a customer escalates, the agent has to read, listen, regulate their own tone, and respond at the same time. The AI taking over the factual load creates room for the rest. Cognitive load drops under support on hard calls, performance lifts, and agents describe themselves as more sure of what to say and steadier on the line. The benefit concentrates where the work is hardest. On calm calls the AI adds little, because the agent already had bandwidth. Switching the assistance on for flagged queues, or once the voice signal indicates pressure, fits the pattern better than running it on every call.

Onboarding lift, conditional on accuracy

The newest agents are the group that benefits most. The first hard calls are the worst, and the assistant fills the gap that experienced colleagues already carry in their heads. The catch is accuracy. New hires lack the experiential filter to catch the system when it is wrong, and they are more likely to read imperfect output aloud.

Knowledge base and ongoing tuning

A knowledge base built for the company is what makes the suggestions specific enough to use in a real call. Generic content cannot do the same job. The surface details (when the assistant speaks up, what its cue looks like, how easy it is to dismiss) only come right through repeated rounds of feedback between agents and developers.

The studies cover task performance, cognitive load, and call duration under controlled conditions. The production KPIs that complete the picture (average handle time at scale, first-call resolution and transfer rates, customer satisfaction) sit on top of those, and are the things to track and confirm once the system is in production.

References

Beatty, Jackson, and Brennis Lucero-Wagoner. 2000. "The Pupillary System." In *Handbook of Psychophysiology, 2nd Ed*, 142–62. New York, NY, US: Cambridge University Press.

Boucsein, Wolfram. 2012. *Electrodermal Activity, 2nd Ed*. Electrodermal Activity, 2nd Ed. New York, NY, US: Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-1126-0.

Grandey, Alicia A., David N. Dickter, and Hock-Peng Sin. 2004. "The Customer Is Not Always Right: Customer Aggression and Emotion Regulation of Service Employees." *Journal of Organizational Behavior* 25 (3): 397–418. https://doi.org/10.1002/job.252.

Hochschild, Arlie Russell. 2012. *The Managed Heart: Commercialization of Human Feeling*. 1st ed. University of California Press. https://www.jstor.org/stable/10.1525/j.ctt1pn9bk.

Paas, Fred, Tamara van Gog, and John Sweller. 2010. "Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives." *Educational Psychology Review* 22 (2): 115–21. https://doi.org/10.1007/s10648-010-9133-8.

Porath, Christine L., and Amir Erez. 2007. "Does Rudeness Really Matter? The Effects of Rudeness on Task Performance and Helpfulness." *Academy of Management Journal* 50 (5): 1181–97. https://doi.org/10.5465/amj.2007.20159919.

Reinhard, Philipp, Nicolas Neis, Lisa Kolb, Dennis Wischer, Mahei Manhai Li, Axel Winkelmann, Frank Teuteberg, Ulrike Lechner, and Jan Marco Leimeister. 2024. "Augmenting Frontline Service Employee Onboarding via Hybrid Intelligence: Examining the Effects of Different Degrees of Human-GenAI Interaction." In *Design Science Research for a Resilient Future*, 384–97. https://doi.org/10.1007/978-3-031-61175-9_26.

Sliter, Michael, Steve Jex, Katherine Wolford, and Joanne McInnerney. 2010. "How Rude! Emotional Labor as a Mediator Between Customer Incivility and Employee Outcomes." *Journal of Occupational Health Psychology* 15 (4): 468–81. https://doi.org/10.1037/a0020723.
Sweller, John. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257–85. https://doi.org/10.1207/s15516709cog1202_4.

Wel, Pauline van der, and Henk van Steenbergen. 2018. "Pupil Dilation as an Index of Effort in Cognitive Control Tasks: A Review." *Psychonomic Bulletin & Review* 25 (6): 2005–15. https://doi.org/10.3758/s13423-018-1432-y.

Authors

Marc Grau is a PhD candidate and Research Associate at the Institute of Information Systems and Digital Business (IWI) at the University of St.Gallen (HSG), Switzerland. His research focuses on human–AI interaction, multi-agent systems, and proactive AI systems for customer service and organisational processes. His work explores how AI can augment human decision-making and support real-time interactions, with publications spanning generative AI architectures, AI-assisted guidance systems, and behavioural mechanisms in human-AI collaboration.

Niklas Weller is a PhD candidate and Research Associate at the Institute of Information Systems and Digital Business at the University of St.Gallen, Switzerland. His work addresses the alignment of AI systems to their organizational application context and explores fundamental limitations of AI automation. His research was presented at leading conferences across information systems and computer science, with publications touching on pluralistic AI alignment, the testing and evaluation of AI systems with synthetic data, and AI-assisted decision-making.

Kevin Schmitt is a Research Associate at the Institute of Information Systems and Digital Business at the University of St.Gallen, Switzerland. His research focuses on the enterprise adoption and scaling of generative AI, organisational structures for deploying large language models, and AI governance.

Prof. Dr. Ivo Blohm is Associate Professor of Information Systems with a focus on Business Analytics at the Institute of Information Systems and Digital Business, University of St.Gallen. His research examines how organizations leverage business analytics, data science, crowdsourcing, and digital platforms to improve decision-making, innovation, and organizational performance. He leads research initiatives in AI, Big Data, and digital transformation, collaborates extensively with industry, and is recognized internationally for his contributions to business analytics and data-driven organizations.

Support

This work was supported as innovation project 104.954 IP-SBM by Innosuisse.

Innovation project supported by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Industry collaboration

This study was conducted independently by researchers at the University of St. Gallen. The study evaluated Agent Assist, an AI-powered support module developed by Spitch, in collaboration with enterprises using the technology in operational settings. Research design, analysis, and interpretation remained the responsibility of the authors.

About Spitch

Spitch, founded in 2014 and headquartered in Switzerland, is a global provider of Collaborative Agentic AI solutions tailored for enterprise and regulated industries, leveraging extensive banking and insurance experience alongside secure deployment options. Spitch's platform and AI Agents combine the power of automation with a human-in-the-loop approach, ensuring people remain in control while AI enhances performance and decision-making, thereby relieving them of routine tasks. Spitch drives AI-powered digital transformation for contact centers, delivering end-to-end solutions, strategic consulting, and professional services.

Contacts

University of St.Gallen
Institute of Information Systems
and Digital Business

University of St.Gallen (HSG)
Institute of Informaton Systems and Digital Business
Müller-Friedberg-Strasse 8
CH-9000 St.Gallen
info@unisg.ch

SPITCH

Spitch AG
Stockerstrasse 10
CH-8002 Zürich
info@spitch.ch
www.spitch.ch.