Ethnographic Perspectives on AI Chatbots in Customer Support

DAVID RHEAMS, *Atlassian*

This study on the integration of artificial intelligence (AI) and large language models (LLMs) presents key findings for the effective and ethical deployment of AI in organizations. It presents original research on the deployment of an AI-powered assistant at Atlassian that was designed to support engineers by summarizing technical documentation and generating example customer responses. Using ethnographic methods and computational analysis, the study examined the nuanced impacts of AI on labor dynamics, efficiency, and empathetic communication within technical support roles. Findings include the importance of understanding the socio-technical environment of AI integration and the ethical considerations associated with AI deployment in customer service settings.

1. Introduction

In March 2023, Atlassian's customer service team began to explore artificial intelligence (AI) and large language models (LLM) chatbot prototypes to reduce the effort required to solve technical customer support issues. The goal was to provide support engineers (SEs) with an AI-powered assistant that summarized and consolidated complex technical documentation while providing an example customer response. My role as a qualitative researcher was to help a team of developers and data scientists build the prototype AI assistant.[2] I designed a research plan focusing on participant observation to understand where the AI assistants fit into their daily work. However, at the time of the study, AI was perceived as a threat to customer service jobs. The newness of the technologies and even my role as an Atlassian employee advising senior leadership presented challenges to ethnographic research.

This case study breaks down the research challenges dealing with sensitive labor issues and AI technologies. It offers insights into how we developed guiding principles, conducted qualitative research, and combined ethnographic practices with quantitative analysis. This approach enabled us to delve into how our engineers use an AI assistant to bolster confidence in their writing, reveal their creativity, and deepen their empathetic communication.

The study ran through several prototype iterations from March through August 2023. By the time it concluded in August 2023, we had identified practices to guide the next stage of product development and developed a perspective on using AI in our customer service teams. Using an ethnographic method, we uncovered product requirements, discovered performance indicators, and cultivated trust with study participants.

Ethnographic practices are essential to understanding AI tools' integration into complex work environments. Conducting interviews to determine what users value and how AI made them more efficient is only one part of the process, and it is insufficient on its own. Understanding the environments we inject AI into and adopting broader, more nuanced research approaches that force questions beyond operational efficiency is crucial. Experimenting with AI assistants in customer service environments is not as low risk as it initially appears. The impact on labor, training, and knowledge structures carries significant ethical and power implications for technical support teams. Writing off customer service as low-risk is not only shortsighted but unethical, given the profound effects AI can have on these workers' roles and livelihoods.

1.1 Customer Service as a Study Site for Organizational Ethnography

Customer service teams within software companies have broad perspectives because the role requires day-to-day interaction with customers, engineers, and management. Those teams capture insights into customer behavior, and they work to fill the gap between how the software works and how customers expect it to work. They are experts in a particular domain, with a deep knowledge of a company's systems and technical architecture. The result is a profession that blends the ability to communicate with a customer, understands abstract business problems, and has technical knowledge to identify software malfunctions. The breadth of their daily work makes them ideal candidates for organizational ethnography, as they encounter customer and coworker requests.

The unique technical support position within a company is one reason organizational anthropologists have studied technical support teams for decades. Julian Orr's *Talking about Machine*s identified technical support as a unique study site in 1990 when technical support was mostly field technicians going into physical business to fix copiers or mainframes. Orr describes the environment of the field technician this way: "The work of technical service involves the community of technicians, the community of users, and their respective corporate entities in addition to the machines, and it occurs in a public arena, the customer's place of business" (Orr 1996, 3). Each of the communities Orr identified was important as we considered how using an AI tool might impact customers and engineers, so we focused on an inclusive approach to capture the voices of these communities in our work.

We were especially attuned to the public discussions around AI, and we recognized that support engineers (like many others in the tech industry) were

sensitive to labor issues in an industry rife with recent layoffs (Lee 2024). If we wanted an accurate picture of how support engineers reacted to an AI assistant, we needed to ensure we were building tools designed to augment their work, not replace it. Most importantly, we required the research participants to understand our goal so that they could speak honestly about AI tools.

2. Study Design

We provided access to the prototype artificial intelligence assistant to 30 study participants. The program team invited participants into a private Slack channel where they interacted with the prototype by typing a message. A Slack automation delivered a response to the SE inside a threaded chat message. However, the prototype did not allow for a ChatGPT-style back-and-forth conversation. Instead, the bot provided answers in four steps inside the chat thread. First, it restated the question to ensure alignment. Second, it provided a short list of summarized sources to formulate the response. Third, it drafted a reply to the question, and fourth, it allowed the support engineer to provide feedback on accuracy. For example, when a support engineer asked, "What prevents a customer from editing a custom field in Jira?" the chatbot replied by restating the question, listing relevant technical documentation, and drafting a response.

The chatbot aimed to save time by eliminating lengthy searches on document repositories and public websites. Additionally, the chatbot could write simple SQL queries or code snippets to resolve issues. After reviewing the answer, the support engineer indicated the accuracy of the response with a *thumbs-up* or *thumbs-down* emoji.

These interactions generated detailed quantitative data on accuracy and time efficiency. We tracked the percentage of tickets where the AI assistant was used and the number of accurate versus inaccurate rankings to build insights on potential time-savings for support engineers. The research team also had access to every ticket that the support engineers worked on and flagged AI-assisted tickets. This extensive tracking produced a wealth of data illustrating when and how engineers interacted with the AI assistant. However, metrics alone cannot tell the whole story of adoption, SE happiness, efficiency, or accuracy. To capture the detailed experiences of support engineers interacting with the AI assistant, we employed ethnographic methods to uncover the nuances of these interactions, working with the 'thick descriptions' of anthropology (Geertz 1973).

2.1. Using Mixed Methods to Set a Baseline

Building thick descriptions requires a nuance and context of social interactions and behaviors. We were as interested in the subjective experiences, emotions, and attitudes towards the AI assistant as we were in the *time-to-resolution* of a support ticket or statistical accuracy. The team quickly recognized that we needed more contextual knowledge of the support engineers' daily work habits, training, and team interactions. Recognizing this gap, we understood that while participant observation and interviews would be the cornerstone of our qualitative research, preliminary work was necessary to build context and better understand the support engineer's specific environment. As we built the qualitative research plan, we arrived at a mixed methods approach, using techniques from both digital and organizational ethnographic practices. Working from Sam Ladner's *Mixed Methods* (Ladner 2019)*,* where she builds on Bryman's (2006) ideas, we combined questionnaires, content analysis, participant observation, and in-depth interviews to build a comprehensive picture of the support engineer's workday.

We used questionnaires three times throughout the study (approximately one per month) to ask contextual questions before we started interviews to establish a baseline for people's perspectives on AI. The questions were a mixture of Likert and open-text, though we found the open-text fields to contain the most helpful information. We extracted anonymous quotes from these text fields and used them to expand the purview of our interview questions.

To further expand our context, we conducted a content analysis of 100 completed support tickets where the engineer used the AI assistant early in the research program. Krippendorff's (2018) comprehensive framework for content analysis guided the approach to systematically examining the textual data from customer support tickets. The analysis compared two components: The original text proposed by the AI assistant versus the text used by the support engineer in the ticket. We reviewed each ticket to identify linguistic patterns between human and AI text.[3]

The content analysis provided context for participant observation and allowed us to validate or disprove early theories. In combination with the ticket analysis, I also assessed the AI-generated responses by observing support engineers (SEs) interacting with the AI chatbot. This approach enabled the identification of patterns in SE behavior, specifically how they selectively incorporated beneficial aspects of the AI's suggestions while disregarding irrelevant or inaccurate parts. Consequently, this process allowed for the identification and categorization of the response

elements that were most valuable to the SEs, helping to corroborate other data sources and develop a vocabulary for coding interviews.

2.2. Interviews & Observation

Our data collection combined primary methods: semi-structured interviews and participant observation through shadowing sessions. Each session was conducted in a recorded Zoom meeting, though all participants were accustomed to Zoom as Atlassian is a fully remote workplace. The interviews were designed with structured questions but kept a conversational tone, allowing for adaptive variations of our questions to fit the natural flow of each conversation, making each interview slightly different. The tactic allowed the conversation to cover multiple topics, drawing out nuances we may have missed otherwise.

We accumulated approximately 17 hours of recorded sessions. While that is quite a bit of content for two researchers to collect and analyze over 12 weeks, we wanted to ensure we had participants from as many countries and Atlassian tenures as possible. To make the work manageable, we used Zoom's recording feature so that a single researcher could conduct the interview without the burden of taking notes.

For interview transcription and data management, we used Dovetail, a research repository platform, which facilitated automated transcriptions and allowed us to highlight and tag key segments of the conversations. The transcription tagging feature was instrumental in coding the data. For example, we tagged the moment in the transcript with "confidence" when a support engineer reported that an interaction with the AI chatbot made them feel more confident in their response.

This approach provided rich contextual and behavioral data that gave quantitative results added depth and revealed contractions between reported behaviors and observed actions. During the shadowing sessions, I was able to watch for these contradictions as the support engineers navigated their daily tasks— resolving tickets, responding to customer inquiries, and troubleshooting issues. My role was to maintain a minimal presence during observation, though I did encourage the support engineers to articulate their decision-making processes and describe their actions in detail as they went through their work.

We held shadowing sessions before the interviews so that I could ask follow-up questions during the interview session. The combination of recorded interviews and observation sessions allowed us to capture detail and nuance that would have been lost otherwise. However, using these features required substantial trust from our support engineer colleagues. One of the factors that made our participants willing to sit down, invite us into their work, and share potentially sensitive information in a recorded Zoom session was the clear guiding principles we shared with participants.

3. Guiding Principles

The program team prioritized writing guiding principles and incorporating them into our documents and practices at the start of the project. We created these principles to ensure we agreed to a specific set of ethical practices, and we wanted to build trust with every team we interacted with at Atlassian. We took a lesson from *Ethnography: Principles in Practice,* emphasizing the importance of building trust with participants and navigating the complexities of hierarchy and team dynamics (Hammersley and Atkinson 2019). We also worked with our internal Responsible Technology team to help craft specific principles for our research project.[4] Establishing these principles early and sharing them with our counterparts in other departments was one of the significant early successes of the program team. We aligned on the following principles: privacy, transparency, and reduction of the potential for bias.

3.1. Support Engineer Privacy

While the program team could not keep all study participants completely anonymous in a corporate environment, we were willing to share results with our internal colleagues. For instance, we removed individual and team names from public reporting on quantitative data. We took the same approach for recorded interviews and shadowing sessions (Murphy, Jerolmack, and Smith 2021, 47). We recorded the sessions but kept them stored on a platform with limited access (managers for study participants did not have access to recorded interviews). Prior to each session, we let the participant know that they were being recorded and that we required their explicit permission before quoting them. In each document, we also clarified that we were studying a feature or system, not an individual's skill. At the start of every recorded session, I stressed this aspect of our research to let the SE know that the conversation was private and would not appear in a performance review. We obtained an agreement from managers that non-participation in the study would not reflect poorly on a support engineer.[5]

3.2. Transparency

Along with allowing users to opt out of the study and limiting direct attribution to their interviews, the program team consciously decided to be as transparent as possible at every study step. Lack of accountability is a pervasive issue in AI systems (O'Neil 2016). Algorithms can create feedback loops that reinforce inequality and operate without transparency. In customer support, AI-driven decisions can be

opaque, making it difficult for customers to challenge or understand the outcomes. This lack of accountability can erode trust in AI systems. It also aligns with one of Atlassian's core values: "Open Company, No Bullshit." While no artificial intelligence system can ever be fully transparent, we strove to clarify how we applied AI and machine learning technologies, that our research process and findings were public, and that the tool provided as much transparency as possible.

The program team believed that we had an obligation to train people on how LLMs processed information to democratize access to the latest AI technologies and make decisions on how to incorporate them into their daily workflows. We created training videos and technical documentation that explain the prototype AI assistant and how to query a large language model to make the technology more approachable and less anxiety-provoking for new users. We hosted group training sessions on Zoom so that our support engineers could ask questions of the developers and researchers working on the AI assistant. The training sessions had the added benefit of improving the chances of a support engineer using the AI assistant. The program team received overwhelmingly positive feedback from these training sessions and videos.

In addition to training, we wanted to ensure transparency in the research methods and processes. Our process documentation and anonymized research results were open to anyone in our customer support teams.[6] We provided clear guidance on the actions the program team was tracking and how we were analyzing results. For instance, we tracked every ticket the SEs worked on during the study and almost every click and input on the prototype AI assistant. Each action was tied back to an individual user but anonymized in our reporting. To build trust, we let the SEs know which performance indicators we were tracking and why we decided to track them.

However, the team recognized that tracking performance metrics may make support engineers nervous in an environment that studies AI. Being transparent with our research participants meant acknowledging that AI has the potential to radically transform the way we work and making sure that the topic of labor was a critical part of the discussion. We encouraged SEs to share concerns or ideas in interviews and questionnaires so that we could address these questions and bring them into public company conversations.

3.3. Reduce Bias

Another way we sought to build trust with our teams was to reduce the potential for bias in our research. The past decade has yielded research and warnings that even seemingly small or insignificant product decisions may have an outsized impact on

the people who use the software (Gebru 2020; Buolamwini and Gebru 2018; Broussard 2018; Broussard 2024). The research program was assigned support teams for the prototype. We divided the teams into control and participant groups so that we could measure changes in behavior.

To account for potential biases, we included teams from multiple countries, people at different stages of their careers, and people with different levels of comfort with AI. We worked with support engineering managers and team leaders to ensure we built the study to incorporate as many perspectives as possible. We ended up with a group that spoke at least seven different languages from five continents. Some of our study participants were early adopters of technology, and others were less interested in artificial intelligence. Participants were also allowed to opt out of the study for any reason on a 'no questions asked' basis so that we could ensure that all employees were willing participants.

We worked with the data scientists on the project to determine the number of study participants and the length of time we needed to ensure statistically significant results for the quantitative A/B test. In an ideal world, the research team would have spent as much time with our control group (without access to the prototype) as we did with the group testing the AI assistant. Additionally, observing meetings where our control and test groups interacted with each other would have yielded context for conducting interviews and shadowing sessions. However, with limited time and a small team, we had to choose where to apply our efforts without risking the guiding principles we established.

When I spoke with the study participants, I built rapport by telling them about my background in technical support and the goal of the study. However, as I was speaking with study participants regularly, both in person and on Slack, I was concerned that I would inadvertently signal my positions on artificial intelligence, Atlassian products, or company goals and influence the participants. To alleviate my concern, I began interviews by acknowledging my position within the company and telling the study participant I was concerned about accidentally influencing their answers to interview questions. This approach opened up a conversation that I found valuable and resulted in some of our most surprising insights.

4. Outcomes

The guiding principles were one reason we achieved positive outcomes for the study. Not only did it help establish trust, but it also helped create a spirit of exploration and a willingness of our participants to try new things, give honest feedback, and help us understand their daily work life. The business goal of the study

was to determine the impact on efficiency, deliver a strategy for future AI assistant customer service tools, and gauge support engineer's happiness when using an assistant. At the end of the study, we had a series of documents that culminated in a clear strategy for the next iteration and a business case for further development in AI assistants for support engineers.

4.1 Efficiency

We collected enough quantitative data to have statistically significant results on efficiency measures and the time required to find answers to support tickets. Our results aligned somewhat with what others in the industry found in early 2023. Harvard Business School published a working paper that showed the effectiveness of AI on a customer support team. Working with Boston University and the Boston Consulting Group, the paper concluded that productivity increased for consultants who used an AI assistant by over 12%. They found that support consultants who were "below the average performance threshold" increased their productivity by 43%. While the specific efficiency gains for our cohorts of support engineers differed, the outcomes were comparable. In short, more experienced support engineers saw fewer gains in efficiency, while support engineers with only two years of Atlassian experience saw the largest bump in productivity (Fabrizio et al., 2023). Our results differed when we measured the productivity of employees with less than a year of experience supporting Atlassian products. The AI assistant was far less helpful to new hires because they needed to have the context to ask the right question and had to put additional effort into validating the assistant's response.

While we could answer the primary questions surrounding efficiency, the quantitative data alone did not explain the behavior of the study participants. The content analysis on customer support tickets found that the percentage of AI-generated text within tickets where we used the AI assistant was less than 20%. We rarely found more than a few phrases or bullet points written by AI in the final product, even when the support engineer flagged the AI answer as correct and helpful. SEs were using AI as a drafting tool rather than relying on it as the final output because writing a ticket is a source of pride for the team. While some SEs were motivated by the quality assurance reviews for Atlassian's support, most wanted to ensure that the customer response was their voice.

We broke a ticket response into four primary sections to understand where a human added the most value and where they used the AI assistant. We found that SEs typically wrote the salutation and restated the issue to ensure they understood the customer's request. There were no instances of AI contributing directly to these sections of a ticket. We found some overlap when the SE wrote directions to a

customer on how to use the product to take steps in a workaround. The SE also wrote all the following steps and salutations. When asked, SEs responded that the AI functioned as a way to write instructions and code snippets efficiently.

4.2. Empathy

Yet, when we watched how support engineers used the tool, some of our participants used AI to assist with translations. We documented three moments within the support process where AI potentially helped with translation. First, some SEs used AI to translate the customer's request if it needed clarification. Second, the SE would translate the customer's request into a more complete AI prompt by providing context or technical information. Some support engineers use AI assistants to translate concepts from their native languages into English (Atlassian's workforce and customers are global, though most customer support communication is in English).

In one instance, a support engineer from Brazil translated a phrase common in Portuguese into English to see if it lost its meaning during translation or risked being misinterpreted by the customer. The translation wasn't simply a word-for-word translation of a technical concept; it was a translation of a pun. The study participant wanted some of his personality to come through on the ticket to build a relationship with the customer. While he felt that the AI assistant's writing could not help him build rapport with a customer, they felt more confident using the AI tool as a sounding board for tricky phrases. The word *confident* surfaced in some capacity in over 80% of the interviews and shadowing sessions. While the direct text may have been discarded, the SEs did report that it helped them cross language barriers and feel better about their writing to show that they empathized with the customer's situation.

The moments where the AI helped the support engineer be more confident also sparked creativity when using the tool. When the AI could simply help consolidate documents and provide a quality summary, the SEs treated it like any other tool that helped them solve an issue. They were pleased but didn't stop and take any particular note. However, when the AI assistant was used more creatively, the support engineers stopped and noted how they were crafting customer communications.

4.3. Beyond Interviews

One of the more common refrains I've heard working on technology products is that "we must talk to the user," accompanied by a push to do user interviews. The impulse to talk to people is admirable and a practice we should continue. However,

one of the realizations I had doing this project was that simply conducting interviews is insufficient when dealing with sensitive topics like artificial intelligence and labor. Despite our early claim that we were "augmenting, not replacing" support engineers, it became increasingly apparent that the complexities of AI's impact on labor required a broad view of the technology beyond the chatbot we were testing. AI is already significantly affecting hiring and labor practices. For example, Intercom's leadership has emphasized the transformative potential of AI in customer service, predicting that AI agents will eventually handle the majority of customer queries, fundamentally changing the nature of human support roles (Adams 2024). Similarly, Microsoft's recent layoffs, attributed to strategic realignments in response to AI advancements, highlight how AI reshapes labor markets (Landymore 2024). These developments underscore that AI, much like past technological advancements such as robotics or the cotton gin, is poised to transform labor markets in unpredictable ways. Ethnographic studies on AI must employ methods that fully recognize and address the impact on labor. However, we found that we needed to create intentional spaces where our participants could freely discuss both the positive and negative aspects of AI, independent of their specific daily tasks, to uncover a broader range of perspectives.

Building a broad perspective also requires time. The additional time spent observing Slack channels, attending study participant team meetings, and having informal conversations was invaluable to the study's success. The openness about potential harms and willingness to slow down and watch participants wrestle with the new technologies provided more precise insights than we may have had otherwise.

About the Authors

David Rheams, PhD is a senior business architect at Atlassian and a lecturer at The University of Texas at Dallas in the Arts, Technology, and Emerging Communication department.

Notes

[1] Atlassian is a software company that builds tools for collaboration and productivity. Their core products are Jira, Confluence, and Trello.

[2] A program team in a software company is a cross-functional group responsible for planning, executing, and managing a project. Our team was composed of software engineers, data scientists, a program manager, customer support leaders, business analysts, and business architects.

[3] We broke the customer service tickets into their components to determine which part of the ticket was used for typical pleasantries, which parts were used for technical information, and which parts were additional customer information. We also compared sentence structure, tone, and directness between AI and human-written text.

[4] Atlassian has an amazing Responsible Technology team who I consulted with at the onset of the project. They were incredibly quick to help shape the specifics we needed for our research group and they play a vital role in development at Atlassian. They can be found here: https://www.atlassian.com/trust/responsible-tech-principles

[5] Some support engineers accepted our offer to opt out of the study because they had other projects to focus on. We risked biasing towards people who enjoyed AI technologies by letting people opt out, but the program team felt it was worth the risk.

[6] We ensured that the guiding principles were the same for both the qualitative and quantitative methods. In particular, our approach stemmed from the principle that our product was not designed to replace customer support professionals; stakeholders reiterated this principle in meetings and strategy planning.

References Cited

Adams, Paul. 2024. "There's No Going Back: AI-first Customer Service Has Arrived." Intercom. https://www.intercom.com/blog/ai-first-customer-service/.

Broussard, Meredith. 2018. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge: MIT Press.

Broussard, Meredith. 2024. *More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. Cambridge: MIT Press.

Bryman, Alan. 2006. "Integrating Quantitative and Qualitative Research: How Is It Done?" *Qualitative Research* 6 (1): 97–113.

Buolamwini, Joy and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15.

Fabrizio, Dell'Acqua. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper 24-013. Harvard Business School. https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf.

Futurism. 2023. "Microsoft's Layoffs and AI's Impact on the Labor Market." Futurism. https://www.futurism.com/microsoft-layoffs-ai-labor-market.

Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.

Gebru, Timnit. 2020. "Race and Gender." In *The Oxford Handbook of AI Ethics*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das, 122–144. Oxford: Oxford University Press.

Hammersley, Martyn, and Paul Atkinson. 2019. *Ethnography: Principles in Practice*. London: Routledge.

Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology*. 4th ed. Thousand Oaks: SAGE Publications.

Ladner, Sam. 2014. *Practical Ethnography: A Guide to Doing Ethnography in the Private Sector*. Walnut Creek: Left Coast Press.

Ladner, Sam. 2019. *Mixed Methods: A Short Guide to Applied Mixed Methods Research*.

Lee, Roger. 2024. "Layoff Tracker." Layoffs.fyi. https://layoffs.fyi.

Marcus, George E., and Michael M. J. Fischer. 2014. *Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences*. Chicago: University of Chicago Press.

Murphy, Alexandra K., Colin Jerolmack, and DeAnna Smith. 2021. "Ethnography, Data Transparency, and the Information Age." *Annual Review of Sociology* 47 (April): 41–61. https://www.annualreviews.org/doi/10.1146/annurev-soc-090320-124805.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing.

Orr, Julian E. 1996. *Talking about Machines: An Ethnography of a Modern Job*. Ithaca: Cornell University Press.