RESEARCH REPORT — NORTH ATLANTA SMALL BUSINESS SERIES

Do Website Chatbots Actually Turn Visitors Into Customers? What the Research Says

A review of controlled experiments, customer surveys, legal decisions and vendor claims on whether chatbots raise conversions, and what the evidence means for local businesses that win customers by phone, form and appointment

Prepared by Make It Loud | North Atlanta Small Business Research Series

Data Sources: Randomized field experiments published in *Marketing Science* (2019), the *Journal of Business Research* (2025) and on arXiv (2025) | *Journal of Marketing* (2022) | Gartner customer surveys (2024, 2026) | British Columbia Civil Resolution Tribunal (2024) | Utah and California bot-disclosure laws | Chat vendor statistics as compiled by JivoChat

+21.7%	−79.7%	64%	0
conversion lift from an AI pre-sale chatbot in the largest randomized test, on an online retail platform	drop in purchases when an AI sales agent said it was a bot before the conversation started	of customers would prefer that companies not use AI for customer service (Gartner, 2024)	published controlled studies we found on chatbots for local service businesses like contractors, clinics or law firms

BOTTOM LINE

Yes, a well-built chatbot can turn more visitors into customers, but the best evidence comes from large online stores, not local service businesses, and the details matter more than the widget. In the most rigorous public test, a 2023–24 randomized experiment with 44,614 shoppers on a large e-commerce platform, an AI pre-sale chatbot raised conversion by 21.7% and sales per shopper by 16.3%, and the gain was larger when the bot could hand customers to a human. Other experiments show that how a bot introduces itself, what language it speaks and whether it can resolve the problem can swing results dramatically, in both directions. Customers are wary: Gartner found 64% would rather companies not use AI for customer service, mostly because they fear losing access to a person. And the eye-catching figures chat vendors publish, from “85% more likely to convert” to “5 times more likely to buy,” compare people who chose to chat with people who didn't, so they cannot show that the chat caused the sale. For a local business, a chatbot is worth testing as a way to capture after-hours leads and answer routine questions quickly, as long as it hands off to a human, never invents answers, and is measured on booked calls and appointments rather than chats.

Executive Summary

Chat widgets and AI assistants are now standard offers from website builders, scheduling platforms and marketing software, and business owners are often told they will “double conversions.” Most of the statistics behind those claims come from the companies selling the software. Independent evidence is harder to find, and it rarely involves the kinds of businesses that make up most of the local economy: plumbers, dentists, med spas, HVAC contractors, attorneys and auto shops, whose customers usually convert by calling, filling out a form or booking an appointment rather than checking out online.

This report reviews what the public research actually shows. We focused on randomized field experiments, in which real customers are randomly assigned to see a chatbot or not, because they are the only studies that can show cause and effect. We also summarize large customer surveys, a peer-reviewed study of thousands of real chatbot conversations, a legal decision on chatbot accuracy, state

disclosure laws and the most widely repeated vendor statistics. We then translate the findings into practical guidance for local businesses.

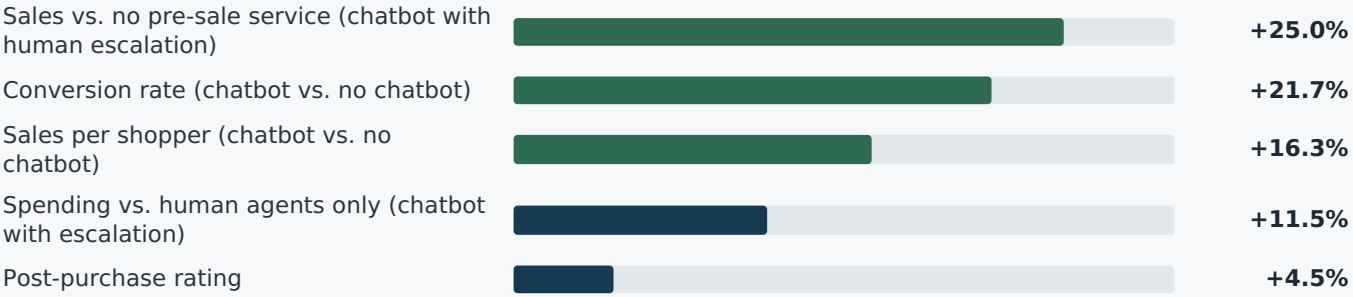
Finding 1 | The Best Controlled Test Found a 21.7% Lift in Conversion

AI pre-sale chatbot vs. no chatbot, 44,614 shoppers: conversion +21.7%, sales per shopper +16.3%, with human escalation +25%

The largest and most recent randomized test of a website sales chatbot was conducted by researchers from Columbia University and Zhejiang University with a large cross-border e-commerce platform between September 2023 and June 2024. Shoppers were randomly assigned to have access to a generative-AI pre-sale chatbot, available 24 hours a day, or not. Among 44,614 consumers, the chatbot raised the conversion rate by 21.7% and sales per shopper by 16.3%, both statistically significant. Customers' post-purchase ratings also rose about 4.5%.

Two details matter for small businesses. First, the chatbot performed best when it could escalate to a human: with escalation, sales were 25% higher than with no pre-sale service at all, and customers in that arrangement spent about 11.5% more than those served by human agents alone. Second, the gains were concentrated among newer and less experienced customers, the people who most need answers before they buy.

Effect of an AI Pre-Sale Chatbot in a Randomized Field Experiment (44,614 Shoppers)



Generative AI and Sales Productivity: Field Experiments in Online Retail (Columbia University, Zhejiang University and Zhejiang University of Finance & Economics; arXiv:2510.12049, 2025). Experiment run September 2023–June 2024 on an unnamed cross-border e-commerce platform.

Finding 2 | How a Bot Introduces Itself Can Make or Break the Sale

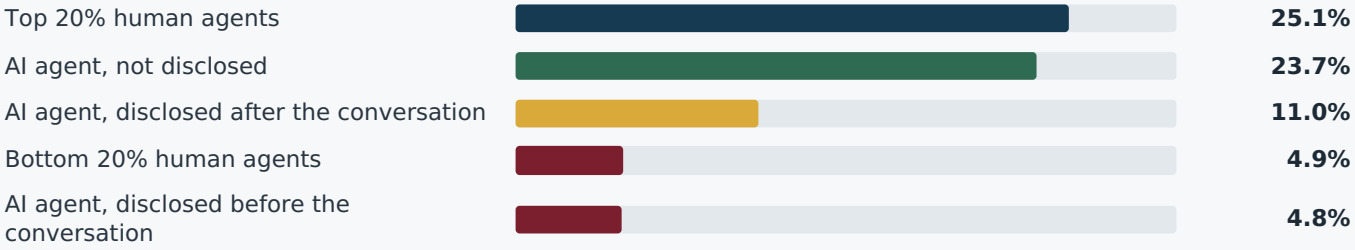
Purchase rate: undisclosed AI agent 23.7%; same agent disclosed up front 4.8% (−79.7%); top human agents 25.1%

A widely cited 2019 study in *Marketing Science* by Xueming Luo, Siliang Tong, Zheng Fang and Zhe Qu tested AI voice agents against human sales staff at a large financial services company, using 6,255 randomly assigned customers who were offered a loan renewal. When customers were not told they were speaking with a bot, the AI agent's purchase rate (23.7%) was statistically the same as the company's best human sellers (25.1%) and nearly five times that of its weakest (4.9%). But when the bot disclosed that it was AI before the conversation began, purchases fell to 4.8%, a 79.7% drop.

The study involved outbound phone calls, not website chat, so it is not a direct measure of website chatbots. Its lesson is about customer expectations: people judged the disclosed bot as less knowledgeable and less empathetic even though it performed the same. That creates a tension for businesses, because a growing number of laws require bots to identify themselves. California's B.O.T. Act, in effect since 2019, requires clear disclosure when a bot is used to encourage a sale, and Utah

requires businesses to disclose generative AI when a consumer asks, and up front in certain high-risk interactions. The practical answer is to disclose and then perform well enough to overcome the skepticism.

Purchase Rates by Agent Type in a Randomized Sales Experiment (6,255 Customers)



Luo, Tong, Fang and Qu, "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases," *Marketing Science* 38(6), 2019. Outbound loan-renewal sales calls at a large Asian fintech company.

Finding 3 | Language and Tone Matter, but So Does Getting the Basics Right

Smaller experiments show that a chatbot's design can change results substantially. In a randomized test published in the *Journal of Business Research* in 2025, a chatbot-based shopping app in India added Hindi alongside English for a random group of 3,952 new users, compared with 8,683 who saw English only. Average daily purchases rose from 0.025 to 0.058 per user, about 132%, with the gains concentrated in higher-involvement categories such as travel and local deals. But the bilingual version also saw sharply higher uninstalls, a reminder that more engagement is not always better engagement.

Adding Hindi to an English-only shopping chatbot raised daily purchases about 132%, but uninstalls rose too

A 2025 field experiment on a U.S. travel-planning site tested the tone of a GPT-4o assistant with 282 users. Users who chatted with an assistant given explicit tone instructions, whether enthusiastic or neutral, subscribed at roughly 14% to 15%, versus about 5.5% for an assistant given no tone guidance. The sample was small and the outcome was subscriptions rather than purchases, but the direction is consistent: the instructions and setup behind a chatbot matter, not just whether one exists. For metro Atlanta businesses serving Spanish-speaking customers, the language finding is especially relevant.

Finding 4 | Customers Are Wary, and a Bad Bot Can Cost You the Sale

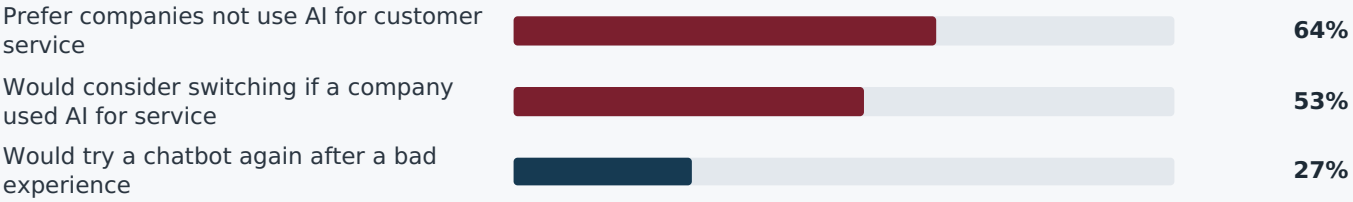
Customer attitudes are a real constraint. In a Gartner survey of 5,728 customers conducted in December 2023, 64% said they would prefer that companies not use AI for customer service, and 53% said they would consider switching to a competitor if a company did. Their top concern was that AI would make it harder to reach a person, followed by fears of job losses and wrong answers. In September 2026, Gartner reported that only 27% of customers would try a chatbot again after a negative experience.

64% of customers prefer companies not use AI for service; 53% would consider switching; only 27% would retry a chatbot after a bad experience

Peer-reviewed research shows why. A 2022 *Journal of Marketing* study by Cammy Crolic, Felipe Thomaz, Rhonda Hadi and Andrew Stephen analyzed about 35,000 real chat sessions at an international telecom company and ran follow-up experiments. When customers were already angry, humanlike chatbots lowered satisfaction, and when problems went unresolved, they hurt evaluations of the company and reduced purchase intentions, because the humanlike design raised expectations the bot

could not meet. Setting expectations clearly at the start of the chat reduced the damage.

What Customers Say About AI in Customer Service (Gartner)



Gartner press releases, July 9, 2024 (survey of 5,728 customers, December 2023) and September 2, 2026 (27% figure as reported in coverage of the release; survey details not published in the coverage reviewed).

Finding 5 | The Most-Cited Chatbot Statistics Can't Show Cause and Effect

Search for chatbot conversion statistics and the same figures appear repeatedly: customers who chat are “5 times more likely to complete a purchase” (attributed to Tidio), “85% more likely to convert” (Intercom), and live chat produces “a 40% increase in conversion rate” (ICMI), among others compiled by chat-software vendors. These numbers are usually undated, come from companies that sell chat tools, and compare visitors who chose to start a chat with visitors who did not.

Vendor claims range from “40% higher conversion” to “5 times more likely to buy,” and none are randomized tests

That comparison has a built-in bias. A visitor who opens a chat to ask about availability or pricing is usually already close to buying; many would have converted with or without the chat. Only randomized experiments, which compare similar visitors with and without access to a chatbot, can isolate the chatbot's effect. The best of those found a 21.7% lift, real and meaningful, but far smaller than “5 times.”

Claim, as commonly cited	Attributed to	What it measures	Shows causation?
Customers who use live chat are 5 times more likely to complete a purchase	Tidio	Chatters vs. non-chatters	No
Chatters are 85% more likely to convert into customers	Intercom	Chatters vs. non-chatters	No
Live chat leads to a 40% increase in conversion rate	ICMI	Company-reported results	No
AI pre-sale chatbot raised conversion 21.7%	Columbia/Zhejiang field experiment (2025)	Randomized: chatbot vs. none	Yes

Vendor claims as compiled by JivoChat, “52 Live Chat Statistics You Need to Know” (accessed November 2026); original publication years were not stated.

Finding 6 | For Local Businesses, Accuracy and Handoff Matter More Than the Widget

The research gap is clear: we found no published randomized study of chatbots on the websites of local service businesses, where most customers convert by phone call, form submission or appointment request rather than online checkout. The closest evidence comes from online retail and financial services, so results for a contractor or clinic may be larger or smaller. What carries over is the pattern: chatbots help most when they answer common questions quickly, are available after hours, serve customers who are new to the business and pass people smoothly to a human.

Moffatt v. Air Canada (2024): a company is responsible for what its chatbot tells customers

Accuracy is a business risk as well as a customer-experience issue. In *Moffatt v. Air Canada* (February 2024), British Columbia's Civil Resolution Tribunal held the airline liable for negligent misrepresentation after its website chatbot gave a customer wrong information about bereavement fares, rejecting the argument that the chatbot was a separate entity. For a local business, a bot that misquotes a price, promises a same-day appointment that does not exist or gives health or legal guidance can cost more than it earns.

What This Means for Local Businesses

1 Treat a chatbot as an experiment, not a sure thing.
Run it for 60 to 90 days and compare booked calls, form submissions and appointments with the same period before, or show it to only half of visitors. Count leads and revenue, not chats.

2 Always offer a human, and make the handoff easy.
The strongest results came from bots that could escalate to a person. Show your phone number, offer a callback and route after-hours chats to a follow-up the next morning.

3 Limit the bot to what it can answer correctly.
Load it with your real hours, service area, pricing ranges and booking rules. Have it say “let me connect you” instead of guessing, especially on price, availability, health or legal questions.

4 Be upfront that it's a bot, and set expectations.
Disclosure is increasingly required and customers are skeptical. A clear, friendly introduction that explains what the assistant can help with reduces frustration.

5 Serve the customers you actually have.
If many of your customers speak Spanish or another language, a multilingual assistant may matter more than any other feature.

The Evidence at a Glance

Study	Setting and design	Sample	Key result
Columbia/Zhejiang field experiments (arXiv, 2025)	Cross-border e-commerce platform; randomized GenAI pre-sale chatbot, 2023–24	44,614 shoppers	Conversion +21.7%; sales per shopper +16.3%; +25% with human escalation
Luo, Tong, Fang & Qu (*Marketing Science*, 2019)	Fintech outbound sales calls; randomized AI vs. human agents and disclosure timing	6,255 customers	Undisclosed AI matched top humans; disclosure before the call cut purchases 79.7%
Nayak & Aravindakshan Nair (*J. Business Research*, 2025)	Chatbot shopping app in India; randomized English vs. English + Hindi	12,635 new users	Daily purchases up about 132%; uninstalls also rose
Jirpongopas et al. (arXiv, 2025)	U.S. travel-planning site; randomized GPT-4o assistant tone	282 chat users	Subscriptions about 14–15% with tone instructions vs. 5.5% without
Crolic, Thomaz, Hadi & Stephen (*J. Marketing*, 2022)	Telecom chat logs plus experiments	~35,000 chats	Humanlike bots hurt satisfaction and purchase intent among angry customers
Gartner customer surveys (2024, 2026)	Surveys of customers	5,728 (2024)	64% prefer no AI in service; 27% would retry a bot after a bad experience

About Make It Loud

Make It Loud Digital Marketing is a Metro Atlanta digital marketing agency serving small and midsize businesses across Gwinnett, Forsyth, Hall and Cherokee counties and beyond. The agency provides search engine optimization, generative engine optimization (GEO) and AI search visibility, Google Business Profile management, website design, pay-per-click advertising, social media, podcast production and AI consulting for home service, healthcare, professional service, automotive and nonprofit clients. More research is published at makeitloud.net/research.

Methodology, Limitations & Sources

Approach. We searched academic databases, preprint servers and industry sources in November 2026 for evidence on whether website or messaging chatbots change purchase or conversion rates. We prioritized randomized field experiments with real customers, then peer-reviewed observational studies, large customer surveys and legal decisions. Figures come from the original papers and releases where available. Vendor statistics are reported as compiled by a chat-software vendor and were not independently verified.

Limitations. None of the experiments involved local service businesses, and several were conducted outside the United States, in different languages and product categories. The e-commerce experiment did not name the platform, and the travel and language studies were small or contain internal inconsistencies noted by reviewers. The 2026 Gartner figure is reported from news coverage of the release. Results from 2019 may not reflect current generative AI assistants. This report is not legal advice; businesses should review disclosure obligations with counsel.

Field experiments. “Generative AI and Sales Productivity: Field Experiments in Online Retail,” arXiv:2510.12049 (2025); Luo X., Tong S., Fang Z., Qu Z., *Marketing Science* 38(6):937–947 (2019); Nayak A., Aravindakshan Nair A., *Journal of Business Research* 190:115158 (2025); Jirpongopas L., et al., arXiv:2509.14259 (2025).

Crolic C., Thomaz F., Hadi R., Stephen A. “Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions,” *Journal of Marketing* (2022).

Gartner; JivoChat. Gartner press releases, July 9, 2024, and September 2, 2026; JivoChat, “52 Live Chat Statistics You Need to Know” (vendor compilation).

Legal. *Moffatt v. Air Canada*, 2024 BCCRT 149; California Business and Professions Code §§ 17940–17943 (B.O.T. Act, SB 1001); Utah Artificial Intelligence Policy Act as amended by SB 226 and SB 332 (2025).