LEADTRUFFLE RESEARCH

First-Reply Benchmark · September 2026

What Gets Home Service Leads to Reply?

A study of 4,699 first responses across Yelp, Thumbtack, and Google Local Services Ads

60.5%
24-hour reply rate for the pre-specified pattern, vs. 50.3% for other replies

3
Independent lead channels analyzed, all pointing the same direction

4,699
Real, redacted lead conversations in the account-capped holdout

Observational study. Not a controlled experiment.

CONTENTS

Table of contents

Click any item below to jump to that section. All charts and data tables are reproduced in full inside each section.

For journalists: section 02 (page 4) is built to be used directly, a citable headline stat ready to drop into coverage.

01 EXECUTIVE SUMMARY

The first reply should prove the request was read

We analyzed 4,699 real, redacted first-reply conversations from Yelp, Thumbtack, and Google Local Services Ads to answer a practical question: after a lead arrives, what should a contractor's first message actually do?

Across all three channels, first replies that did three specific things were associated with more customers responding within 24 hours:

- ☐ Referenced the specific job the customer asked about, not a generic phrase like "your request"
- ☐ Asked exactly one focused question, not zero and not several
- ☐ Did not immediately ask for a phone number or contact information

Messages built this way received a 60.5% reply rate within 24 hours, compared with 50.3% for other replies. After comparing results within the same contractor account, the adjusted difference was smaller but still positive on every channel checked: Yelp, Thumbtack, and Google LSA.

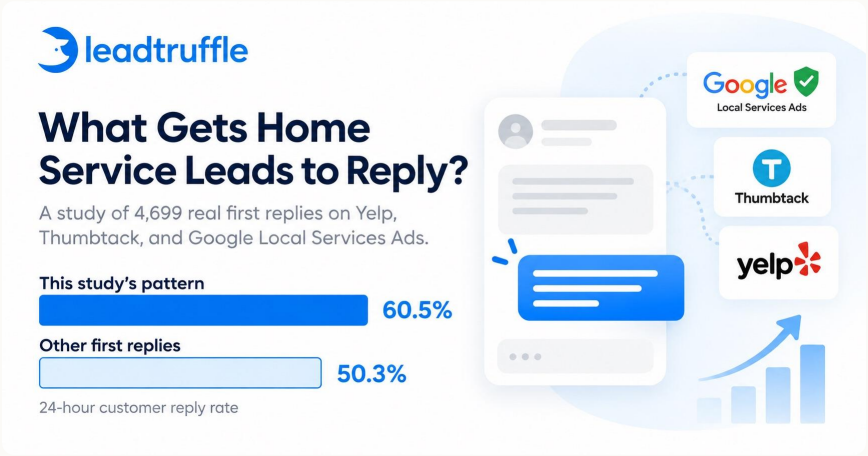
> **Why this holds up:** this composite pattern was identified in an earlier 1,000-message discovery sample, then re-tested against a much larger, separate 4,699-message holdout before publication. The direction replicated on all three channels.

This is an observational study, not a controlled experiment. It measures whether a customer replied, not whether the lead became a booked job. The confidence intervals for any single channel are wide enough to include zero. What is consistent is the direction: the same pattern was never associated with fewer replies on any channel tested.

> *"By the time a third-party lead reaches a contractor, the customer has usually already explained what they need. The first response should prove the business read that, not force the customer to repeat it or start over."*
>
> Aaron Decker, co-founder, LeadTruffle

02 KEY FINDINGS AT A GLANCE

The numbers a journalist needs



60.5% vs 50.3%

24-hour reply rate for the pre-specified pattern vs. all other first replies, pooled across all three channels

+3.2 pp

Adjusted difference after comparing message styles within the same contractor account, pooled across channels

4,699

Conversations in the deterministic, account-capped holdout: 1,760 Yelp, 1,492 Thumbtack, 1,447 Google LSA

87 accounts

Real contractor accounts represented across the three channels, capped per account to avoid one business dominating results

Adjusted lift by channel

Channel	24h reply, pattern	24h reply, rest	Pooled diff	Adjusted diff
ALL	60.5%	50.3%	+10.2 pp	+3.2 pp
Yelp	48.8%	41.3%	+7.5 pp	+1.3 pp
Thumbtack	56.8%	51.2%	+5.6 pp	+4.2 pp
Google LSA	73.8%	64.8%	+9.0 pp	+5.4 pp

"Adjusted pp" (percentage points) compares the pattern against other messages within the same contractor account and platform, so differences in account quality or speed do not get credited to the message style. It is a more conservative, more trustworthy estimate than the raw pooled difference.

Raw 24-hour reply rate: pattern vs. all other replies

Chart 1. Unadjusted 24-hour reply rate, pooled across Yelp, Thumbtack, and Google LSA, n = 4,699.

What this does and doesn't prove: the pattern is associated with more replies, on every channel tested, after adjusting for account-level differences. It does not establish that the message style caused the difference, and it does not measure bookings or revenue.

One line for a headline

"First replies that named the job, asked one clear question, and skipped the immediate phone-number ask were associated with a 60.5% 24-hour reply rate, compared with 50.3% for other replies, across 4,699 real home-service lead conversations."

03 THE RESEARCH QUESTION

Why the first reply matters more than it looks

Contractors pay for the opportunity to reach a customer on Yelp, Thumbtack, or Google Local Services Ads. Once the lead arrives, a lot depends on what happens in the first message back. A generic acknowledgment can lose a customer's attention immediately. A message that clearly shows the business read the request can keep the conversation moving.

LeadTruffle set out to answer a specific, practical question: **what first-reply construction is associated with a useful customer response on Yelp, Thumbtack, or Google Local Services Ads?**

We used two outcomes to measure this. The primary outcome is any recorded customer reply in the same conversation medium within 24 hours. A stricter, secondary outcome is a model-classified response that meaningfully advances the project or scheduling conversation, not just an acknowledgment.

What we measured

Structural features of the first reply: whether it referenced the specific job, how many questions it asked, whether it requested contact details immediately, and whether it read as templated or personal.

What we did not measure

Booking outcomes, revenue, or lead quality by platform. Booking data across this dataset was too sparse and incomplete to support a reliable claim.

> *"Generic first replies may be wasting the opportunity a paid lead represents. We wanted real data, not a guess, on what actually keeps a conversation moving."*
>
> Aaron Decker, co-founder, LeadTruffle

04 METHODOLOGY

How the study was built

The dataset

The sample is a deterministic, source-targeted, account-capped holdout drawn from real LeadTruffle contractor accounts, covering a cohort from February through August 2026 with a 30-day outcome maturity window. It excludes a conservative superset of the 1,000 conversations used in an earlier discovery study, so the holdout is a genuinely separate replication sample, not a re-analysis of the same rows.

Each contractor account was capped at 100 messages per platform so that no single business could dominate the results. The sample was filled round by round across accounts, targeting up to 2,000 messages per source, and retaining only conversations with defensible, exact linkage.

Channel	Reply medium	Linkage	Sample	Accounts
Yelp	Marketplace message	Exact provider conversation	1,760	32
Thumbtack	Marketplace message	Exact provider conversation	1,492	24
Google LSA	Email	Exact email conversation	1,447	31

> **A correction we're disclosing:** an earlier internal draft incorrectly analyzed 213 Google LSA-derived SMS rows. Google LSA replies actually travel through LeadTruffle's email gateway. The corrected Google cohort uses exact email-conversation linkage, and every Google LSA figure in this report reflects that correction.

Extraction and privacy

All extraction ran inside a read-only PostgreSQL transaction, verified before any query, with statement and idle-transaction timeouts and a single database connection. Names, phone numbers, emails, URLs, street-like addresses, and ZIP codes were redacted locally before any text was sent to a classification model. No fuzzy message or identity matching was used for platform linkage. Raw and redacted row-level data is never published; only aggregate results are tracked in this report.

04 METHODOLOGY, CONTINUED

How replies were classified and adjusted

The classifier

Every first reply was labeled by an AI classifier that received only the platform and redacted message text. It never saw whether the customer replied, when they replied, or any later message in the conversation, so labels could not be shaped by the outcome they were later compared against. A separate, second classifier later labeled the customer's actual response, using the inquiry and that first reply only, after outcomes were already extracted deterministically from the database.

Adjusting for account differences

Raw reply-rate differences can be misleading. Some contractors get better leads, respond faster, or work in more responsive trades regardless of message wording. To account for this, we compared the same message style within the same contractor account and platform, and required at least three examples of each style before a comparison counted. The adjusted numbers throughout this report come from that within-account comparison, not the raw pooled rates.

Reliability, not ground truth

The classifier's labels are model-generated measurements, not human-verified ground truth. To check how stable those labels are, an independent 180-message sample was relabeled by the same classifier under the same conditions.

Label	Agreement
Pre-specified composite pattern	86.7%
Exact question count	88.9%
Primary question type	93.9%
Asks for contact details	98.9%
Inquiry intent (exploratory only)	65.6%

What we did with low-agreement labels: inquiry intent showed materially lower test-retest agreement, so it was excluded from the primary, headline comparisons and is reported separately as exploratory context only.

05 THE PRE-SPECIFIED FINDING

One pattern, tested twice, holding up on all three channels

Before this holdout was ever examined, an earlier 1,000-message discovery study identified one actionable composite: **a specific reference to the requested work, exactly one question, and no immediate request for contact details.** That composite was locked in as the primary hypothesis before the larger 4,699-message holdout was analyzed, which is what makes this a genuine replication rather than a pattern found by searching the data after the fact.

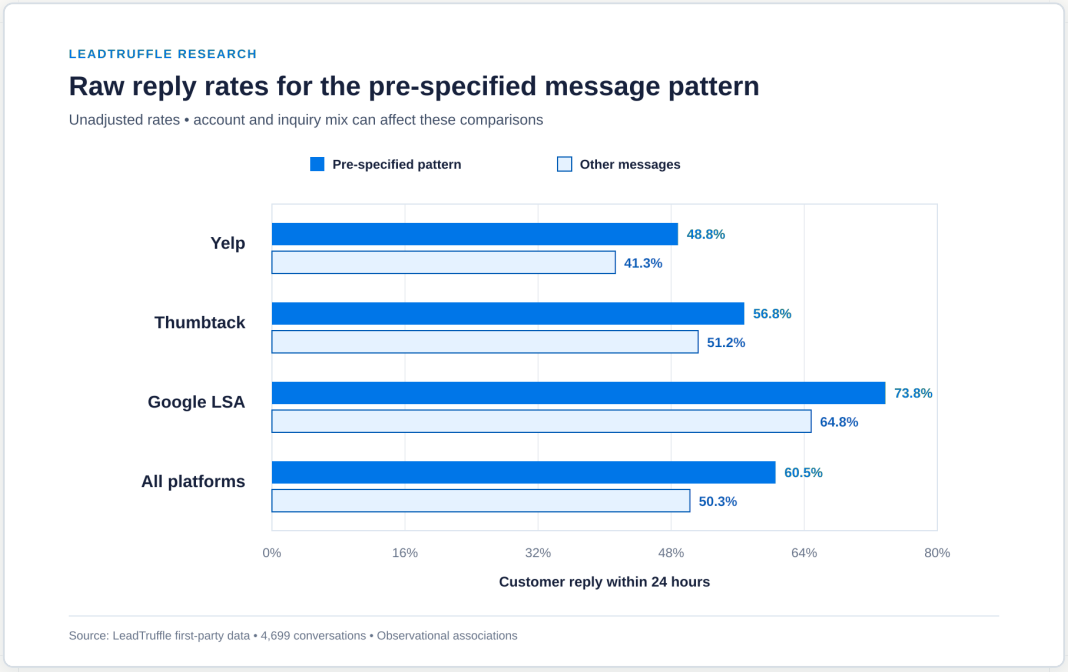


Chart 1. Unadjusted 24-hour reply rate for the pre-specified pattern versus other messages, by channel.

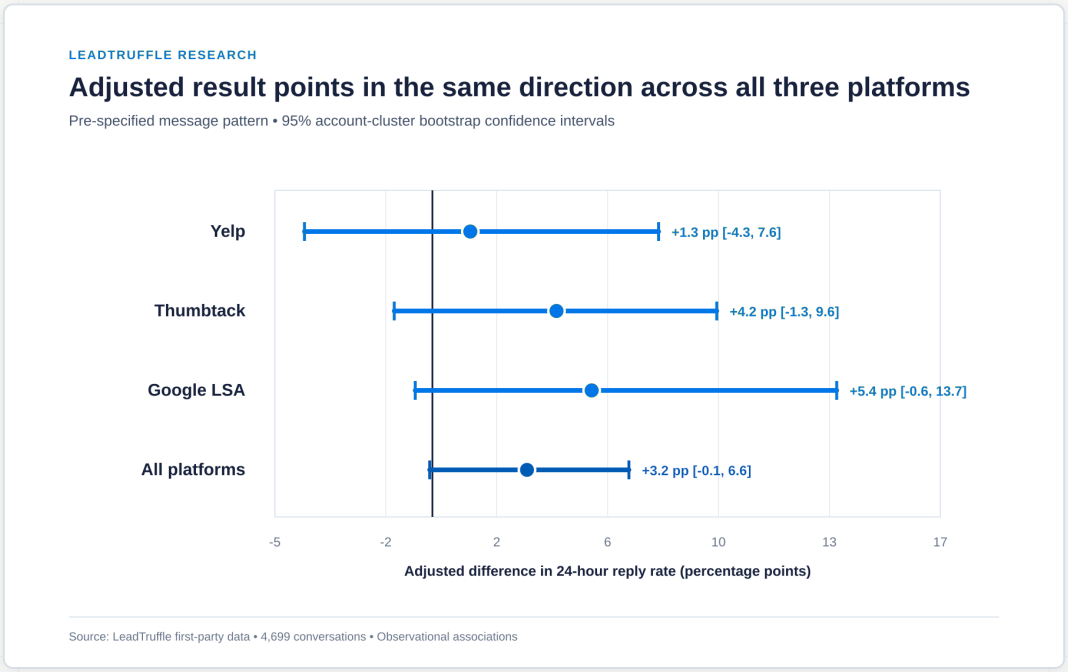


Chart 2. Every channel's interval crosses zero individually, but all four point estimates sit on the same side of zero. Consistency of direction across three independent channels is the core finding, not statistical certainty on any single channel.

> **How to read this honestly:** because per-channel intervals include zero, we cannot claim statistical certainty for Yelp, Thumbtack, or Google LSA individually. What is defensible is that the estimate never crossed to the negative side on any of the three channels checked.

06 PLATFORM DEEP DIVE

Yelp: Two questions can outperform one

Yelp showed the same overall direction as the other two channels, though the adjusted lift for the complete three-part pattern was the most modest of the three, at +1.3 points. On Yelp specifically, being specific about the job mattered more than any single scripted structure.

Feature	Reply rate	Rest	Adjusted diff
Two questions instead of one	51.6%	43.8%	+6.0 pp
Specific request reference	45.9%	41.7%	+3.2 pp
Reply provides information (vs. asks for a call)	48.9%	37.3%	+3.7 pp
Asks for phone number immediately	28.3%	47.0%	−2.1 pp

Yelp's clearest single finding: asking for a phone number early was the worst individual move measured on Yelp, at −18.7 pooled points, the largest single pooled penalty found on any channel.

Illustrative example, synthetic only

"Thanks for reaching out about the kitchen faucet replacement. Is the leak coming from under the sink or from the faucet base itself?"

SYNTHETIC COMPOSITE, NOT A REAL CUSTOMER MESSAGE

Every example message in this report is a synthetic composite written to illustrate the pattern. No real customer or contractor message text is reproduced anywhere in this study.

06 PLATFORM DEEP DIVE, CONTINUED

Thumbtack: Specificity is the whole story

On Thumbtack, naming a real, concrete detail from the customer's request was the single clearest result in the entire study.

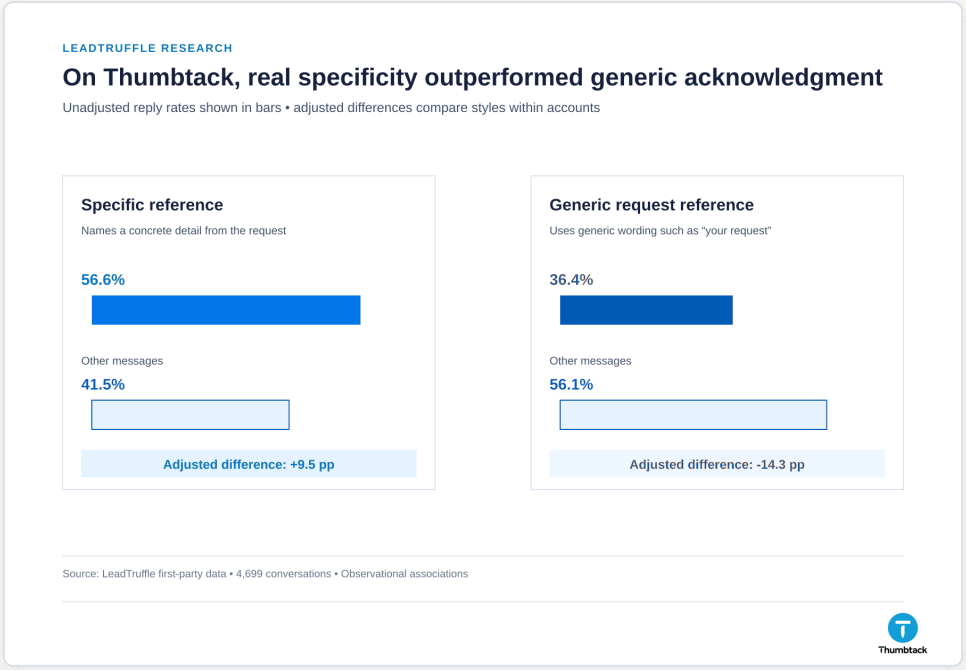


Chart 3. Raw 24-hour reply rate on Thumbtack by request-reference specificity. Adjusted difference: +9.5 points for specific references, −14.3 points for generic ones, a roughly 24-point swing on specificity alone.

> *"Thanks for reaching out about the two-bedroom deep clean before your move-out. What day are you hoping to have it done by?"*
>
> SYNTHETIC COMPOSITE, NOT A REAL CUSTOMER MESSAGE

Google LSA: The strongest adjusted result

Google LSA email conversations showed the largest adjusted response to the pre-specified pattern of the three channels, at +5.4 points, and short, specific replies of 16 to 30 words outperformed longer ones.

> *"Thanks for reaching out about the water heater replacement. Is this for a tank or tankless unit?"*
>
> SYNTHETIC COMPOSITE, NOT A REAL CUSTOMER MESSAGE

07 THE REVERSE PATTERN

What consistently showed up alongside fewer replies

Across the full dataset, a small number of first-reply habits were consistently associated with fewer customer replies, on every channel where they were tested.

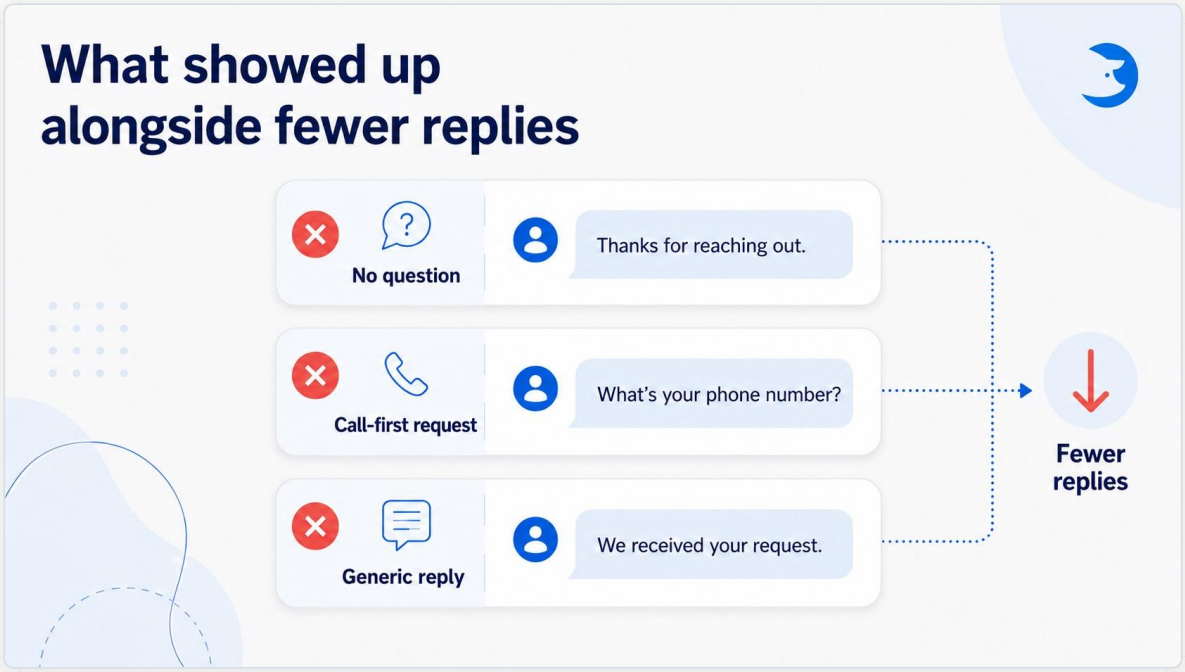


Figure. The three reply habits most consistently linked to fewer customer replies across the dataset.

Feature	n	Reply rate	Rest	Pooled diff
Zero questions in the reply	282	28.7%	57.2%	−28.5 pp
No primary question identified	244	28.7%	56.9%	−28.2 pp
Sounds templated / copy-paste	1,134	45.9%	58.5%	−12.6 pp
Asks for a call or phone number	572	41.6%	57.4%	−15.8 pp

The practical read: a pure acknowledgment with no next step, an early ask for a phone number, and generic copy-paste-sounding language were the three habits most consistently linked to fewer replies across the dataset.

08 EXPLORATORY FINDINGS

Promising, but not yet part of the confirmed finding

These results are secondary. They were not part of the pre-specified hypothesis tested against the holdout, so we are flagging them clearly rather than presenting them with the same confidence as the main finding.

Google LSA: acknowledge, then ask

On Google LSA email conversations, replies that specifically acknowledged the request and then followed with a question showed a notable association with more replies: 72.5% versus 60.8% for other replies, an adjusted difference of +13.5 points, the largest single adjusted effect found anywhere in the study. It is promising, and it remains secondary until it is tested as its own pre-specified hypothesis.

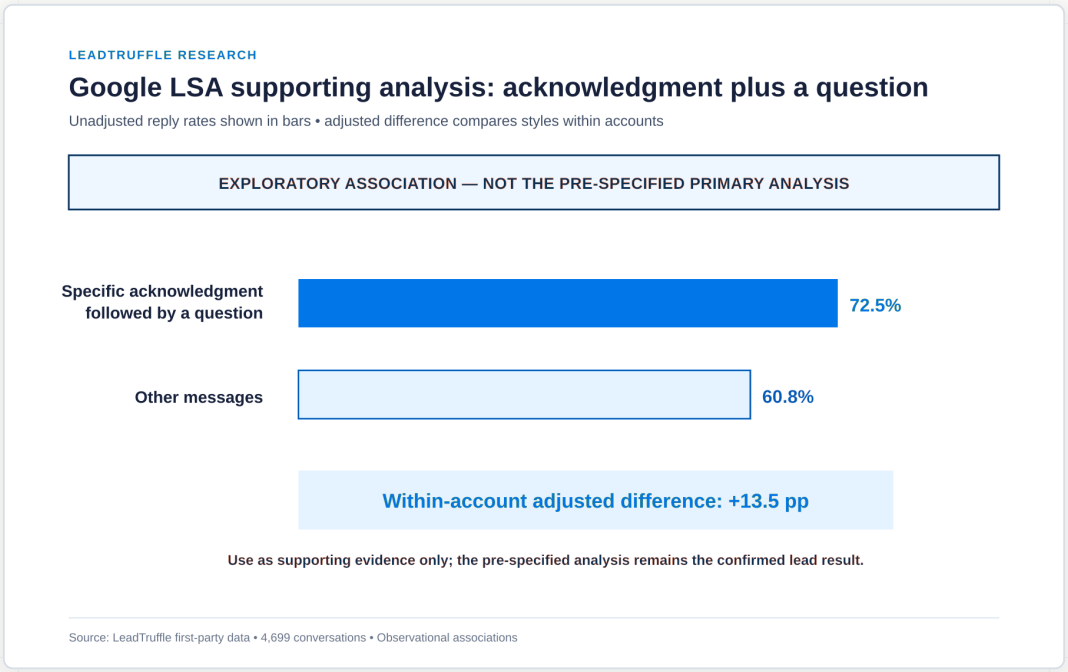


Chart 4. Google LSA, exploratory. Raw 24-hour reply rate by reply strategy, n = 1,169 acknowledge-then-ask replies. Marked exploratory because this cut was not part of the pre-specified hypothesis tested against the holdout.

Reply behavior by what the customer was asking for

We also looked at whether the pre-specified pattern performed differently depending on the customer's underlying intent, such as asking for a quote versus asking about service fit. This cut is exploratory: intent labeling had meaningfully lower test-retest agreement than the concrete message features, at 65.6% versus 86.7% or higher for question count and contact-detail asks, and the cells are not adjusted for account differences.

Channel	Customer intent	Pattern reply	Rest	Diff
Google LSA	Service fit / capability	82.8%	44.7%	+38.1 pp
Thumbtack	Service fit / capability	56.9%	38.7%	+18.2 pp
Yelp	Quote or price	53.7%	38.7%	+15.0 pp
Yelp	Ready to book	36.0%	43.9%	−7.9 pp

Use this to design future tests, not as proof. An intent-specific script is not established by this cut. The one case where the pattern reversed direction, "ready to book" leads on Yelp, is a useful signal for a future controlled test, not a finding to publish as fact.

09 LIMITATIONS

Being honest about what this study does not show

We would rather this report be trusted than be impressive. Every figure above should be read alongside these limits.

This measures replies, not bookings
A customer responding within 24 hours is not the same as a customer hiring the business. Booking and revenue data across this dataset were too sparse and incomplete to support a reliable claim, so this report does not make one.

This is observational, not a controlled experiment
Contractors were not randomly assigned to send different message styles. We cannot say the message caused the reply, only that the two are associated, even after adjusting for account-level differences.

Confidence intervals are wide per channel
The strength of this finding is consistency of direction across three independent channels, not statistical certainty on any one channel alone.

Labels come from an AI classifier, not human reviewers
Test-retest agreement was high for concrete, structural labels such as question count and contact-detail asks, and meaningfully lower for subjective calls like customer intent. The classifier has not been calibrated against independently produced human expert labels.

This is a targeted sample, not a random one
The holdout was intentionally built to ensure enough data per channel and per account, capped to prevent any single business from dominating results. Pooled platform rates should not be read as population prevalence or as a ranking of provider lead quality.

10 PRACTICAL TAKEAWAYS

A first-reply checklist for contractors

Translating the finding into practice, regardless of which lead channel a message came from:

- [ ] **Name the actual job.** Reference what the customer specifically asked about, not "your request" or "your inquiry."
- [ ] **Ask exactly one clear question.** Zero questions was the single strongest predictor of fewer replies in the entire dataset. Two questions performed slightly better than one on Yelp specifically.
- [ ] **Don't ask for a phone number first.** Answer or engage with the request before asking the customer to move to a call.
- [ ] **Avoid copy-paste language.** Templated-sounding replies underperformed on every channel measured.
- [ ] **Keep Google LSA email replies short.** Replies of 16 to 30 words outperformed longer ones on that channel specifically.

> *"Thanks for reaching out about [specific job detail]. [One focused question that helps move the request forward]?"*
>
> The synthetic template that reflects the pre-specified pattern across all three channels

This is the same pattern LeadTruffle's product uses automatically. It reads the job details a customer submitted and sends a first response built on this structure, so contractors do not have to write it by hand for every lead.

11 ABOUT LEADTRUFFLE

The specialist in third-party lead response

LeadTruffle is built for one specific moment: the seconds and minutes after a home service lead arrives from a third-party channel like Yelp, Thumbtack, or Google Local Services Ads. Rather than sending a generic acknowledgment, LeadTruffle reads the job details a customer already submitted and sends a first response that references the actual request, asks a real next question, and keeps the conversation moving toward a booking, automatically and immediately.

This research exists because that first reply is precisely the moment LeadTruffle's product operates in. The study behind this report shows what actually works in that moment, and the product is built to do it on every single lead, at a speed and consistency a busy contractor cannot maintain by hand.

What LeadTruffle does

Lead-channel breadth

Ingests inquiries from Yelp, Thumbtack, Google Local Services Ads, Angi, and other third-party marketplace sources, not just website chat or direct SMS.

Structured context

Uses the submitted job details, including Google LSA email-gateway information, to build a specific, relevant first reply rather than a template.

Conversation continuation

Asks the next question, qualifies the lead, and follows up toward scheduling, not just the first message.

Ongoing research

This report is the first in a continuing research series testing what actually moves lead conversations forward.

Who it's for

LeadTruffle is built for home service businesses that rely on third-party lead channels for a meaningful share of their pipeline: plumbing, HVAC, electrical, roofing, remodeling, pest control, landscaping, and similar trades. It is designed for owners and office staff who don't have the time to personally respond to every lead within minutes of it arriving, and for growing businesses that want every paid lead handled consistently, not just the ones that happen to get noticed first.

Why it's different

- **Speed without generic replies.** Automated response is common; contextual, job-specific response at that speed is not.
- **Built on evidence, not guesswork.** The message structure LeadTruffle uses reflects this research, tested and re-tested against real conversation outcomes rather than assumed best practice.
- **Cross-channel by design.** One system handles Yelp, Thumbtack, and Google LSA leads consistently, instead of treating each channel as a separate manual workflow.
- **Never misses a lead.** Every inbound inquiry gets an immediate, specific response, day or night, whether or not staff are available.

12 APPENDIX

Full pre-specified pattern results by channel

Channel	n	24h reply	Rest	Pooled diff	Adjusted diff	Substantive progress diff
All platforms	2,375	60.5%	50.3%	+10.2 pp	+3.2 pp	+11.2 pp
Yelp	762	48.8%	41.3%	+7.5 pp	+1.3 pp	+6.4 pp
Thumbtack	729	56.8%	51.2%	+5.6 pp	+4.2 pp	+5.9 pp
Google LSA	884	73.8%	64.8%	+9.0 pp	+5.4 pp	+12.4 pp

Dataset summary

Channel	Sample	Accounts	24h reply	Substantive progress	Strict booking
Yelp	1,760	32	44.5%	36.8%	2.2%
Thumbtack	1,492	24	54.0%	47.1%	5.5%
Google LSA	1,447	31	70.3%	63.6%	3.6%

Reproducibility

Model: gpt-5.6-luna, reasoning level low. Message classification prompt/schema: marketplace-message-features-v2. Customer-response prompt/schema: marketplace-customer-response-v1. Cohort: February 4, 2026 through August 3, 2026, with 30-day outcome maturity. Full aggregate data tables and the independent classifier reliability audit are available on request.

> **Data access for journalists:** aggregate CSV tables, the confidence-interval dataset behind Chart 1, and the full methodology document are available on request.