From Manual Lead Management to AI-Assisted Sales Operations: A Case Study of a B2B SaaS Company

Uluç Muslu

November 2026

Online version: ulucmuslu.com/research/ai-lead-qualification-b2b-saas

Abstract

Small business-to-business (B2B) software firms are starting to use large language models (LLMs) to qualify and prioritize sales leads. Most published evidence on lead scoring, however, comes from firms with long data histories, and claims about AI agents in sales are rarely checked against operational records. This paper examines how AI-assisted lead qualification changed the sales operations of a Turkish B2B software-as-a-service (SaaS) company that sells to independent insurance agencies. Using a single-case insider design, it compares a manual process, in which one or two people made the first calls to more than 2,400 leads, with an AI-assisted process introduced in April 2026. The new process combined a conversational qualification agent on WhatsApp and Instagram, LLM-based scoring, a purpose-built customer relationship management (CRM) system, and automated feedback to the advertising platform. Sources include the CRM's code, an internal pipeline report, and acquisition records. The change reduced first-response time from the next working day to a few minutes and moved screening ahead of the first call. The pipeline records show, however, that the main causes of lost leads lay outside the first response: almost one in three leads with a negative outcome was not an insurance agency, and commercial fit, above all price, accounted for an even larger share of the recorded reasons. The paper argues that for small firms, the value of AI qualification depends less on the model than on explicit criteria, structured outcome data, deliberately retained human decision points, a link back to acquisition, and a measurement plan.

Keywords: lead qualification, lead scoring, large language models, sales operations, B2B SaaS, small and medium-sized enterprises

1. Introduction

In B2B markets, marketing generates leads, and sales turns some of them into customers. The handoff between the two has long been a weak point. Salespeople do not follow up every lead that marketing sends them, and whether they do depends partly on how well the leads were qualified (Sabnis et al., 2013). The value of a lead also appears to fall quickly when the first contact is delayed (Oldroyd et al., 2011). For a small firm, the underlying constraint is capacity. When one or two people must call every inquiry, response time and the quality of follow-up depend on how many leads arrived that day.

Large language models appear to remove this constraint. An LLM-based agent can reply to a prospect at any hour, ask qualifying questions, classify the answers, and rank the lead before a salesperson sees it. Vendors describe such systems in strong terms, but the academic evidence is narrower. Research on predictive lead scoring mostly relies on models trained on historical CRM data (D'Haen & Van den Poel, 2013; Nygård & Mezei, 2020), which small firms often do not have. One of the main field studies of generative AI at work concerns customer support rather than sales (Brynjolfsson et al., 2025). Less is known about what changes in daily operations when a small firm with a short data history introduces an LLM-based qualification system.

This paper examines one such case: a Turkish B2B SaaS company that sells sales and quoting software to independent insurance agencies. From March 2025 to June 2026, I was the company's sales and marketing manager. I built its paid acquisition and, in April 2026, the AI-assisted qualification system described here. The paper asks the following research question:

> *How does AI-assisted lead qualification influence operational efficiency and sales processes in a small B2B SaaS company?*

Two narrower questions follow from it. First, which parts of the lead-handling process changed, and which did not? Second, which of these effects can be measured with the records a small firm actually keeps?

The paper makes one deliberate choice that shapes the analysis. It does not attribute the company's revenue or customer numbers to the AI system. When the system went live, the database already held 2,415 lead records, close to the roughly 2,500 leads in the company's acquisition records, so nearly all of the leads and customers counted there predate the system. The period after go-live was also too short for a fair comparison of conversion rates. The analysis therefore focuses on process-level effects that can be observed directly, and it treats the limits of measurement as a finding in its own right.

2. Literature Review

2.1 Lead qualification and the marketing-sales handoff

Lead qualification is the step that decides which inquiries deserve a salesperson's time. Jolson (1988) distinguished between "tight" and "loose" approaches to qualifying leads. The choice between them is a trade-off. Tight qualification protects sales time but risks discarding real buyers, while loose qualification passes more leads to sales and moves the cost of filtering to them. It is, in effect, a decision about where scarce capacity is spent.

The handoff between the two functions is also a known source of friction. Kotler et al. (2006) traced the tension between sales and marketing to both economic and cultural causes. Sabnis et al. (2013) examined what this means for individual leads. Using data from 461

salespeople in four firms, they showed that the follow-up of marketing leads depends on the salesperson's situation, and that experienced salespeople respond less to managerial tracking and more to how well the leads have been prequalified. Ohiomah et al. (2019) found a related effect for sales technology: the use of a lead management system improved inside sales performance indirectly, through more intensive follow-up and more adaptive selling. Both studies suggest that the quality of qualification, and the systems built around it, shape whether leads are worked at all.

2.2 Response time

The timing of the first contact is a second variable that firms can manage directly. Oldroyd et al. (2011) audited how 2,241 U.S. firms responded to a web inquiry. Thirty-seven percent responded within an hour, 24% took more than 24 hours, and 23% never responded. In a separate analysis of 1.25 million leads, firms that tried to contact a lead within an hour were nearly seven times as likely to qualify it, meaning to have a meaningful conversation with a decision-maker, as firms that waited even one hour longer. These figures come from a practitioner study co-authored by the head of a sales software company, not from peer-reviewed research, and they measure qualification rather than sales. Peer-reviewed evidence on lead response time remains scarce, which is one reason the effect of faster responses deserves careful measurement.

2.3 From scoring rules to predictive models

Lead scoring ranks leads by their likelihood to convert. Monat (2011) identified 16 commonly available lead characteristics that stand in for the determinants of industrial purchase decisions and turned them into a scoring model, while arguing that no single quantitative model fits every firm. D'Haen and Van den Poel (2013) proposed an iterative framework that is particularly relevant to small firms. When only data on existing customers are available, prospects are scored by their similarity to those customers; statistical models are added once outcomes for non-customers accumulate. Nygård and Mezei (2020) showed that supervised machine learning methods, such as random forests, can estimate purchase likelihood from historical behavioural data. A systematic review of 44 studies found decision trees and logistic regression to be the most common methods (Wu et al., 2024). A later study by the same authors notes that most machine learning scoring depends on behavioural data that exist only after the first contact, and it proposes profiling leads by firm characteristics before ranking them (Wu et al., 2026).

Most of these approaches need a history of leads with known outcomes. LLM-based scoring does not fit neatly into this lineage. An LLM can apply written criteria to unstructured answers without being trained on the firm's past outcomes. Agrawal et al. (2018) describe AI

as a technology that makes prediction cheap, while judgment, the decision about what each outcome is worth, remains with people. An LLM score built without outcome data is therefore closer to a rule-based score that encodes a manager's judgment than to a statistical prediction.

2.4 AI and human work in sales

Syam and Sharma (2018) examined how machine learning and AI affect each step of the selling process and set out an agenda for research on these changes. Paschen et al. (2020) mapped, for each stage of the B2B sales funnel, the tasks that AI can support and the role that people keep, an approach they call collaborative intelligence. Raisch and Krakowski (2021) argued that automation and augmentation depend on each other, and that emphasizing either one too strongly creates self-reinforcing problems.

Field evidence on generative AI supports a collaborative view. In a study of 5,172 customer support agents, access to an AI assistant increased the number of issues resolved per hour by 15% on average, with the largest gains among less experienced workers (Brynjolfsson et al., 2025). Adoption, however, depends on trust. People tend to abandon an algorithm after seeing it make mistakes, even when it performs better than a human (Dietvorst et al., 2015). In a field experiment with a predictive sales tool, Habel et al. (2024) found that its effect depended on the characteristics of both customers and salespeople. Their tool predicted customer churn rather than lead conversion, but the implementation problem is similar.

Research on small and medium-sized enterprises adds a further constraint. A review of 106 studies grouped the factors that shape AI adoption in such firms into eight clusters, including resources, knowledge, and infrastructure (Schwaeke et al., 2025). Small firms rarely have the data histories, analysts, or budgets that most lead scoring research assumes. This case adds a process-level view from that setting, in which the same person may design the system, run the advertising, and close the sales.

3. Research Design

3.1 Approach

The study uses a single-case design, which suits "how" questions about a contemporary process in its real setting (Yin, 2018). The case is revelatory in Yin's sense: it gives full access to the design and records of a kind of system that outside researchers rarely see in a small firm. Single-case designs have also been used to study marketing automation in B2B firms (Järvinen & Taiminen, 2016). Because I designed and operated the system, this is insider research. Coghlan (2019) describes the advantages of researching one's own organization, such as access and knowledge of how work is actually done, and its risks, such as role duality and a preunderstanding that can narrow what the researcher sees. I addressed these risks in four ways. All quantitative claims come from documents created at the time

rather than from memory. Several sources are used together, as Yin (2018) recommends. Recollection is used only for process timings and team practices that were not logged. Findings that are unfavourable to the system are reported alongside favourable ones.

3.2 Data sources

Table 1 lists the sources. Figures on response time and team capacity come from my direct observation as the operator of the system. They were not logged automatically and are reported as typical values.

Table 1

Data Sources

Source	Period	Content
CRM source code and database schema	April to June 2026	Pipeline stages, reason codes, integrations, feedback rules
Data migration log	April 2026	2,415 lead records moved from the previous database
Internal pipeline report generated from the CRM	24 March to 24 April 2026	232 leads with a negative outcome and their coded reasons; 17 open proposals
Aggregate acquisition records	2025 to 2026	Leads, product demonstrations, new customers, advertising spend
Direct observation as system operator	March 2025 to June 2026	Response times, team capacity, working practices

3.3 Analysis

The analysis compares two phases. The manual phase runs from March 2025 to early April 2026. The AI-assisted phase runs from the go-live of the new system in the second week of April 2026 to June 2026, about two months. The phases are compared on five operational dimensions: response time, the effort spent on screening, the structure of the records, the feedback from sales to acquisition, and the points at which people make decisions. In this paper, operational efficiency refers to how quickly, and with how much staff effort, leads are screened and routed. Staff hours and cost per qualified lead were not recorded separately, so efficiency is assessed through response time and the allocation of screening work. The reasons recorded for negative outcomes in the pipeline report are then analyzed descriptively.

The company and all individuals are anonymized. Only aggregated figures are reported, and no customer data appear in the paper.

4. Case Context

The company sells a web-based platform that helps independent insurance agencies manage their customers and renewals and obtain quotes from many insurers on one screen. The platform connects to more than 30 insurance companies through an integration layer developed by its parent company. The target customer is an agency with roughly 3 to 15 employees, usually run by its owner, who also makes the purchase decision. The sales process consists of a first call, a product demonstration, a written proposal, and a contract. Customers could choose between a subscription and a one-time purchase.

Leads came mainly from lead forms on Facebook and Instagram, with search advertising on Google in a supporting role. Over about ten months of paid acquisition, the acquisition records I kept show roughly 2,500 leads, 300 product demonstrations, and 40 new customer accounts, for a total advertising spend of about USD 10,400. This equals about USD 4.20 per lead and about USD 260 of advertising per new customer. Twelve percent of leads reached a demonstration, and 13% of demonstrations became customers.

During the manual phase, one or two people made the first calls to new leads; the sales CRM later had five users in total. I ran the advertising and creative work, joined sales meetings, and closed a share of the accounts myself.

5. Findings

5.1 The manual process

In the manual phase, form submissions were collected in a shared database in Notion. One or two people called each new lead by phone. Typically, the first contact took place on the next working day. Leads that arrived in the evening or on the weekend waited longest, and on busy days the backlog grew.

Qualification happened during the first call. The caller had to establish whether the company was an insurance agency, how many insurers it worked with, how many people would use the system, and which software it already used. Every lead, relevant or not, required a call before it could be screened. The database did have structured fields for the pipeline stage, the reasons for a negative outcome, the insurers an agency worked with, and its web-service permissions. These fields were filled in by hand after each call. The database kept no structured history of stage changes, so the time a lead spent at each stage could not be measured, and records were not linked to the advertising platform's lead IDs. By April 2026 the database held 2,415 lead records.

No information flowed back from sales to advertising. The advertising platform was optimizing for form submissions, so it had no information about which of those people ran an agency or bought the product. Whether a lead was useful became known only on the phone, after the advertising cost had been paid.

5.2 The AI-assisted process

In April 2026, using an AI coding assistant, I built a CRM around the company's pipeline. The first working version, including the migration of all 2,415 records, was deployed within two days. The integration with the advertising platform followed within the next four days. The system had five components:

1. **Intake.** New form leads entered the CRM automatically through a workflow tool as they arrived, with a daily synchronization as a fallback. Duplicate submissions were rejected by their advertising lead ID.
2. **Conversational qualification.** An LLM-based agent sent the first message on WhatsApp or Instagram within a few minutes, at any hour. It asked how many insurers the agency worked with, how many people would use the system, which software the agency used, and whether it already had web-service access from its insurers. It also searched the company name on the web to check whether the lead was an insurance agency.
3. **Scoring.** An LLM scored each lead against written criteria, using the information collected in the conversation. The score set the order in which the sales team made calls.
4. **Structured pipeline.** Each lead was assigned one of 21 defined stages in four groups (to be called, evaluation, proposal and contract, and outcome), and a negative outcome was recorded with one or more of 12 coded reasons. The stages and reasons were carried over from the earlier database. What was new was a timestamped history of every stage change and a link between each new form lead and its advertising lead ID.
5. **Feedback to advertising.** Through Meta's Conversions API for CRM integration (Meta, n.d.), the CRM sent a "Lead" event when a salesperson moved a lead into an evaluation or proposal stage, and a "Purchase" event carrying the monthly contract value when a contract was signed. A custom "LeadDisqualified" event was sent only when a lead was rejected because it was not an insurance agency, and that contact was also added to an audience excluded from future advertising. Other negative reasons, such as price or no response, were treated as "soft" negatives and not reported, since a real agency that declines on price is still the right audience. The feedback was meant to let the advertising platform optimize for lead quality rather than for form volume.

People kept every decision that involved a commitment. The agent did not quote prices or make promises. Calls, demonstrations, proposals, and pricing remained with the sales team. Table 2 summarizes the differences between the two phases.

Table 2

Lead Handling Before and After the Change

Dimension	Manual phase (March 2025 to April 2026)	AI-assisted phase (April to June 2026)
First response	Phone call, typically on the next working day	WhatsApp or Instagram message within a few minutes, at any hour
Screening	During the first call, by a salesperson	Before the first call, by the agent and the scoring step
Call order	Order of arrival	Score
Record structure	Stages and reason codes filled in by hand; no stage history; no link to advertising lead IDs	Same stages and reason codes, plus a timestamped stage history and advertising lead IDs for form leads
Feedback to advertising	None	Qualified, non-agency, and purchase events sent automatically
Human decisions	All steps	Calls, demonstrations, proposals, and pricing

5.3 What the pipeline records show

The internal report for 24 March to 24 April 2026 is the most detailed record from the period of the change. It lists 232 leads with a negative outcome and 17 leads waiting for a response to a proposal. Of the 232 leads with a negative outcome, 211 had at least one recorded reason, and some had more than one. Together with the 21 leads that had none, Table 3 counts 253 entries.

Table 3

Recorded Reasons for Negative Outcomes, 24 March to 24 April 2026 (n = 232 leads, 253 entries)

Group	Reasons included	Count	Share of entries
Not eligible	Not an insurance agency	76	30.0%
Commercial fit	Price too high (51); works with too few insurers (24); product does not meet the need (15); insufficient capital (7)	97	38.3%
Could not be reached	No response (28); number not in use (5)	33	13.0%
Not interested	Not interested (17); personal reasons (8)	25	9.9%
No reason recorded	None given	21	8.3%
Bought elsewhere	Bought from another provider	1	0.4%

Note. Percentages do not sum to 100 because of rounding.

The largest single reason was that the lead was not an insurance agency. These leads included banks, courier companies, accountants, driving schools, construction firms, and universities, all of which had filled in a form aimed at agencies. Almost one in three leads with a negative outcome (76 of 232) therefore pointed to a targeting problem in advertising rather than a problem in sales.

Commercial fit was the largest group of reasons. Price alone appeared in 22% of the leads with a negative outcome, and small agencies that worked with few insurers formed a further segment for which the product was too large. Faster qualification can identify such leads earlier, but it does not address price.

Finally, 21 of the 232 leads (9.1%) had no reason recorded. A structured field therefore did not guarantee complete data.

The 17 open proposals show where work was waiting at the time of the change. Eleven of them came from leads that arrived before the new system went live. Four leads had stopped answering after receiving a proposal. In three cases the decision-maker was travelling or had left the agency (one lead appears in both groups), three leads were weighing the cost or postponing, and six had received their proposal in the previous few days. None of these delays concerned the first response.

5.4 Effects on operations

The following effects could be observed directly, without additional measurement.

Response time. The first response moved from the next working day to a few minutes, and leads that arrived outside office hours were handled in the same way as leads that arrived during them. The comparison has a limit. Oldroyd et al. (2011) measured attempts to reach a decision-maker, while the agent's first message was not yet a conversation with one. Even so, this is the change most likely to matter for conversion, although this study cannot measure that effect.

Screening. Screening moved from the first call to the agent and the scoring step. Calls began with known facts about the agency, and the order of calls followed the score rather than the order of arrival. In effect, the agent collected a simple version of the firm-level profile that Wu et al. (2026) recommend building before leads are ranked. For the one or two people making first calls, this changed the content of the calls; whether it saved time was not measured.

Disqualification as a signal. A negative outcome had previously been a field in a database that only the sales team read. For leads that were not insurance agencies, it now also became an event sent back to the advertising platform, together with an exclusion from future audiences. A lead rejected for that reason became a data point that the advertising system could also use.

A likely shift of the bottleneck. The theory of constraints holds that improving one step of a process moves the limit to the next weakest step (Goldratt & Cox, 2004). Once the first response was fast, proposals, pricing, and decision-makers would be expected to become the limit. The open proposals in the April report are consistent with this expectation, but most of them came from leads first contacted by phone, so they cannot confirm it.

Some things lay outside the reach of the system. It did not address price objections or the mismatch with small agencies, which sit in the product and its pricing, and the observation window was too short to show whether the feedback to advertising had reduced the share of non-agencies arriving through the forms. Several effects could also not be measured. The window after go-live was about two months, while the path from lead to signed contract often took several weeks. Response times in the manual phase were not timestamped. The LLM score was not validated against closed outcomes, because too few outcomes accumulated within the window. The stage history recorded by the new CRM was not extracted before the observation period ended, so stage durations are not reported, and no figures on the agent's own performance, such as its reply rate or the number of non-agencies its web check identified, are available for this study.

6. Discussion

6.1 Speed is the easiest gain but not a sufficient one

The clearest effect of the system was speed, which fits the evidence on response time. Speed, however, acts only on leads that are worth reaching. In this case, most negative outcomes had causes that sat before or after the first response: the wrong audience reached the form, or the product and price did not suit the agency. Faster qualification would bring these facts to light sooner, although this study did not measure how much sooner.

6.2 An LLM score encodes judgment

The scoring step is better described as encoded judgment than as prediction. The criteria were my own, written down and applied by the model to every lead, although the consistency of its scores was not tested. This has real value. The criteria became explicit, every lead was assessed against the same written standard, and the assessment happened at any hour. It also carries a risk of false precision, because a number attached to a lead looks more objective than the judgment behind it. The research on algorithm aversion suggests a second risk: if salespeople see the score fail on a few visible cases, they may stop using it altogether (Dietvorst et al., 2015). The framework of D'Haen and Van den Poel (2013) indicates the way forward for a small firm. A score can begin as judgment and become a prediction only as outcomes accumulate and the score is checked against them. That step requires the structured, timestamped records described in Section 5.2.

6.3 The feedback loop

The largest single reason for negative outcomes pointed to targeting rather than sales. Qualification at the end of the funnel identifies such leads only after the firm has paid for them. Sending quality signals back to the advertising platform, and excluding non-agencies from future audiences, moves the correction to the start of the funnel. AI lead qualification therefore depends as much on passing outcome data back to marketing as on the conversational agent itself. Coded reasons also help with the friction between sales and marketing described by Kotler et al. (2006), because both functions can then discuss lead quality in the same terms.

The loop has limits that matter for small firms. Meta's documentation recommends this kind of integration for advertisers that generate at least 200 leads a month, and it expects the stage being optimized to be reached within 28 days of the lead (Meta, n.d.). The acquisition records show about 2,500 leads over ten months, or about 250 a month, so the company was above the volume threshold, but not by a wide margin. In a sales cycle of several weeks, a signed contract may also arrive too late to serve as the optimization signal, which leaves the

earlier "Lead" stage as the practical one. A smaller firm might automate qualification in the same way and still lack the volume, or the speed of outcomes, that the advertising platform needs to learn from the feedback.

6.4 Augmentation in a small firm

The system gave a small team more coverage, especially outside office hours. People kept pricing, proposals, and every promise made to a customer. This is closer to the collaborative model of Paschen et al. (2020) than to full automation. Keeping people in the conversations in which agencies explained why they bought or declined also preserved the learning that Raisch and Krakowski (2021) warn can be lost when a task is fully automated.

One question this case did not examine is disclosure. In a field experiment, Luo et al. (2019) found that customers bought less when they were told at the start of a conversation that they were talking to a chatbot. For an agent that opens the conversation on WhatsApp, how and when it identifies itself is both an ethical and a commercial design decision, and one that deserves its own study.

6.5 Implications for small firms

The case suggests five practical considerations for small firms that plan to introduce AI into lead qualification:

1. Write the qualification criteria down before automating them. The agent can only apply criteria that someone has made explicit.
2. Build structured outcome records first. Stages and coded reasons are what make later evaluation possible. In this case they already existed in the earlier database, which made a fast migration possible.
3. Keep people at the points of commitment, such as pricing, proposals, and promises, and make this clear to the team.
4. Connect qualification back to acquisition, so that the cause of poor leads can be corrected where it arises, and check whether lead volume is high enough for the advertising platform to use the feedback.
5. Plan the measurement before launch. Timestamps on every first contact, a period in which a share of the leads is handled without the agent, and a fixed date for checking scores against outcomes would make the effect of the system measurable.

7. Limitations

This is a single case in one industry and one country, and its findings cannot be generalized statistically. Eisenhardt (1989) recommends between four and ten cases for building theory, and a single case can only suggest patterns for later testing. As an insider study, this paper is exposed to the bias that comes from evaluating one's own design. The

quantitative sources were created during operations rather than for research. The period after go-live was short, the pipeline report covers mostly the weeks before and during the transition, and response times in the manual phase were not logged. Other changes in the same period, such as new advertising creative, pricing discussions, and seasonal demand, may also have affected the pipeline. There was no control group. For these reasons, the paper describes changes in process and does not estimate effects on conversion or revenue.

8. Conclusion

In this case, AI-assisted lead qualification changed the front of the sales process. First responses came within minutes, screening moved from the first call to the system, and leads rejected as non-agencies became signals that flowed back to advertising. The pipeline records show that the main causes of lost leads, poor targeting and commercial fit, lay outside the first response, and the period of observation was too short to measure the system's effect on sales. The answer to the research question is therefore that, in the sense of efficiency used here, the system made first contact faster and moved screening ahead of the first call, which changed where the sales team spent its effort. Its effect on staff hours and on sales outcomes remained unmeasured.

For small firms, the main lesson is that the value of AI qualification depends less on the model than on the operational design around it: explicit criteria, structured outcome data, deliberately retained human decision points, a link back to acquisition, and a measurement plan. Future research could test such systems with controlled rollouts, in which a random share of leads receives the agent's first response, validate LLM scores against closed outcomes over a full sales cycle, and compare several small firms in different industries.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.

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

Coghlan, D. (2019). *Doing action research in your own organization* (5th ed.). SAGE.

D'Haen, J., & Van den Poel, D. (2013). Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. *Industrial Marketing Management, 42*(4), 544–551. https://doi.org/10.1016/j.indmarman.2013.03.006

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

Goldratt, E. M., & Cox, J. (2004). *The goal: A process of ongoing improvement* (3rd rev. ed.). North River Press.

Habel, J., Alavi, S., & Heinitz, N. (2024). Effective implementation of predictive sales analytics. *Journal of Marketing Research, 61*(4), 718–741. https://doi.org/10.1177/00222437221151039

Järvinen, J., & Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. *Industrial Marketing Management, 54*, 164–175. https://doi.org/10.1016/j.indmarman.2015.07.002

Jolson, M. A. (1988). Qualifying sales leads: The tight and loose approaches. *Industrial Marketing Management, 17*(3), 189–196. https://doi.org/10.1016/0019-8501(88)90002-8

Kotler, P., Rackham, N., & Krishnaswamy, S. (2006). Ending the war between sales and marketing. *Harvard Business Review, 84*(7/8), 68–78.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science, 38*(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

Meta. (n.d.). *Conversions API for CRM integration*. Meta for Developers. Retrieved November 2, 2026, from https://developers.facebook.com/documentation/ads-commerce/conversions-api/conversion-leads-integration

Monat, J. P. (2011). Industrial sales lead conversion modeling. *Marketing Intelligence & Planning, 29*(2), 178–194. https://doi.org/10.1108/02634501111117610

Nygård, R., & Mezei, J. (2020). Automating lead scoring with machine learning: An experimental study. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 1439–1448). https://hdl.handle.net/10125/63916

Ohiomah, A., Andreev, P., Benyoucef, M., & Hood, D. (2019). The role of lead management systems in inside sales performance. *Journal of Business Research, 102*, 163–177. https://doi.org/10.1016/j.jbusres.2019.05.018

Oldroyd, J. B., McElheran, K., & Elkington, D. (2011). The short life of online sales leads. *Harvard Business Review, 89*(3), 28.

Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons, 63*(3), 403–414. https://doi.org/10.1016/j.bushor.2020.01.003

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Sabnis, G., Chatterjee, S. C., Grewal, R., & Lilien, G. L. (2013). The sales lead black hole: On sales reps' follow-up of marketing leads. *Journal of Marketing, 77*(1), 52–67. https://doi.org/10.1509/jm.10.0047

Schwaeke, J., Peters, A., Kanbach, D. K., Kraus, S., & Jones, P. K. (2025). The new normal: The status quo of AI adoption in SMEs. *Journal of Small Business Management, 63*(3), 1297–1331. https://doi.org/10.1080/00472778.2024.2379999

Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management, 69*, 135–146. https://doi.org/10.1016/j.indmarman.2017.12.019

Wu, M., Andreev, P., & Benyoucef, M. (2024). The state of lead scoring models and their impact on sales performance. *Information Technology and Management, 25*(1), 69–98. https://doi.org/10.1007/s10799-023-00388-w

Wu, M., Andreev, P., & Benyoucef, M. (2026). Profiling before scoring: A two-stage predictive model for B2B lead prioritization. *Journal of Personal Selling & Sales Management, 46*(3), 313–329. https://doi.org/10.1080/08853134.2026.2633231

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.