

White Paper

Improve Data Quality to Support Quality AI Outcomes

Sponsored by: Epsilon

Lynne Schneider

July 2025

IDC OPINION

The evolving landscape of artificial intelligence is shifting from broad, generic discussions to a more nuanced focus on foundational data quality as a key differentiator. As organizations increasingly integrate AI into their operations, data quality becomes paramount in driving meaningful insights and achieving competitive advantage. According to IDC's research, the emphasis on data quality is not merely a technical requirement but a strategic imperative that influences the effectiveness of AI applications across various sectors. The success of these initiatives hinges on the ability to manage and leverage high-quality data, resolve identity, and deliver AI-ready data products. It is imperative to have a strategy that refines the data your enterprise gathers into a clean and consolidated foundation for enrichment, analysis, and action.

IN THIS WHITE PAPER

We cannot be satisfied with the status quo on data quality when the process and technology continue to advance. Every enterprise can point to a situation where AI and other insight efforts could have been more successful if only their customer and prospect records were better consolidated and streamlined, if the identities had been better aligned, and if there was better insight to plan campaigns with better outcomes.

Key points discussed in this white paper include:

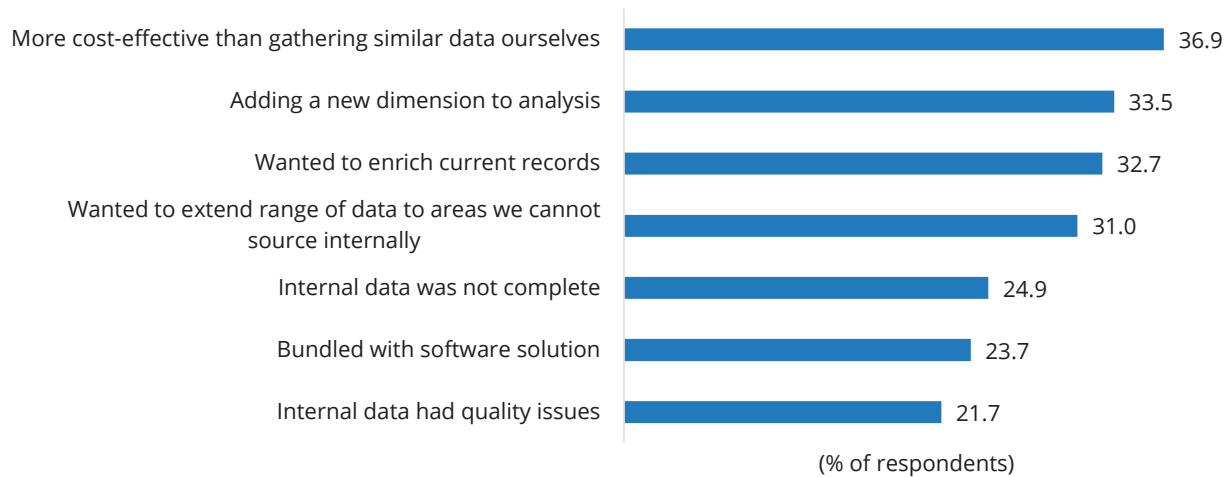
- Robust data, including enrichment, feeds quality foundational data. Leveraging external data, such as third-party data and partner data, includes many benefits (see Figure 1).

- Broad foundational data leads to higher-quality identity resolution, supporting more trusted and valuable insights.
- Higher-quality identity resolution fuels a modern AI-enabled marketing function.
- AI-enabled marketing execution ultimately drives better business outcomes.
- Data is fluid and constantly changing — and your data strategy needs to take an agile approach.

FIGURE 1

External Data Additive to Internal Data

Q. *Why did you or would you begin using external data sets?*



n = 511

Note: Data is weighted using "country IT spend."

Source: IDC's *External Data Sourcing and Collaboration Survey*, June 2024

SITUATION OVERVIEW

AI is top of mind for marketers everywhere. It is supposed to be the key to more productive, better targeted, and more personalized content — and it is also supposed to make all the members of the marketing and business operations teams more efficient. However, deploying predictive and generative AI (GenAI) without creating the foundation of high-quality, identity-resolved data can only provide efficiency without a higher level of effectiveness. To change the outcomes, you need to increase your data quality.

No one plans to have low-quality data, but not enough organizations plan and carry out the activities needed to raise data quality. How can we go about increasing data quality, knowing that we plan to use this data to fuel our newer AI-infused way of operating?

First, we need to acknowledge that first-party data alone cannot provide a robust enough view of customers and prospects. Even with dedicated efforts to encourage current customers and prospects to share information, this data is likely very static. It can be okay when it comes to demographic data or information about past behavior, but it is unlikely to give a dynamic, recent view of what the audience is actually doing here and now. Many marketers want to add intent data, data from partners, and other types of attributes to get more information for effective segmenting and targeting.

This is where many run into a secondary issue — how to reconcile disparate data sources to unified identities. The goal is to take customer data and combine it with the other sources to create a unified, holistic view. This is absolutely necessary before predictive and generative AI enters the picture. AI doesn't think; its strength is how it predicts and uncovers insights in the data presented to it. If it is not presented with a high-fidelity view of your customers, it diminishes the accuracy of understanding how, when, where, and with what messages to reach people. Having a lot of unmatched data doesn't provide an advantage; it can even work against you.

In marketing, the success of AI-driven initiatives is directly tied to the quality of the data that powers them. When marketers rely on clean, current, and well-organized data, predictive and generative AI tools can effectively segment audiences, personalize messaging, and optimize campaign performance. On the other hand, if the underlying data is inaccurate, incomplete, or outdated, AI models may misidentify customer preferences, target the wrong segments, or deliver irrelevant content — ultimately leading to wasted resources and missed opportunities.

The consequences of poor data quality, and therefore poor identity resolution, in marketing are far reaching. Campaigns may fail to reach their intended audiences, personalized offers may fall flat, and customer trust can erode if communications feel off-target or repetitive. These challenges highlight the necessity for strong data foundations. Marketers must not only collect and store data responsibly but also use technology to regularly clean, update, and enrich their databases to ensure ongoing accuracy.

Building a High-Quality Data Foundation

A robust data foundation begins with a thorough review and organization of your first-party data. This means collecting all customer information from different touch points — such as email lists, purchase histories, and website interactions — and bringing it together into a single, cohesive system. Once consolidated, it's important to clean the data by removing duplicates, correcting typos, and ensuring consistency in how information is recorded. This process not only streamlines your database but also helps you build a clearer, more accurate picture of your customers.

Maintaining internal data hygiene is an ongoing responsibility. Without regular upkeep, errors and outdated details can quickly accumulate, undermining your marketing efforts and leading to wasted resources. To combat this, organizations should establish clear guidelines for how data is entered and maintained and use automated tools to flag potential issues. Periodic reviews — such as spot checks or scheduled audits — can catch problems early and keep your data in top shape.

Identifying and resolving data issues requires both technology and teamwork. Automated solutions can help spot inconsistencies or missing fields, while team members can address more complex discrepancies that require human judgment. Encouraging everyone who handles data to follow best practices and communicate about potential problems creates a culture of accountability. With these steps in place, your data will remain reliable and ready to support advanced marketing strategies and AI-driven insights.

The Strategic Value of Holistic Identity Resolution and Third-Party Data

Many organizations quickly discover the limitations of relying solely on first-party data. Customer information collected directly from owned channels — such as website registrations, purchase records, or email signups — often suffers from gaps, outdated details, or inconsistencies. People change addresses, update their contact information, or shift their preferences, and unless customers proactively share these changes, your records can become stale. This makes it difficult to maintain a clear, accurate view of each individual, which in turn hampers effective personalization and targeting.

To overcome these challenges, it's essential to understand what constitutes high-quality data. Data quality isn't just about having a lot of information; it's about ensuring that each piece is accurate, complete, timely, consistent, and relevant to

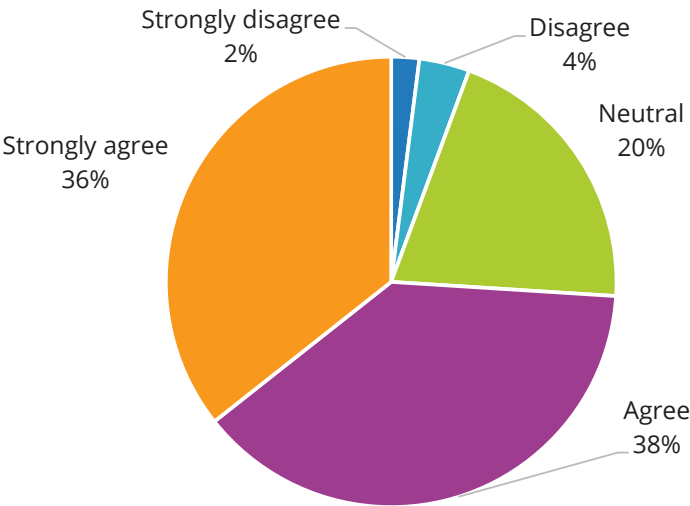
your goals. When any of these elements are missing, your marketing efforts risk missing the mark. This is where identity resolution and the use of third-party data become strategic assets. By integrating external data — such as demographic updates, lifestyle changes, or purchase behaviors — you can bridge the gaps in your own records, creating a more holistic and up-to-date profile of your customers. This not only sharpens your targeting but also sets you apart from competitors who may be working with less complete information (see Figure 2).

Enhancing your first-party data with second- and third-party sources is no longer a luxury; it's a necessity in today's fast-moving market. External data can provide valuable enrichment, such as the latest address information, insights into changing consumer preferences, or behavioral trends that signal shifting interests. For example, supplementing your database with third-party updates might reveal that a customer has recently moved, started a new job, or shown interest in a new product category — details that can dramatically improve the relevance of your outreach. By making data augmentation a core part of your strategy, you ensure your marketing efforts remain accurate, timely, and impactful, positioning your business for sustained growth and differentiation (see Figure 3).

FIGURE 2

Generative AI Initiatives Leading to More Data Sourcing

Our generative AI efforts have led us to bring in more external data



n = 511

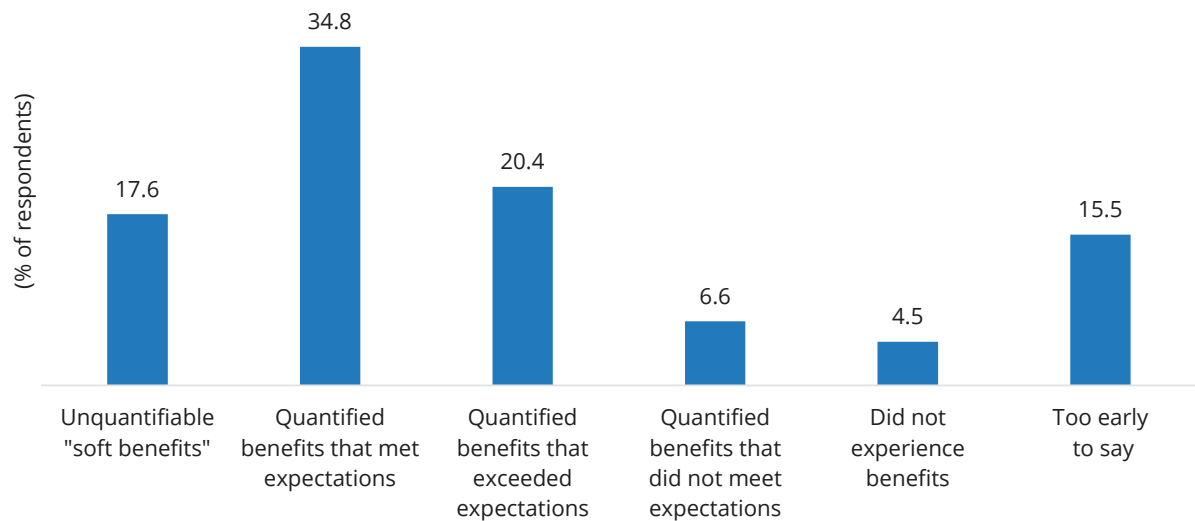
Note: Data is weighted using "country IT spend."

Source: IDC's *External Data Sourcing and Collaboration Survey*, June 2024

FIGURE 3

Most Enterprises See Benefits from External Data

Q. *How has the marketing process benefited from the application of external data sets?*



n = 511

Note: Weighted using "country IT spend."

Source: IDC's *External Data Sourcing and Collaboration Survey*, June 2024

How Might This Work?

What if a popular women's retailer, with plenty of first-party data on their current customers, wanted to reach newer, younger customers during a pivotal shopping season? How could they use the power of AI to better understand this audience and reach them with the right message at the right time?

In this situation, there might be some first-party data the retailer could use. A team could analyze generative AI insights to uncover what audiences in this age group like to do, what product categories and products they purchased last year, what they've been buying recently, and whether they use iOS or Android devices. Using this input, they could develop a key audience profile. For this situation, the team would be using external data and creating identities to create a quality data set for analysis.

The next step would be to use the audience data and AI insights to develop a solution that would address both the needs of this audience and the products

and services of the retailer. In an iterative fashion, AI could also be used to measure and course correct on the activation.

Without quality data, the team wouldn't have insight into what would appeal to this new target audience. They would not be able to create a campaign with the same level of success and would not be able to measure as effectively.

FUTURE OUTLOOK

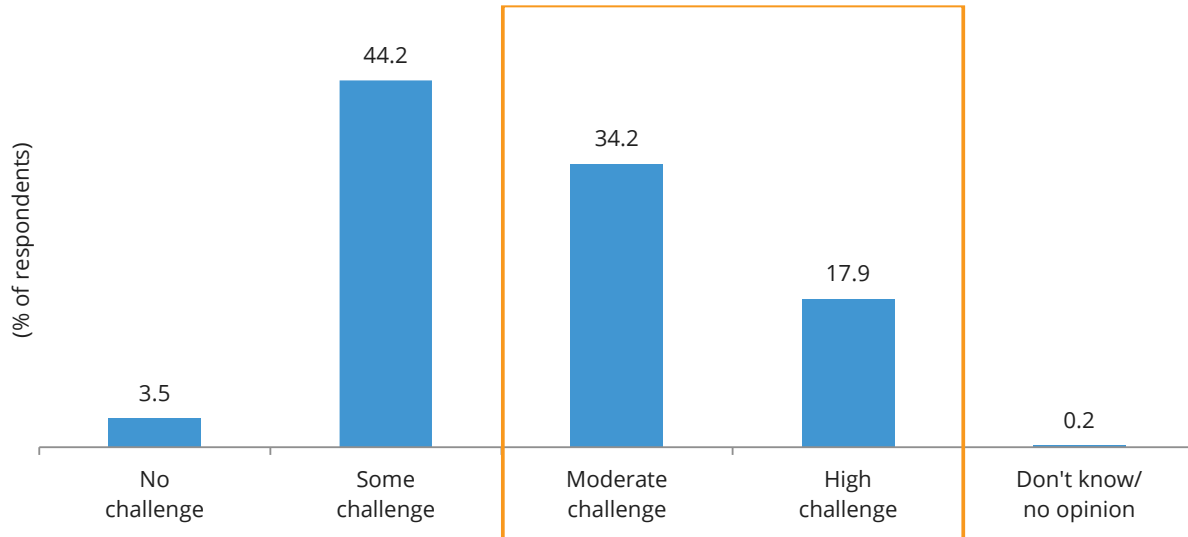
IDC research highlights that organizations across industries expect to significantly increase their use of predictive and generative AI in the coming years, but the value of these technologies depends entirely on the breadth and quality of data available to them. While AI technologies open new possibilities for business transformation — ranging from customer service and fraud detection to automated maintenance — these systems are only as effective as the data they can access and learn from.

Without high-quality, well-governed, and contextually relevant data, organizations risk amplifying errors and missing out on the promised ROI of AI investments. As AI adoption accelerates, IDC's research consistently points to the need for scalable data infrastructure, strong governance, and continuous data quality management to ensure that AI delivers meaningful, reliable business outcomes.

Trust will always be an issue with predictive and generative AI. Today, many face challenges getting others in their organizations to trust external data. This pattern can be even more intense when it comes to AI. Many see it as a black box that they don't understand, and as the capabilities grow, the possibilities for misuse and mistrust grow as well. Leading marketers are going to need to continuously educate their stakeholders about how high-quality data and reliable AI tools are being used to benefit them (see Figure 4).

FIGURE 4

Users of External Data Faces Challenges Getting Others in the Organization to Trust the Data



n = 511

Source: IDC's *External Data Sourcing and Collaboration Survey*, June 2024

MARKET CHALLENGES/OPPORTUNITIES

When considering a revamped approach to data quality and the steps from improving foundational data quality, to high fidelity identity resolution, assurance that your data is AI ready — all while remaining agile, IDC sees some common challenges and opportunities.

Overcoming Implementation Challenges

- Create a complete data quality improvement process; don't just address issues as they come up.
- Resource efforts appropriately; understanding the varied internal data sources and systems is key. Be realistic about the maturity of the organization's data infrastructure and data science expertise, and choose partners and technologies that can support your journey.
- Address common obstacles such as integration complexity, data privacy concerns, and organizational silos (challenges that arise with any data source).

- Prioritize ongoing data governance and continuously monitor data quality.
- Treat third-party data enhancement and identity resolution as essential components, not optional add-ons, to fully unlock the potential of AI-driven marketing strategies. Work with trustworthy partners.
- Plan the data and tech stack; will it be a comprehensive ecosystem, point solutions, or a hybrid? Choose the level of flexibility your organization requires.

Opportunities

- **Scalable data strategies:** Establishing robust data cleaning and enrichment processes lays the groundwork for identity resolution and scalable, repeatable AI initiatives that can adapt as data sources and business needs evolve.
- **Competitive differentiation:** By augmenting first-party data with high-quality third-party sources, organizations can access insights that competitors may lack, leading to smarter AI models and differentiated customer experiences.
- **Improved decision-making and ROI:** Reliable, comprehensive data creates a comprehensive understanding of each customer, which supports better AI-driven insights, leading to more effective marketing strategies, optimized spend, and improved return on investment.
- **Building trust and reducing waste:** High data quality reduces errors, increases trust in AI outputs, and minimizes wasted marketing spend caused by inaccurate targeting or duplicate communications.

VENDOR PROFILE: EPSILON

Epsilon is a global data, technology, and services company that powers the marketing and advertising ecosystem.

Key Considerations

- **Comprehensive identity resolution:** Epsilon's CORE ID allows for more accurate targeting, segmentation, and measurement, reducing duplication and inaccuracies often found in solutions relying on weaker identifiers such as email addresses or cookies.
- **Data clean room capabilities:** It supports both marketers and data scientists with a user-friendly interface for audience creation and prebuilt models, as well as advanced analytical tools for custom modeling.

- **Third-party data access:** It provides access to curated third-party data, compatible with CORE ID to help clients gain a deeper understanding of consumer attributes, including transactional, demographic, and psychographic data.
- **AI:** Epsilon's predictive and generative AI capabilities work across their solutions to help brands find and deliver the right message to more in-market consumers by optimizing to the individual in real time as consumer behavior changes.

CONCLUSION AND NEXT STEPS

As we've explored, high-quality data is the foundation upon which effective identity management and AI-driven marketing are built. When your data is accurate, current, and well organized, it strengthens your understanding of customer identities — enabling AI to deliver more relevant, timely, and personalized marketing experiences. In essence, data fuels your identity, identity powers your AI, and AI drives your marketing execution. This chain reaction ultimately leads to stronger engagement, higher conversion rates, and better business outcomes.

The future of data-driven AI in marketing and business is exciting — but only for organizations that treat data as a dynamic asset. Data is fluid: customer behaviors, preferences, and circumstances are constantly evolving. Relying solely on static, first-party data can leave your marketing strategies out of sync with real-world audience shifts. That's why leading companies are increasingly leveraging external partner and third-party data sources to supplement and refresh their customer insights. As highlighted in recent industry research, organizations are seeking data partners to provide richer, more up-to-date perspectives on consumer activity, intent, and trends.

Regularly enriching your first-party data with trusted external sources ensures that your AI models have the most current and comprehensive view of your audience. This approach not only sharpens your targeting and personalization but also helps future proof your marketing efforts in a rapidly changing landscape.

Next steps should include:

- Assess your current baseline, including the state of your data, data hygiene practices, and current identity and data partnerships

- Identify AI-enabled data strategy goals, incorporating both the data and the tools that will be needed to achieve those goals
- Connect with peers and others that can inform your marketing AI journey
- Choose trusted partners and technology to support your effort and ongoing operations
- Collaborate across your organization to improve data quality, manage/monitor process, and develop effective change management

In summary, the value of AI for marketing is only as strong as the data that feeds it. By prioritizing data quality, embracing a fluid data strategy, integrating external data sources, and forming identities, you empower your AI to deliver the insights and outcomes that drive sustained business success.

ABOUT IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2025 IDC. Reproduction without written permission is completely forbidden.