When AI Acquires Data: Strategic Complementarities in M&A*

Yue Fang† Vidhan K Goyal‡ Yingxuan He§ Zilong Zhang¶

August 9, 2026

Abstract

We examine whether complementarities between artificial intelligence (AI) capabilities and data assets shape mergers and acquisitions. Using employees' job skills to measure firm-level AI capability, we show that AI-capable firms are more likely to acquire data-intensive targets. These transactions generate higher announcement returns and larger combined shareholder gains without higher acquisition premia. Post-merger, AI–data mergers exhibit substantial increases in patenting, citations and patent value, particularly in AI-related technologies. Overall, our findings identify AI–data complementarity as a central mechanism driving acquisition selection, value creation, and post-merger innovation.

Keywords: Mergers and Acquisitions, Artificial Intelligence, Data Assets, Human Capital, Innovation

JEL Codes: G34, O14, O31, O33, M15, L25

*We are grateful for helpful comments from Tania Babina (discussant), Andrey Golubov (discussant), Yanru Han, Gerry Hoberg, Kai Li, and participants at the AFA, JAAF, MARC Conference, University of Macau, and Westminster Business School Symposium.

†Zhejiang University, e-mail: yfang39@zju.edu.cn

‡HKUST and ABFER, e-mail: goyal@ust.hk

§Singapore Management University, e-mail: yingxuan.he.2025@phdacc.smu.edu.sg

¶Zhejiang University, e-mail: zilzhang@zju.edu.cn

1 Introduction

How does artificial intelligence (AI) reshape the value and reallocation of data assets within firms? Data has emerged as a central input in modern production, driving innovation, productivity growth, and strategic decision-making across industries.[1] Yet, the economic value of data depends critically on a firm's ability to transform raw information into actionable insights. Firms with strong AI expertise are better positioned to extract value from large datasets (Farboodi and Veldkamp, 2021; Abis and Veldkamp, 2024; Mihet, Rishabh, and Gomes, 2026), while the development and improvement of AI models themselves rely on large volumes of training data to enhance algorithmic performance and spur innovation (Beraja, Yang, and Yuchtman, 2023; Bartlett, McCrary, and O'Hara, 2025). Together, these forces imply a fundamental complementarity between AI capabilities and data assets. More broadly, AI has been characterized as a general-purpose technology that reshapes production processes and organizational boundaries (Agrawal, Gans, and Goldfarb, 2019).

If AI and data are complementary inputs, reallocating data across firms should generate value when it moves to owners with stronger AI capabilities. Mergers and acquisitions (M&As) provide a natural setting to examine this mechanism. Through acquisition, firms can internalize and integrate complementary intangible assets that may be difficult to coordinate through arm's-length contracting. The joint value of AI capability and data assets may require coordinated investments in model development, data infrastructure, and organizational routines that are difficult to specify ex ante. Theories of firm boundaries emphasize that when complementary investments are hard to contract upon, ownership facilitates integration and surplus internalization (Aghion and Holden, 2011; Rhodes-Kropf and Robinson, 2008). In this framework, the value of data depends on the acquiring firm's AI capability, and ownership structures that combine complementary intangible assets within a single organization should generate a greater surplus.

These considerations yield clear empirical implications. First, AI-capable firms should be more likely to acquire data-intensive targets. Second, transactions that com-

[1] See, for example, Begenau, Farboodi, and Veldkamp (2018), Zhu (2019), Brynjolfsson and McElheran (2019), Cong, Xie, and Zhang (2021), Goldstein, Spatt, and Ye (2021, 2025), He, Huang, and Zhou (2023).

bine strong acquirer AI capability with substantial target data assets should generate greater shareholder value. Third, if complementarity operates through knowledge production, such mergers should result in stronger post-merger innovation outcomes. We take these predictions to the data.

This paper studies whether AI–data complementarity shapes acquisition choices, market valuation, surplus allocation, and post-merger innovation. We construct firm-level measures of AI capability based on workforce skill composition following Babina et al. (2024) and develop novel measures of target data intensity using textual disclosures and large language models. To examine acquisition selection, we adopt a conditional logit framework comparing realized acquirer–target pairs to matched counterfactual combinations, following Bena and Li (2014).

We document consistent evidence across multiple dimensions. First, acquisition patterns reflect systematic AI–data alignment. At both the industry and firm levels, merger activity is disproportionately concentrated between high-AI acquirers and data-intensive targets. AI-capable firms are significantly more likely to acquire data-rich targets relative to otherwise similar firms in matched choice sets. These findings indicate that AI capability expands the set of economically attractive acquisition targets by increasing the marginal value of data assets.

Second, AI capability is positively associated with acquisition performance, and this relationship is driven by complementarity with data assets. Acquirers with stronger AI capabilities experience significantly higher cumulative abnormal returns (CARs) around deal announcements. However, these gains are concentrated in transactions involving data-intensive targets: mergers combining high-AI acquirers with high-data targets generate substantially larger announcement returns than other acquirer–target combinations. Deals involving AI capability or data intensity in isolation do not exhibit comparable valuation effects. In addition, the combination of high-data acquirers and high-AI targets does not generate significant acquirer CARs. These findings suggest that markets recognize and price acquirer AI - target data complementarities at the time of acquisition announcements.

Importantly, these valuation gains reflect greater surplus creation rather than sim-

ple redistribution. Transactions undertaken by high-AI acquirers generate larger combined wealth effects without paying higher acquisition premia. As a result, AI-capable acquirers capture a larger share of total merger gains, consistent with stronger bargaining positions associated with complementary capabilities.

A central concern is that AI-capable firms may differ systematically from other acquirers along unobserved dimensions. To address this issue, we exploit the geographic variation in the local AI talent supply, following Babina et al. (2024), as an instrument for firm-level AI capability. The instrumental variable estimates reinforce the baseline findings, supporting a causal interpretation of the relationship between AI capability and acquisition performance. One may also be concerned that our data intensity measure captures factors beyond firms' data assets. To address this, we replicate the baseline regressions with an alternative GPT-based measure that is designed to capture data dependence rather than risk-related language in targets' 10-K filings and find consistent results. In addition, to confirm that our results are not driven by firms' vulnerability to cybersecurity concerns, we replace our baseline data intensity measure with a standard cybersecurity proxy provided by Florackis et al. (2023). We find that our baseline results are not replicated using this cybersecurity measure.

We then examine whether these valuation gains translate into real innovative outcomes. Prior research suggests that AI promotes innovation by leveraging large datasets to uncover patterns and opportunities (Cockburn, Henderson, and Stern, 2019). To isolate integration effects from selection, we adopt a quasi-experimental design comparing completed mergers with withdrawn bids, following Bena and Li (2014) and Seru (2014). Using patent applications and citations from the Extended KPSS dataset (Kogan et al., 2017), we find that completed mergers combining high-AI acquirers with data-intensive targets exhibit substantially larger proportional increases in post-merger patenting, citations and patent value relative to withdrawn bids and other transaction types. The patent applications and patent value gains are also pronounced in AI-related patenting, consistent with the integration of newly acquired data into AI research programs. In contrast, mergers involving AI capability or data assets in isolation do not generate comparable innovation responses.

Across acquisition selection, market valuation, surplus allocation, and post-merger innovation, the evidence consistently points to AI–data complementarity as the central mechanism shaping merger outcomes. Value creation is concentrated in transactions that jointly combine strong acquirer AI capability with substantial target data assets, suggesting that AI reshapes firm boundaries by increasing the returns to internalizing complementary data and analytical capabilities within a single organization.

This paper contributes to the corporate finance literature in three primary ways. First, we identify AI–data complementarity as a novel and economically meaningful source of merger gains. While prior work emphasizes technological similarity and innovation-based synergies in mergers (Bena and Li, 2014), and shows that technologically advanced acquirers facilitate technology adoption at targets (Ma, Ouimet, and Simintzi, 2025), we demonstrate that the value of data assets depends critically on complementary AI capabilities at the acquiring firm. Complementarity drives not only acquisition selection but also announcement returns, combined wealth effects, and post-merger innovation.

Second, we contribute to the growing literature on artificial intelligence in corporate finance. Prior research shows that AI adoption improves internal decision-making, including corporate governance (Erel et al., 2021), asset management (Sheng et al., 2024), and forecasting accuracy (Cao et al., 2024). Related work emphasizes that firms' ability to transform data into knowledge shapes productivity and innovation (Abis and Veldkamp, 2024). We examine a distinct margin: how AI capability affects firms' external investment decisions and surplus allocation in M&A transactions.

Third, our findings inform the literature on intangible assets and firm boundaries. The rise of intangible capital—including software, data, and organizational knowledge—has reshaped firm value and competitive dynamics (Haskel and Westlake, 2017; Crouzet et al., 2022). Data has been conceptualized as a productive, non-rival input generating increasing returns through reuse and recombination (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2021), with important implications for market structure and valuation (Begenau, Farboodi, and Veldkamp, 2018; Veldkamp and Chung, 2024). We provide transaction-level evidence that the economic returns to data depend on complementary

AI capabilities and that mergers serve as a mechanism for internalizing these complementarities. By linking acquisition matching, market valuation, surplus allocation, and post-merger innovation, we show that AI capability fundamentally alters the incentives to integrate data assets within firm boundaries.

2 Data and Variables

We construct our sample from the SDC Platinum database and focus on U.S. acquisitions announced between January 1, 2009, and December 31, 2019. Following Masulis, Wang, and Xie (2007) and Gokkaya, Liu, and Stulz (2023), we apply the following filters: (1) the target is a U.S. firm; (2) the acquirer holds less than 50% of the target prior to the transaction and at least 50% afterward; (3) the disclosed deal value exceeds $1 million and is at least 1% of the acquirer's market capitalization, measured 11 trading days prior to the announcement; (4) the transaction is neither an intended deal[2] nor a remaining-interest acquisition; and (5) the acquirer does not operate in the finance, insurance, and real estate sector (FIRE), as defined by SIC codes. After applying these filters, the resulting SDC sample contains 5,378 unique transactions with complete deal information, including announcement dates, deal values, payment methods, and acquirer ownership. We then merge the filtered SDC transactions with Compustat and CRSP.[3] Compustat provides firm-level financial variables, and CRSP daily returns are used to compute cumulative abnormal returns (CARs).

2.1 Measuring AI Capabilities of Firms

Given the central role of human capital in deploying AI systems, we measure firms' AI capabilities based on the prevalence of AI-related skills in their workforce, based on LinkedIn data.[4] The dataset contains self-reported information on education (institu-

[2] In SDC, an *intended deal* indicates that the acquirer has announced that it proposes or expects to make an acquisition (a designation often used for repurchases).

[3] We use the updated SDC–Compustat matching provided by Michael Ewens https://github.com/michaelewens/SDC-to-Compustat-Mapping. For details, see Ewens, Peters, and Wang (2025), Phillips and Zhdanov (2013).

[4] LinkedIn is the largest professional networking platform and had 169.9 million U.S. users in 2019.

tions, fields of study, enrollment periods), work experience (employers, titles, dates), skills, and other attributes. We match acquirers and targets to LinkedIn firm pages in two steps. First, we use company website URLs from Compustat as the primary identifier to link firms to their LinkedIn corporate profiles. When URLs are missing, we supplement with manual web searches. The remaining unmatched firms are hand-matched using name, location, founding year, and industry. This procedure yields an overall match rate of approximately 80%. 13,263,435 employment records from 8,521,432 LinkedIn individual profiles are matched.

For matched firms, we construct AI capability measures using employee profile data. AI-related skills are identified using the taxonomy developed by Babina et al. (2024), which includes 67 skills such as *Deep Learning, XGBoost, NLP,* and *Machine Learning*. An employee is classified as an AI employee if they report at least one AI-related skill. For example, an employee listing both *Deep Learning* and *XGboost* is counted as one AI employee with two AI skills.

We track employees and reported skills at a monthly frequency and construct firm-level measures of total employees, AI employees, and AI skills. For each acquirer and target firm, we compute average monthly values over the 12 months preceding the announcement. This yields measures of average monthly employees, AI employees, and AI skills. For each transaction, *Acq_AISkill* denotes the number of AI skills and *Acq_AIEmployee* denotes the number of AI employees at the acquirer, averaged over the 12 months preceding the announcement. Target firm measures are constructed analogously. After excluding observations with missing values, the final sample comprises 4,471 unique transactions.

— Figure 1 about here —

Figure 1 maps the geographic distribution of AI talent across U.S. counties using work location information reported by employees on their LinkedIn pages. We geolocate AI skills and aggregate counts at the county level. The resulting heatmap shows strong concentrations in established tech hubs and university centers—such as the San Francisco Bay Area, Seattle, Boston, New York City, Austin, Denver, and Washington, D.C.—which

closely mirrors the geography of U.S. tech labor markets and support the validity of our LinkedIn-based AI measures.

2.2 Measuring Data Intensity

We develop a measure of data intensity based on the frequency of data security-related keywords in Item 1A risk-factor disclosures from firms' 10-K filings. This approach is motivated by the idea that firms handling large or complex datasets face greater data security risks because they must protect sensitive information, comply with evolving data regulations (e.g., GDPR, CCPA), and mitigate potential breach-related costs. Ramadorai, Uettwiller, and Walther (2025) show that, as these regulations have become more binding, firms' data privacy disclosures have become more detailed and more truthful. As data volumes increase, exposure to these risks also rises, making the extent of data security disclosures a useful proxy for data intensity.

In modern regulatory environment, the value of data is inseparable from the cost of protecting it: Firms that hold substantial amounts of data also need to safeguard them from breach risks. As a result, the demand for data security increases with data intensity. To construct this measure, we compile a list of data security keywords from the Cloud Security Alliance's *Data Security Glossary*. Specifically, we employ a custom Python program to calculate the frequency of these keywords in item 1A risk-factor disclosures, where firms are required to report material risks, including threats of data breaches. Our data intensity measure, *DataIntensity*, is defined as the total count of these keywords in a firm's Item 1A disclosure. This measure captures the emphasis placed on data-related risks and serves as a proxy for the scale and complexity of the firm's data assets.

For robustness, we construct an alternative measure of data intensity using a large language model (LLM). Following Eisfeldt et al. (2026), we use their prompt verbatim and GPT-4.1 to analyze the first 3,000 words of Item 1 (business descriptions) in target firms' 10-K filings to assess their potential access to data. For the hyperparameters, we set the temperature to zero to minimize randomness and ensure consistent outputs. The remaining hyperparameters are left at their default values. Specifically, we instruct the LLM to

evaluate the extent to which each filing addresses six data-related dimensions: (i) the nature of the business, (ii) firm scale and reach, (iii) data collection mechanisms, (iv) data utilization practices, (v) data infrastructure and management, and (vi) data regulation and privacy. Each dimension is assigned a score from 0 to 3, where 0 denotes no data-related content and 1, 2, and 3 correspond to low, medium, and high levels of data relatedness, respectively. The overall data relatedness measure is defined as the sum of all six dimension-specific scores.

2.3 Other Variables

We follow the standard M&A event-study methodology to estimate acquirer cumulative abnormal returns (CARs). We estimate the market model over the $[-240, -11]$ trading-day window preceding the announcement. The event window spans $[-3, +3]$ trading days around the announcement date. We compute abnormal returns (ARs) as the difference between the actual and predicted returns, and define CARs as the cumulative sum of ARs over the event window. For robustness, we also consider alternative event windows of $[-4, +4]$ and $[-5, +5]$.

We measure deal-level innovation using firm-level patent data from the Extended KPSS dataset (Kogan et al., 2017). We define *Patent Applications* as (i) the sum of patents applied for by the acquirer and the target in the five years preceding the acquisition, and (ii) the number of patents applied for by the combined firm in the five years following the acquisition. The variable *Patent Citations* is constructed analogously using forward citation counts. To examine AI-related innovation, we use the Artificial Intelligence Patent Dataset (AIPD)[5] and match patents to firms using the USPTO PatentsView and Google Patents Public datasets. *AI Patent Applications* is defined as (i) the sum of AI patents applied for by the acquirer and the target in the pre-acquisition period and (ii) the number applied for by the combined firm in the post-acquisition period. Patents are classified as AI-related if the AIPD indicator *predict93_any_ai* equals one. The variable *predict93_any_ai* equals one if a patent is predicted to be AI-related by any of the technology-

[5]Developed by the USPTO's Office of the Chief Economist, the AIPD uses machine-learning models to identify AI-related patents and pre-grant publications in U.S. patents from 1976–2023 (Pairolero et al., 2025).

component models under the 93% threshold, and zero otherwise (Pairolero et al., 2025; Giczy, Pairolero, and Toole, 2022). *Patent Value* is defined as the natural logarithm of one plus total value creation of patents received (i) by the acquirer and the target in the five years preceding the acquisition, and (ii) by the combined firm in the five years following the acquisition. Patent value is calculated by KPSS (`xi_real`), and is deflated to 1982 (million) dollars using the CPI. Together, these measures capture both the quantity and the quality of innovation around each deal.

Our empirical specifications include a standard set of control variables. At the firm level, we control for average employee number, size, Tobin's Q, leverage, ROA, the intangible-asset ratio, and R&D expense, as well as the whether the target is a public company. Deal-level controls include the payment method (*AllStock*), relative deal value (*Rel_DealVal*), an indicator for challenged deals (*Challenge*), a cross-border indicator (*Crossborder*), diversification (*Conglomerate*), and an indicator for high-technology deals (*Hightechdeal*), defined following Loughran and Ritter (2004) and Li et al. (2018)). Detailed variable definitions are provided in Appendix A.

2.4 Summary Statistics

Table 1 reports summary statistics across five panels. Panel A reports statistics for acquiring firms. The average acquirer employs approximately 1,455 workers; however, firm size is highly skewed, with the largest acquirers employing more than 80,000 individuals. On average, acquirers employ about 23 AI-skilled workers and report 32 AI-related skills, though the large standard deviations indicate substantial heterogeneity in AI capability across firms.[6] The log-transformed AI measures, $LnAcqAIEmp$ and $LnAcqAISkill$, have means of 1.28 and 1.38, respectively. Acquirers exhibit moderate leverage (mean 0.26) and hold a sizable proportion of intangible assets (mean intangible ratio 0.28). R&D intensity averages 3%, while profitability (ROA) averages 2%. The mean Tobin's Q is 2.00, consistent with relatively strong growth opportunities among acquiring firms.

Panel B reports statistics for target firms. Targets are substantially smaller than ac-

[6]Employment is recorded monthly, and annual averages are reported over the 12 months preceding the transaction.

quirers, with an average of 293 employees. AI capability among targets is limited relative to acquirers: the average target employs fewer than four AI-skilled workers and reports fewer than five AI-related skills. The mean $LnTarAISkill$ is 0.48. Approximately 15% of targets are publicly listed. Data intensity measures, available for a subsample of targets, display considerable dispersion. The mean value of the continuous data-intensity measure is around 650. The GPT-based data intensity measure ranges from 0 to 18, with a mean value of 8.06. To examine whether our baseline results are driven by cybersecurity rather than data intensity, we adopt a cybersecurity measure provided by Florackis et al. (2023). This measure has a mean value of 0.26. Overall, these statistics indicate that acquirers are larger and possess stronger AI-related human capital, consistent with evidence that larger firms have better access to financing and greater capacity to invest in proprietary technologies (Archer and Faerber, 1966; Kahle and Stulz, 2013; De Guevara, Maudos, and Salvador, 2022).

— Table 1 about here —

Panel C describes transaction characteristics. Only 4.0% of deals are financed entirely with stock, and challenged transactions are rare, accounting for about 1% of the sample. Approximately 39% of transactions are conglomerate mergers, while 5% involve cross-border acquisitions. The average relative deal value is 0.25, implying that the typical transaction represents one-quarter of the acquirer's total assets. High technology deals account for 21% of the sample, defined as transactions in which both the acquirer and target operate in high tech industries. Acquirer announcement returns are positive on average: mean CARs equal 1% across the $[-3, +3]$, $[-4, +4]$, and $[-5, +5]$ windows, with substantial dispersion. For the subsample with available target return data, targets experience large positive announcement effects, with mean target CARs of 14% over both $[-2, +2]$ and $[-3, +3]$ windows. Combined announcement returns average 3%. The acquirer's share of total synergy is close to zero on average, though value appropriation varies substantially across deals.

Panel D summarizes post-merger innovation outcomes. In the subsample with patent data, combined firms file an average of 34.55 AI-related patent applications

and 99.89 total patent applications, and receive an average of 412.56 forward citations. Large standard deviations indicate substantial heterogeneity in innovative activity across merged firms.

3 Empirical Results

3.1 The Pursuit of Data by High-AI Acquirers

3.1.1 Industry distribution of merger pairing

We test whether AI–data complementarity shapes M&A activity by examining whether firms with stronger AI capabilities are more likely to acquire data-intensive targets. Figure 2 summarizes merger frequency using an industry-level pairing approach based on SDC industry classifications. We retain the 15 acquirer industries and 15 target industries with the highest merger frequencies and rank them by acquirer AI capability and target data intensity. Merger propensity is measured as the number of deals for each acquirer–target industry pair. The vertical axis ranks acquirer industries by AI capability (highest at the top), and the horizontal axis ranks target industries by data intensity (highest on the left).

— Figure 2 about here —

The figure reveals a clear association between acquirer AI capability, data intensity, and M&A activity. Merger activity is concentrated in the upper-left region of the matrix, where high-AI acquirer industries pair with high-data-intensity target industries. Darker cells (i.e., higher deal counts) in this region indicate that firms with strong AI capabilities disproportionately acquire data-rich targets, consistent with AI–data complementarity rather than random industry matching.

Although many deals occur within industry—reflecting the well-known tendency toward within-industry consolidation—there is also substantial cross-industry activity. This pattern is especially pronounced among high-AI acquirers: their cross-industry

mergers disproportionately involve data-intensive targets, suggesting strategic motives beyond industry consolidation.

We also observe a pronounced asymmetry between high- and low-AI acquirer industries. High-AI industries engage in acquisitions across a broad range of target industries, particularly those with high data intensity. In contrast, low-AI industries complete fewer deals overall, and their activity is more evenly distributed or concentrated in low-data-intensity industries. This asymmetry suggests that AI capability expands the set of feasible and attractive targets. Consistent with this interpretation, data-intensive industries such as software, business services, and telecommunications attract a disproportionate share of acquisitions by high-AI acquirers.

Overall, the figure provides visual evidence that M&A activity is systematically patterned around AI–data complementarities. Firms with strong AI capabilities are more likely to acquire data-intensive targets, both within and across industries. This pattern supports the view that AI talent and data assets jointly shape merger decisions, beyond traditional motives related to within-industry consolidation.

3.1.2 Firm-level analysis of merger pairing

To formally examine whether AI-capable acquirers are more likely to pair with data-rich targets, we estimate a firm-level selection model using a conditional logit framework following Bena and Li (2014). The model compares realized acquirer–target pairs to matched counterfactual pairs constructed from plausible alternative acquirers and targets. Identification comes from variation within deal-specific choice sets, allowing us to assess which combinations of acquirer and target characteristics are more likely to be observed relative to other feasible matches within the same deal.[7]

We construct three control samples, each defining a set of plausible alternative acquirers and targets for the firms involved in observed deals. For each actual acquirer and target, we identify up to five matched hypothetical firms, generating a set of potential merger combinations. We then pool realized and hypothetical pairs to examine which

[7] In a conditional logit model with deal fixed effects, coefficients are identified from within-deal variation and can be interpreted as changes in the relative odds that a given acquirer–target pair is chosen over other feasible pairs in the same deal-specific choice set.

characteristics predict observed M&A matches. Control firms are drawn from Compustat/CRSP as of the fiscal year-end preceding the announcement and must not have participated in any M&A activity—as either acquirers or targets—during the prior three years.

The first control group matches firms on industry and size, capturing the well-documented clustering of M&A activity within industries (Andrade, Mitchell, and Stafford, 2001; Harford, 2005). For each actual firm, we identify up to five matching firms based on SIC industry classification and total assets. Industry matching begins at the four-digit SIC level and is progressively broadened to the three- and two-digit levels if available peers are fewer than five.[8] Within the matched industry pool, we select the five firms with the smallest absolute differences in size relative to the actual firm. Firm size is measured by total assets at the fiscal year-end preceding the announcement. We refer to these firms as *Industry-Size-Matched* controls.

The second control group additionally conditions on operating performance by incorporating return on assets (ROA), a key determinant of M&A activity (Healy, Palepu, and Ruback, 1992). Using propensity-score matching, we estimate a probit model with size and ROA as predictors and match each actual firm to its five nearest neighbors within the industry-matched pool.

The third and most stringent control group additionally incorporates Tobin's Q to account for differences in growth opportunities, valuation, and strategic fit—factors known to influence merger decisions (Andrade, Mitchell, and Stafford, 2001; Shleifer and Vishny, 2003; Rhodes-Kropf and Viswanathan, 2004; Rhodes-Kropf and Robinson, 2008). We add Tobin's Q to the propensity-score model used in the second control group and select the five nearest firms from the industry-matched pool as control firms.

We then estimate the following conditional logit model:

$$\begin{aligned} Acquirer\text{-}Target_{ijm} = \alpha &+ \beta_1\, HighAcqAI_{im} \times HighTarData_{jm} \\ &+ \beta_2\, HighAcqAI_{im} + \beta_3\, HighTarData_{jm} \\ &+ \beta_4\, \mathbf{X}^A_{im} + \beta_5\, \mathbf{X}^T_{jm} + \beta_6\, \mathbf{X}^D_{ijm} \\ &+ \mu_m + \varepsilon_{ijm}, \end{aligned} \tag{1}$$

[8] We exclude deals for which the relevant firm has fewer than five peers at the two-digit SIC level.

where i and j index acquirers and targets, respectively, and ij denotes all possible acquirer-target pairings within deal m. The dependent variable, $Acquirer\text{-}Target_{ijm}$, equals one for realized matches and zero for counterfactual pairs. Control variables include acquirer and target firm fundamentals (size, Tobin's Q, leverage, ROA, sales growth, and R&D expense) and deal-level indicators for same-state location and cross-industry deals. We restrict the analysis to transactions in which both acquirer and target are publicly traded.

We estimate equation (1) using three matched samples: *Industry-Size-Matched*, *Industry-Size-ROA-Matched*, and *Industry-Size-ROA-Tobin's Q-Matched*. In each sample, every target is paired with up to five matched counterfactual acquirers, and each acquirer is paired with up to five matched counterfactual targets. We retain only deal choice sets with at least three pseudo-acquirers and three pseudo-targets. The resulting samples include 533, 515, and 508 realized deals, respectively, and, when combined with matched controls, each sample contains more than 4,000 observations.

— Table 2 about here —

Table 2 reports estimates of the acquirer–target pairing model across three increasingly stringent matched samples. Across all specifications, the interaction between high acquirer AI capability and high target data intensity is positive and statistically significant. This indicates that, within a given deal choice set, pairings that combine AI-capable acquirers with data-intensive targets are more likely to be observed than alternative feasible pairings.

To aid interpretation, consider Column (3). For high-AI acquirers, moving from a low-data to a high-data target increases the relative odds of a match by a factor of approximately $e^{\beta_3+\beta_1} \approx e^{0.487} \approx 1.63$. Within a given choice set, high-data targets have approximately 63% higher odds of being matched with high-AI acquirers than low-data targets. In contrast, for low-AI acquirers, target data intensity has little effect on relative match odds. Equivalently, the interaction term implies that the effect of target data intensity on match odds is nearly twice as large for high-AI acquirers as for low-AI acquirers ($e^{\beta_1} \approx 1.86$), with all comparisons made within the same choice set.

The interaction coefficient is stable across the industry-size, industry-size-ROA, and industry-size-ROA–Tobin's Q matched samples, suggesting that the pattern is not driven by differences in scale, operating performance, or growth opportunities. The main effects of high acquirer AI capability and high target data intensity are negative or statistically insignificant. In the presence of the interaction, these coefficients capture relative attractiveness evaluated at the omitted category of the interacting variable and should not be interpreted as standalone predictors of acquisition likelihood.

As an additional robustness check, we re-estimate the model excluding the matching variables used in each respective matching procedure, with results reported in Columns (4)–(6). The estimates remain quantitatively similar. Overall, the results provide firm-level evidence that merger pairing is systematically related to the alignment between acquirer AI capability and target data assets. Within deal-specific choice sets, AI-capable acquirers exhibit a stronger propensity to match with data-intensive targets, consistent with AI–data complementarities shaping merger selection.

3.1.3 Supplementary evidence on human capital

To provide supplementary evidence on AI–data complementarity, we examine changes in acquirers' workforce composition around transactions, focusing on data-analytics roles. We expect AI-intensive firms to maintain larger stocks of data-related human capital prior to acquiring data-rich targets, strengthening their capacity to integrate and deploy those data assets after acquisition. To test this hypothesis, we examine whether high-AI firms employ more workers in data-oriented roles—such as *data analysts* and *data scientists*—identified by job titles containing the keyword "data" (e.g., "senior data analyst") in the years leading up to acquisitions. We construct two measures: (i) the total number of data-analytics workers and (ii) the subset of these workers who also report at least one AI-related skill, as defined above. These measures capture differences in firms' data-oriented staffing and the extent to which data roles are combined with AI expertise.

— Figure 3 about here —

Figure 3 plots industry-adjusted data-analytics employment over a 13-year window

$[-10, +2]$ around deal announcements (year 0), comparing high-AI and low-AI acquirers. The two panels provide complementary evidence on how AI-intensive firms accumulate human capital consistent with AI–data complementarity, both before and after acquisitions.

Panel A shows a clear and persistent divergence between high-AI and low-AI acquirers in industry-adjusted data-analytics employment over the years surrounding the deal. High-AI acquirers exhibit a steady upward trend in data-analytics employment beginning many years before the transaction, consistent with a gradual buildup of data-related human capital. Importantly, these patterns emerge well before the deal year (year 0), suggesting that differences in data-analytics staffing are unlikely to reflect only short-run adjustments immediately around the acquisition. The absence of a sharp break at the deal year is consistent with this interpretation. Rather than exhibiting a discrete post-merger spike, data-analytics employment evolves smoothly through and after the acquisition, consistent with the view that data capabilities function as a durable organizational asset.

Panel B focuses on data-analytics workers who also possess AI-related skills. The divergence between high-AI and low-AI acquirers is even more pronounced for this subgroup. High-AI acquirers steadily increase the number of workers with combined AI and data expertise, while low-AI acquirers exhibit a persistent decline. This pattern suggests that high-AI firms are not expanding generic data functions alone; instead, they disproportionately accumulate talent that complements AI development, such as model training, feature engineering, and advanced analytics. Overall, the evidence suggests that AI-intensive firms build hybrid skill sets that enable effective use of large-scale data in AI applications, rather than relying solely on traditional data roles.

3.2 Acquirer AI and Announcement Returns

3.2.1 Baseline result

Building on evidence that AI capabilities and data assets are aligned in M&A decisions, we next examine whether this alignment is reflected in market value at deal announcement.

Specifically, we estimate how acquirer AI capability relates to cumulative abnormal returns (CARs) around deal announcements.

$$CAR_{i,t} = \alpha + \beta_1 AcquirerAI_{i,t-1} + \mathbf{X}'_{i,t-1}\boldsymbol{\beta}_2 + \mathbf{Z}'_{j,t-1}\boldsymbol{\beta}_3 + \mathbf{D}'_{ij,t}\boldsymbol{\beta}_4 + \text{Industry FE} + \text{State} \times \text{Year FE} + \varepsilon_{i,t}. \tag{2}$$

The key independent variable, $AcquirerAI$, measures the acquirer's AI capability in the 12 months preceding the announcement. We proxy for AI capability using two log-transformed measures: $LnAcqAISkill$ (average number of AI skills) and $LnAcqAIEmp$ (average number of AI-skilled employees), both computed over the 12-month pre-announcement window.

Following the M&A literature, we include acquirer controls $\mathbf{X}$ (average employee number, size, Tobin's Q, leverage, ROA, intangible asset ratio, and R&D expense), target controls $\mathbf{Z}$ (target public status), and deal controls $\mathbf{D}$ (payment method, relative deal value, an indicator for challenged deals, a cross-border indicator, a diversification indicator, and an indicator for high-technology deals). Regressions include acquirer industry fixed effects (four-digit SIC) and state-by-year fixed effects.

— Table 3 about here —

Table 3 reports announcement-return regressions relating acquirer AI capability to short-window stock market reactions to M&A announcements. Columns (1)–(3) use AI skill intensity, measured by $LnAcqAISkill$. The coefficients on *LnAcqAISkill* are positive and statistically significant across all three event windows. A one-standard-deviation increase in $LnAcqAISkill$ is associated with a 0.795 percentage point increase in $CAR[-3,+3]$, which is approximately 79.5% of the sample mean. The magnitude remains similar as the event window expands from $[-3,+3]$ to $[-5,+5]$. This pattern is consistent with the market viewing acquisitions by AI-capable firms more favorably, potentially reflecting expectations of greater synergy realization.

Columns (4)–(6) use an alternative measure of AI capability—the log number of AI-skilled employees—and yield similar results. The coefficients on $LnAcqAIEmp$ are positive and statistically significant across event windows, indicating that the positive an-

nouncement effects are not driven by a particular measure of AI capability. Consistency across skill- and headcount-based measures strengthens the interpretation that AI human capital is value-relevant in M&A.

Overall, the results suggest that acquirer AI capability is positively priced by financial markets at deal announcements. Investors appear to expect that AI-intensive firms realize greater synergies, consistent with the view that AI human capital is a value-relevant asset in M&A.

3.2.2 Instrumental variable approach

A potential concern is that firms with stronger measured AI capabilities may differ along other dimensions that also predict acquisition performance, generating higher announcement returns for reasons unrelated to AI. To address this concern, we implement an instrumental variables (IV) strategy following Babina et al. (2024), exploiting plausibly exogenous variation in firms' access to local AI talent. This approach leverages regional labor market conditions that affect firms' ability to recruit AI-skilled workers and, conditional on controls and fixed effects, are plausibly unrelated to deal-specific value creation.

We construct a distance-based instrument that captures firm-level exposure to local AI talent supply, measured using the share of AI graduates from nearby universities and weighted by university size. We construct the instrument in three steps. First, using data from Babina et al. (2024), we identify "AI-strong" universities as those in the top 25% by AI researcher count during 2004–2008. This pre-sample period helps mitigate concerns that university AI strength is endogenous to subsequent firm behavior. Second, we measure each firm's proximity to these universities using NBER county-level distance data and restrict attention to universities located within a 60-mile radius of the firm's headquarters to capture localized labor-market effects. Third, we compute a size-weighted average of AI graduate shares from these nearby universities. Formally, the instrument is defined as:

$$IV_{i,t} = \sum_{j \in J_{it}} \left(\frac{AI_Grads_{jt}}{Total_Grads_{jt}} \right) \times Weight_Size_{jt}, \tag{3}$$

where AI_Grads_{jt} is the number of AI-related first-job graduates from university j in year

t, $Total_Grads_{jt}$ is the total number of graduates from university j in year t, and J_{it} is the set of AI-strong universities located within a 60-mile radius of the county containing firm i's headquarters. The weight term, $Weight_Size_{jt}$, reflects university size:

$$Weight_Size_{jt} = \frac{\ln(Total_Grads_{jt})}{\sum_{\ell \in J_{it}} \ln(Total_Grads_{\ell t})}. \tag{4}$$

This instrument captures firm-level exposure to plausibly exogenous variation in local AI talent supply. We measure the instrument using values from the year preceding the announcement to align with the timing of acquirer AI capability in the first stage. The first-stage regression is:

$$\begin{aligned} LnAcqAISkill_{i,t-1} = \beta_0 + \beta_1 IV_{i,t-1} + \mathbf{X}'_{i,t-1}\boldsymbol{\beta}_2 + \mathbf{Z}'_{j,t-1}\boldsymbol{\beta}_3 + \mathbf{D}'_{ij,t}\boldsymbol{\beta}_4 \\ + \text{Industry FE} + \text{State} \times \text{Year FE} + \nu_{i,t}. \end{aligned} \tag{5}$$

We then use the fitted values from the first stage in the second stage to estimate the effect of AI capability on announcement returns:

$$\begin{aligned} CAR_{i,t} = \alpha + \beta_1 \widehat{LnAcqAISkill}_{i,t-1} + \mathbf{X}'_{i,t-1}\boldsymbol{\beta}_2 + \mathbf{Z}'_{j,t-1}\boldsymbol{\beta}_3 + \mathbf{D}'_{ij,t}\boldsymbol{\beta}_4 \\ + \text{Industry FE} + \text{State} \times \text{Year FE} + \varepsilon_{i,t}. \end{aligned} \tag{6}$$

By exploiting plausibly exogenous variation in local AI talent supply, this two-stage least squares (2SLS) approach strengthens identification and mitigates concerns that the baseline CAR results reflect unobserved differences in acquirer quality.

— Table 4 about here —

Table 4 reports the instrumental variable results linking acquirer AI capability to announcement returns. Columns (1) and (2) present the first-stage regressions without and with controls for industry fixed effects and state×year fixed effects, respectively. In both specifications, the local AI talent instrument is strongly and positively associated with $LnAcqAISkill$. The first-stage F-statistic is 23.84 without fixed effects and 41.66 with fixed effects, exceeding conventional thresholds for weak instruments, indicating strong

instrument relevance. This result is consistent with geographic exposure to AI-skilled labor being a key determinant of firms' measured AI capability.

Columns (3)–(5) report the second-stage 2SLS estimates for CARs over the $[-3, +3]$, $[-4, +4]$, and $[-5, +5]$ event windows, corresponding to specifications without fixed effects in the first stage. Across all event windows, the coefficient on instrumented $LnAcqAISkill$ is positive and statistically significant. Columns (6)–(8) report the corresponding second-stage 2SLS estimates with the full set of fixed effects included in the first stage. The estimated coefficients remain positive and statistically significant across all event windows, indicating that the baseline IV results are robust to alternative fixed-effects specifications.

The control variables are broadly consistent with prior evidence in the M&A literature. Firm size enters negatively and significantly, consistent with larger acquirers experiencing lower announcement returns. Leverage and profitability (ROA) are positively associated with CARs, consistent with financially stronger acquirers receiving more favorable market reactions. Target public status is also negatively related to announcement returns, consistent with greater competition and lower surplus capture in acquisitions of publicly traded targets.

Overall, the IV results reinforce the main conclusion: acquirer AI capability is positively related to acquisition performance, and the estimates are consistent with a causal interpretation under the IV identifying assumptions. By exploiting plausibly exogenous variation in local AI talent supply, this approach mitigates concerns that AI-intensive firms are simply better acquirers for unrelated reasons and strengthens the interpretation that AI capabilities contribute to superior M&A outcomes.

3.2.3 Combined Wealth Effects and Acquisition Premium

We next investigate whether high-AI acquirers are associated with greater combined wealth effects in mergers. Examining both acquirer and target returns allows us to assess overall value creation associated with acquirer AI capability. This analysis is limited to transactions involving publicly traded targets.

To measure the combined wealth effect, we construct *CombinedCAR* as the market-

capitalization-weighted average of the target's CAR over $[-41, +3]$ and the acquirer's CAR over $[-3, +3]$, using market capitalizations as of day -42 (target) and day -4 (acquirer) as weights, respectively (Dessaint, Eckbo, and Golubov, 2025). We then re-estimate our baseline CAR regression using *CombinedCAR* as the dependent variable.

— Table 5 about here —

Table 5 examines how acquirer AI capability relates to total value creation, value division between acquirers and targets, and target announcement returns for deals involving public targets. The results suggest that acquirer AI capability is associated with higher overall merger synergies and a larger share of gains accruing to acquirers, while also benefiting target shareholders.

Column (1) shows that *LnAcqAISkill* is positively and significantly associated with *CombinedCAR*, suggesting that mergers undertaken by more AI-capable acquirers are associated with greater total value for the combined firm. Column (2) shows no statistically significant association between *LnAcqAISkill* and *Premium*, suggesting that AI-capable acquirers do not systematically pay higher premiums to target shareholders. Column (3) shows a positive and marginally significant association between *LnAcqAISkill* and the acquirer's share of synergy gains. Together, these results are consistent with AI capability strengthening the acquirer's bargaining position, allowing acquirers to retain a larger share of merger gains without systematically transferring value through higher premia.

Columns (4) and (5) show that target CARs are significantly higher when acquirers have stronger AI capabilities. Thus, even though premia are not higher on average, target shareholders benefit from transactions involving AI-capable acquirers, potentially reflecting expectations of greater long-term value creation.

Taken together, the results suggest that acquirer AI capability is associated with higher total merger value and a larger share of gains accruing to acquirers, while still generating positive announcement effects for targets. These findings are consistent with AI capability being a source of economically meaningful value in M&A.

3.2.4 AI–Data Complementarity and Value Creation in Mergers

We next examine whether announcement returns differ systematically across acquirer–target combinations defined by acquirer AI capability and target data intensity. Building on our earlier evidence on merger pairing and workforce composition, this analysis asks whether transactions that align high acquirer AI capability with data-intensive targets are associated with more favorable market reactions than other deal types.

— Table 6 about here —

Table 6, Panel A reports regressions of acquirer announcement returns on indicators for mutually exclusive acquirer–target combinations based on acquirer AI capability, target AI capability, and target data intensity. Because these indicators are mutually exclusive and collectively exhaustive, several deal types are omitted as reference categories, and the reported coefficients should be interpreted relative to those omitted combinations.

The coefficient on $HighAcqAI \times HighTarData$ is positive and statistically significant for event windows $[-4, +4]$ and $[-5, +5]$. In Columns (4)–(6), which additionally include target industry-by-year fixed effects, the coefficient on the interaction term remains positive and statistically significant across all event windows. This indicates that deals combining AI-capable acquirers with data-intensive targets are associated with higher announcement returns relative to the omitted reference categories. The magnitude of the coefficient remains stable as the event window expands from $[-3, +3]$ to $[-5, +5]$, suggesting that the market reaction is not driven by short-lived price pressure.

By contrast, other acquirer–target combinations do not exhibit similarly large or statistically significant announcement gains. Transactions in which high-AI acquirers acquire low-data targets, or in which low-AI acquirers acquire data-intensive targets, are associated with smaller and statistically insignificant returns. Deals involving low-AI acquirers and high-AI targets are associated with negative announcement returns in the shorter event windows, although these estimates are imprecise. Taken together, these patterns indicate that neither acquirer AI capability nor target data intensity alone is sufficient to generate positive market reactions; rather, announcement gains are concen-

trated in transactions that combine strong acquirer AI capability with substantial target data assets.

To quantify these differences more precisely, we conduct pairwise tests comparing announcement returns across acquirer–target combinations. Relative to transactions in which low-AI acquirers acquire high-data targets, deals in which high-AI acquirers acquire data-intensive targets are associated with announcement returns that are higher by approximately 16.7 percentage points ($t = 3.17$). Relative to transactions in which high-AI acquirers acquire low-data targets, the high-AI–high-data combination is associated with announcement returns that are higher by approximately 9.9 percentage points ($t = 2.11$).

To examine whether our results are driven by the interaction between high acquirer data intensity and high target AI capability, we construct indicator variables, $HighAcqData$ and $LowAcqData$, which equals one if the acquirer's data intensity is above (below) the sample median. We then interact these variables with indicators of target AI capability and add their interactions into the baseline specification. As shown in Panel B, the coefficient on $HighAcqAI \times HighTarData$ remains positive and statistically significant across event windows and fixed-effects specifications. In contrast, $HighAcqData \times HighTarAI$ is negative and statistically insignificant. In some specifications, $LowAcqData \times HighTarAI$ is also negative and statistically significant. Deals in which both the acquirer and the target have high AI capability generate positive announcement returns, but the effects are not robust across specifications. Overall, among all two-way combinations, only transactions involving high acquirer AI capability and high target data intensity consistently yield significantly positive announcement returns.

Overall, the results are consistent with AI–data complementarity shaping market perceptions of merger value. Investors appear to respond more favorably to acquisitions in which AI-capable acquirers obtain data-intensive targets, while other combinations are associated with muted or negative announcement effects.

3.3 Robustness Tests Using Alternative Data Intensity Measures

The baseline results rely on a measure of target data intensity constructed from the frequency of data-related and data-security terms in the target firm's Item 1A risk disclosures. A potential concern is that this measure may capture firms' exposure to cybersecurity risk, rather than their underlying data intensity per se.

To address this concern, we conduct two robustness tests. First, we construct an alternative measure of data intensity using a generative AI-based approach. Specifically, following Eisfeldt et al. (2026), we use GPT-4.1 and their prompt (applied verbatim) to evaluate the extent to which a target firm's annual report reflects reliance on data as a key input into business operations, focusing on data generation, storage, and usage rather than risk disclosure. The resulting measure is designed to capture economically meaningful data dependence rather than cybersecurity exposure. Second, we directly employ the cybersecurity risk measure developed by Florackis et al. (2023) and investigate whether this measure can reproduce our baseline results, to examine if our findings are not driven by variation in cybersecurity risk exposure. Using these two alternative measures, we re-estimate both the merger-pairing analysis and the announcement return analysis. The results are reported in Table 7 and Table 8, respectively.

In the pairing analysis in Table 7, Columns (1)–(3) report results using the GPT-based measure of target data intensity, while Columns (4)–(6) present results using the cybersecurity-risk-based measure. The coefficient on $HighAcqAI \times HighTarData$ is positive and statistically significant only when data intensity is measured using the GPT-based approach. In contrast, all specifications using the cybersecurity-risk-based measure yield insignificant estimates.

In the deal announcement return analysis in Table 8, when using the GPT-based measure of data intensity, the coefficient on $HighAcqAI \times HighTarData$ is consistently positive and statistically significant across specifications, and is statistically larger than that on $HighAcqAI \times LowTarData$. In contrast, when data intensity is proxied by the cybersecurity-risk-based measure, the coefficient on $HighAcqAI \times HighTarData$ remains positive but becomes statistically insignificant. In addition, the coefficient on

$HighAcqAI \times LowTarData$ is of similar magnitude and becomes statistically significant in the $[-5, +5]$ window. The comparison using cybersecurity measure suggests that, conditional on high acquirer AI capability, variation in target cybersecurity exposure does not explain cross-sectional differences in announcement returns.

Overall, the results highlight that our baseline findings are robust to an alternative LLM-based measure of data intensity, while they are not replicated by a cybersecurity-risk-based proxy. This assures that the main effects are driven by economically meaningful data intensity rather than cybersecurity risk exposure.

— Tables 7 and 8 about here —

4 Innovation Outcomes following AI–Data Mergers

4.1 Post-Merger Innovation and AI–Data Complementarity

Our earlier results show that mergers combining high-AI acquirers with data-intensive targets are associated with more favorable announcement returns. We next examine whether these valuation gains are reflected in post-merger innovation outcomes. Innovation is a central driver of firm value, and prior work highlights the role of data assets and analytical capabilities in knowledge creation. Because AI technologies extract patterns and insights from large datasets, integrating AI capability with data-rich targets may enhance innovative activity after merger completion.

A central challenge in assessing post-merger innovation is endogeneity. Firms with strong AI capabilities and data assets may self-select into mergers based on unobserved innovation potential, and observed post-merger changes may reflect pre-existing trends rather than the causal effects of integration. To address these concerns, we adopt a quasi-experimental design that compares completed mergers (treatment group) to withdrawn bids (control group), following prior work (Bena and Li, 2014; Seru, 2014), which treats bid withdrawal as a plausibly exogenous source of variation in merger completion. This approach allows us to difference out common innovation trends around merger an-

nouncements while isolating differential innovation effects associated with deal completion.

We identify withdrawn bids between 2009 and 2019 and match each withdrawn bid to completed mergers based on acquirer–target industry pairs and announcement timing, selecting the closest matches based on relative deal size. The final sample includes 58 withdrawn and 188 completed mergers; sample construction details are provided in Internet Appendix Table IA. Patent application and citation data are drawn from the Extended KPSS database (Kogan et al., 2017). Using three-, four-and five-year windows before and after the announcement year, we estimate a difference-in-differences specification in both a negative binomial framework, which is well-suited for count outcomes such as patent filings and citations, and an OLS framework for patent value which is continuous variable.

The specification includes interactions between post-announcement periods, merger completion, acquirer AI capability, and target data intensity. Coefficients on interactions involving the treatment indicator capture differential post-merger changes for completed deals relative to withdrawn bids, while interactions that exclude the treatment indicator absorb common post-announcement trends across deal types. All specifications include deal and year fixed effects and control for size-weighted acquirer–target characteristics.

— Table 9 about here —

Table 9, 10 and 11 report post-merger innovation results. Table 9 examines *Patent Application*, while Table 10 and Table 11 examine *Patent Citation* and *Patent Value* respectively, capturing both the quantity and impact of innovation. In all the Tables, columns (1)–(2) use three-year pre- and post-announcement windows, and Columns (3)–(6) extend the analysis to four- and five-year windows. Across all specifications, the coefficient on the AI–data interaction term,

$$After_{ij,t} \times Treat_{ij} \times HighAcqAI_i \times HighTarData_j,$$

is positive and statistically significant. This indicates a larger proportional post-merger

increase in innovation for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations.

For Table 9 and 10, because the model is estimated using a negative binomial specification, coefficients can be interpreted as semi-elasticities of expected patent counts. In Column (2) of Table 9, which includes the full set of controls, the estimate on

$$After \times Treat \times HighAcqAI \times HighTarData$$

is 1.38. This coefficient measures the differential post-merger change in patenting for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and net of common post-announcement effects and lower-order interactions. Interpreted in proportional terms, this estimate implies that expected patent applications increase by a factor of $e^{1.38} \approx 3.98$. Thus, these mergers experience a post-merger increase in patenting that is approximately four times larger than that observed for otherwise comparable transactions. Table 10 shows a similar pattern for patent citations, indicating that these gains extend to higher-impact innovation.

— Table 10 about here —

For Table 11, *Patent Value* is specified as natural logarithm of one plus total value creation of patents received by the acquirer and the target. In Column (2) of Table 11, which includes the full set of controls, the coefficient estimate for the variable of interest is 2.055, indicating that mergers between high-AI acquirers and data-rich targets generate approximately $e^{2.055} \approx 7.81$ times greater post-merger patent value than comparable withdrawn deals.

In contrast, lower-order treatment interactions involving acquirer AI capability or target data intensity alone are small or statistically insignificant. This pattern suggests that merger completion is not associated with comparable post-merger innovation gains when AI capability or data assets are present in isolation. Instead, innovation increases are concentrated in transactions that combine strong AI capability on the acquirer side with substantial data assets on the target side.

To validate these findings, we conduct a falsification test using placebo announcement years three years after the actual deals, and re-estimate post-merger *Patent Application* and *Patent Value* using five-year pre- and post-announcement windows. Columns (7)–(8) of Table 9 and 11 show no treatment effects under placebo conditions, confirming that the observed innovation gains are causally linked to mergers.

— Table 11 about here —

4.2 AI-Specific Innovation Following AI–Data Mergers

We next examine whether the documented innovation gains are concentrated in AI-related technologies. Building on Beraja, Yang, and Yuchtman (2023), we focus on patents classified as artificial intelligence using the USPTO Artificial Intelligence Patent Dataset (AIPD) (Pairolero et al., 2025), which identifies patents related to machine learning, natural language processing, computer vision, and AI hardware. To mitigate noise arising from the relative scarcity of AI patents, we restrict the sample to transactions in which both the acquirer and the target held AI patents prior to the merger, ensuring that post-merger changes reflect the expansion of existing AI research programs rather than entry into AI activity.

We replicate the withdrawn-bid research design and matching procedure used in the broader innovation analysis, replacing overall patenting with AI patent applications as the outcome. As before, we estimate a difference-in-differences specification with deal and year fixed effects, using negative binomial and OLS framework for *AI-related Patent Application* and *AI-related Patent Value*, respectively. Consistent with the development cycle of AI technologies, we define the post-merger period to begin in $t+2$ rather than $t+1$ in overall patenting, reflecting the time required to integrate newly acquired data, retrain models, and generate patentable outputs.

— Table 12 about here —

Table 12 reports post-merger AI patent application results. Across all specifica-

tions, the AI–data interaction term,

$$After \times Treat \times HighAcqAI \times HighTarData,$$

is positive and statistically significant, implying a larger proportional increase in AI patent applications for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations.

In Column (2), which includes the full set of controls, the estimated coefficient on the AI–data interaction term is 1.172, corresponding to an incidence rate ratio of $e^{1.172} \approx 3.23$. In Column (4) and Column (6), which extends the event window, the coefficient rises to 1.333 and 1.205, corresponding to an incidence rate ratio of $e^{1.332} \approx 3.79$ and $e^{1.203} \approx 3.33$. The larger effects over longer horizons are consistent with the gradual nature of AI innovation, which requires time for data assimilation, model development, and validation. Table 13 reveals a similar pattern for the AI patent value, confirming that the increase in AI patent filings is accompanied by a rise in innovation value.

To further validate the results of AI-related innovation, we conduct a falsification test by artificially shifting the merger announcement dates three years forward and re-estimating the effect on post-merger *AI Patent Application* and *AI Patent Value* over five-year windows around these placebo events. The results, displayed in Columns (7)–(8) of Table 12 and 13, show no significant treatment effects under the placebo, confirming that the observed increases in AI-related patent filing and value are indeed attributable to the actual mergers.

— Table 13 about here —

Taken together, Tables 9 – 13 show that post-merger innovation gains are concentrated in transactions that combine strong AI capability on the acquirer side with substantial data assets on the target side. Across both overall patenting and AI-specific innovation, merger completion is associated with large proportional increases in innovative activity when AI capability and data assets are jointly integrated within the firm. In contrast, mergers involving AI capability or data intensity in isolation do not exhibit comparable

post-merger innovation responses. These findings suggest that AI–data complementarity is an important mechanism linking the positive market reactions documented earlier to subsequent increases in both the quantity and impact of post-merger innovation.

5 Conclusion

This paper examines how corporate AI capability shapes the value and reallocation of data assets through mergers and acquisitions. We argue that AI and data are complementary forms of intangible capital: the value of data depends on the firm's ability to analyze and deploy it, while AI systems themselves rely on access to large and diverse datasets. If this complementarity is economically meaningful, mergers that combine strong AI capability with substantial data assets should create value.

We provide consistent evidence supporting this view. First, acquisition patterns reflect systematic AI–data alignment: AI-capable firms are significantly more likely to acquire data-intensive targets. Second, transactions that combine high-AI acquirers with data-rich targets generate higher announcement returns and larger combined wealth effects. These gains are not driven by higher acquisition premia, implying greater surplus creation rather than simple redistribution. Indeed, AI-capable acquirers capture a larger share of merger gains, consistent with enhanced bargaining positions associated with complementary capabilities. Deals involving AI capability or data intensity in isolation do not exhibit comparable valuation effects. In addition, the combination of high-data acquirers and high-AI targets does not generate significant announcement returns. Third, using a withdrawn-bid research design, we show that completed mergers combining AI capability and data assets lead to substantial increases in post-merger innovation, both in overall patenting and AI-related patenting.

Taken together, our findings indicate that AI–data complementarity is a central mechanism shaping merger outcomes. The value of data is not fixed but depends on the capabilities of the firm that owns it. By reallocating data assets to AI-capable acquirers, mergers serve as a mechanism for internalizing complementary intangible capital and enhancing innovation.

More broadly, our results suggest that AI is reshaping corporate investment decisions and firm boundaries. As data continues to accumulate and AI technologies advance, the returns to integrating complementary analytical capabilities within a single organization may increase. Understanding how firms combine AI expertise and data assets is therefore central to explaining patterns of capital allocation, value creation, and technological progress in the modern data economy.

References

Abis, S., and L. Veldkamp. 2024. The changing economics of knowledge production. *Review of Financial Studies* 37:89–118.

Aghion, P., and R. Holden. 2011. Incomplete contracts and the theory of the firm: What have we learned over the past 25 years? In R. Gibbons and J. Roberts, eds., *Handbook of Organizational Economics,* 353–402. Princeton University Press.

Agrawal, A., J. Gans, and A. Goldfarb, eds. 2019. *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Andrade, G., M. Mitchell, and E. Stafford. 2001. New evidence and perspectives on mergers. *Journal of Economic Perspectives* 15:103–20.

Archer, S. H., and L. G. Faerber. 1966. Firm size and the cost of externally secured equity capital. *Journal of Finance* 21:69–83.

Babina, T., A. Fedyk, A. He, and J. Hodson. 2024. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics* 151:103745–.

Bartlett, R. P., J. McCrary, and M. O'Hara. 2025. The market inside the market: Odd-lot quotes. *Review of Financial Studies* 38:661–711.

Begenau, J., M. Farboodi, and L. Veldkamp. 2018. Big data in dinance and the growth of large firms. *Journal of Monetary Economics* 97:71–87.

Bena, J., and K. Li. 2014. Corporate innovations and mergers and acquisitions. *Journal of Finance* 69:1923–60.

Beraja, M., D. Yang, and N. Yuchtman. 2023. Data-intensive innovation and the state: Evidence from AI firms in China. *Review of Economic Studies* 90:1701–23.

Brynjolfsson, E., and K. McElheran. 2019. Data in action: Data-driven decision making and predictive analytics in U.S manufacturing. Working paper, Standford University and Rotman School of Management.

Cao, S., W. Jiang, J. Wang, and B. Yang. 2024. From man vs. machine to man+ machine: The art and AI of stock analyses. *Journal of Financial Economics* 160:103910–.

Cockburn, I. M., R. Henderson, and S. Stern. 2019. The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*, 115–46. University of Chicago Press.

Cong, L. W., D. Xie, and L. Zhang. 2021. Knowledge accumulation, privacy, and growth in a data economy. *Management science* 67:6480–92.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou. 2022. The economics of intangible capital. *Journal of Economic Perspectives* 36:29–52.

De Guevara, J. F., J. Maudos, and C. Salvador. 2022. Firms' investment, indebtedness and financial constraints: Size does matter. *Finance Research Letters* 46:102240–.

Dessaint, O., B. E. Eckbo, and A. Golubov. 2025. Bidder-specific synergies and the evolution of acquirer returns. *Management Science* 71:1391–417.

Eisfeldt, A. L., G. Schubert, B. Taska, and M. B. Zhang. 2026. Generative AI and firm values. *Journal of Finance* forthcoming.

Erel, I., L. H. Stern, C. Tan, and M. S. Weisbach. 2021. Selecting directors using machine learning. *Review of Financial Studies* 34:3226–64.

Ewens, M., R. H. Peters, and S. Wang. 2025. Measuring intangible capital with market prices. *Management Science* 71:407–27.

Farboodi, M., and L. Veldkamp. 2021. *A growth model of the data economy*. National Bureau of Economic Research.

Florackis, C., C. Louca, R. Michaely, and M. Weber. 2023. Cybersecurity risk. *Review of Financial Studies* 36:351–407.

Giczy, A. V., N. A. Pairolero, and A. A. Toole. 2022. Identifying artificial intelligence (AI) invention: A novel ai patent dataset. *Journal of Technology Transfer* 47:476–505.

Gokkaya, S., X. Liu, and R. M. Stulz. 2023. Do firms with specialized M&A staff make better acquisitions? *Journal of Financial Economics* 147:75–105.

Goldstein, I., C. S. Spatt, and M. Ye. 2021. Big data in finance. *Review of Financial Studies* 34:3213–25.

———. 2025. The next chapter of big data in finance. *Review of Financial Studies* 38:605–22.

Harford, J. 2005. What drives merger waves? *Journal of Financial Economics* 77:529–60.

Haskel, J., and S. Westlake. 2017. *Capitalism without capital: The rise of the intangible economy*. Princeton University Press.

He, Z., J. Huang, and J. Zhou. 2023. Open banking: Credit market competition when borrowers own the data. *Journal of Financial Economics* 147:449–74.

Healy, P. M., K. G. Palepu, and R. S. Ruback. 1992. Does corporate performance improve after mergers? *Journal of Financial Economics* 31:135–75.

Jones, C. I., and C. Tonetti. 2020. Nonrivalry and the economics of data. *American Economic Review* 110:2819–58.

Kahle, K. M., and R. M. Stulz. 2013. Access to capital, investment, and the financial crisis. *Journal of Financial Economics* 110:280–99.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman. 2017. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics* 132:665–712.

Li, P., F. Li, B. Wang, and Z. Zhang. 2018. Acquiring organizational capital. *Finance Research Letters* 25:30–5.

Loughran, T., and J. Ritter. 2004. Why has IPO underpricing changed over time? *Financial Management* 5–37.

Ma, W., P. Ouimet, and E. Simintzi. 2025. Mergers and acquisitions, technological change, and inequality. *Journal of Financial Economics* 172:104136–.

Masulis, R., C. Wang, and F. Xie. 2007. Corporate governance and acquirer returns. *Journal of Finance* 62:1851–89.

Mihet, R., K. Rishabh, and O. Gomes. 2026. Is it AI or data that drives firm market power? *Journal of Monetary Economics* 157:103878–.

Pairolero, N. A., A. V. Giczy, G. Torres, T. Islam Erana, M. A. Finlayson, and A. A. Toole. 2025. The artificial intelligence patent dataset (AIPD) 2023 update. *Journal of Technology Transfer* 1–24.

Phillips, G. M., and A. Zhdanov. 2013. R&D and the incentives from merger and acquisition activity. *Review of Financial Studies* 26:34–78.

Ramadorai, T., A. Uettwiller, and A. Walther. 2025. Privacy policies and consumer data extraction: evidence from U.S. firms. *Review of Finance* 29:1337–67.

Rhodes-Kropf, M., and D. T. Robinson. 2008. The market for mergers and the boundaries of the firm. *Journal of Finance* 63:1169–211.

Rhodes-Kropf, M., and S. Viswanathan. 2004. Market valuation and merger waves. *Journal of Finance* 59:2685–718.

Seru, A. 2014. Firm boundaries matter: Evidence from conglomerates and R&D activity. *Journal of Financial Economics* 111:381–405.

Sheng, J., Z. Sun, B. Yang, and A. L. Zhang. 2024. Generative AI and asset management. *Review of Financial Studies* forthcoming.

Shleifer, A., and R. W. Vishny. 2003. Stock market driven acquisitions. *Journal of Financial Economics* 70:295–311.

Veldkamp, L., and C. Chung. 2024. Data and the aggregate ecomomy. *Journal of Economic Literature* 62:458–84.

Zhu, C. 2019. Big data as a governance mechanism. *Review of Financial Studies* 32:2021–61.

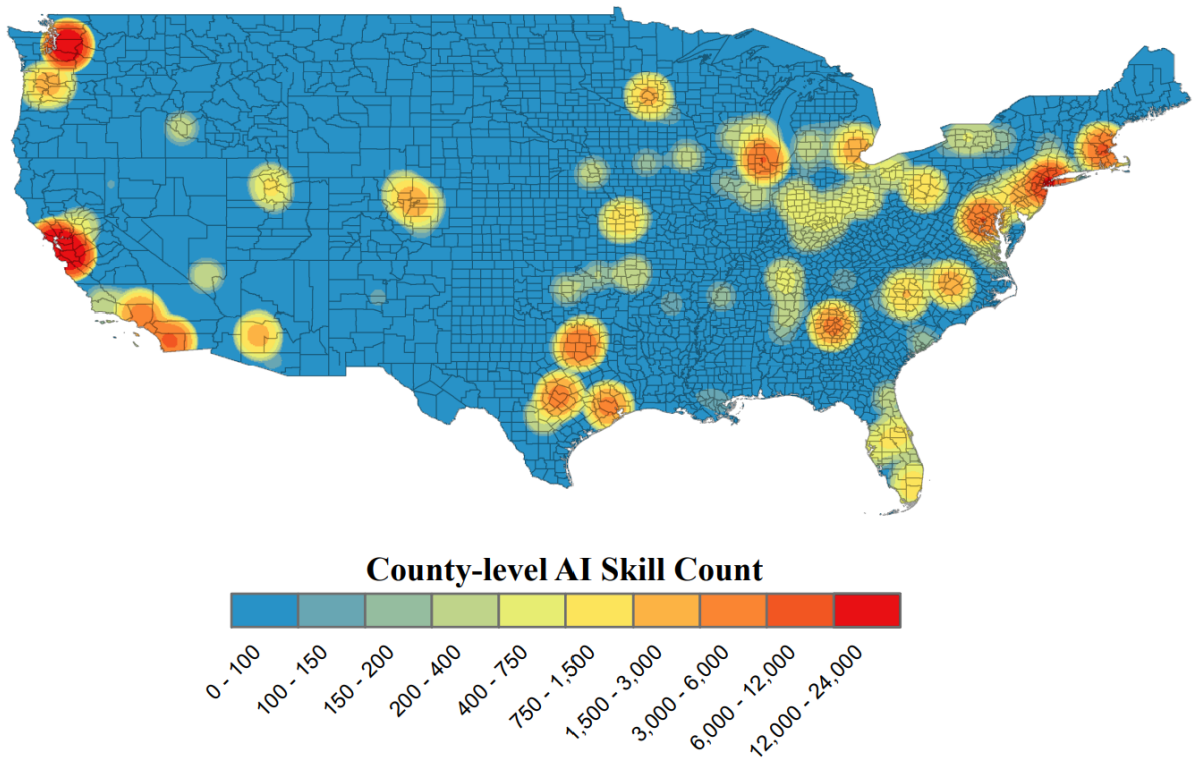


Figure 1: Geographic Distribution of AI Skills
The county-level AI skill count measures the total number of AI-related skills held by employees in our sample, aggregated at the county level. Skills are identified from LinkedIn profiles based on employees' primary reported work locations. Warmer colors on the map indicate higher concentrations of AI talent.

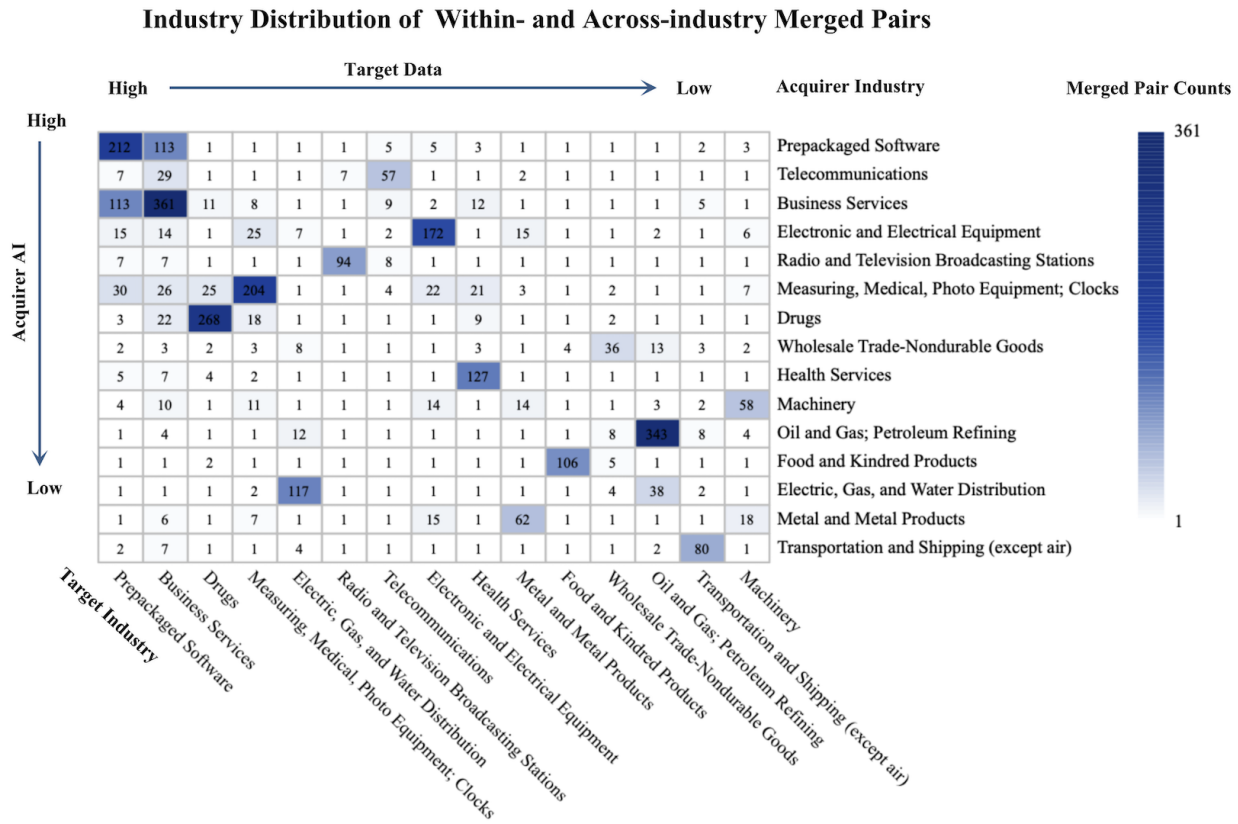


Figure 2: Industry Distribution of Within- and Across-industry Merged Pairs
This figure shows the industry distribution of all merged pairs, including both within- and cross-industry transactions. We select the 15 most frequent acquirer industries and the 15 most frequent target industries, ranking acquirers by their AI capability and targets by data intensity. The vertical axis lists acquirer industries in descending order of AI capability, while the horizontal axis displays target industries ordered by data intensity. Each cell reports the number of observed deals for a given acquirer-target industry pair, capturing merger activity across industry combinations.

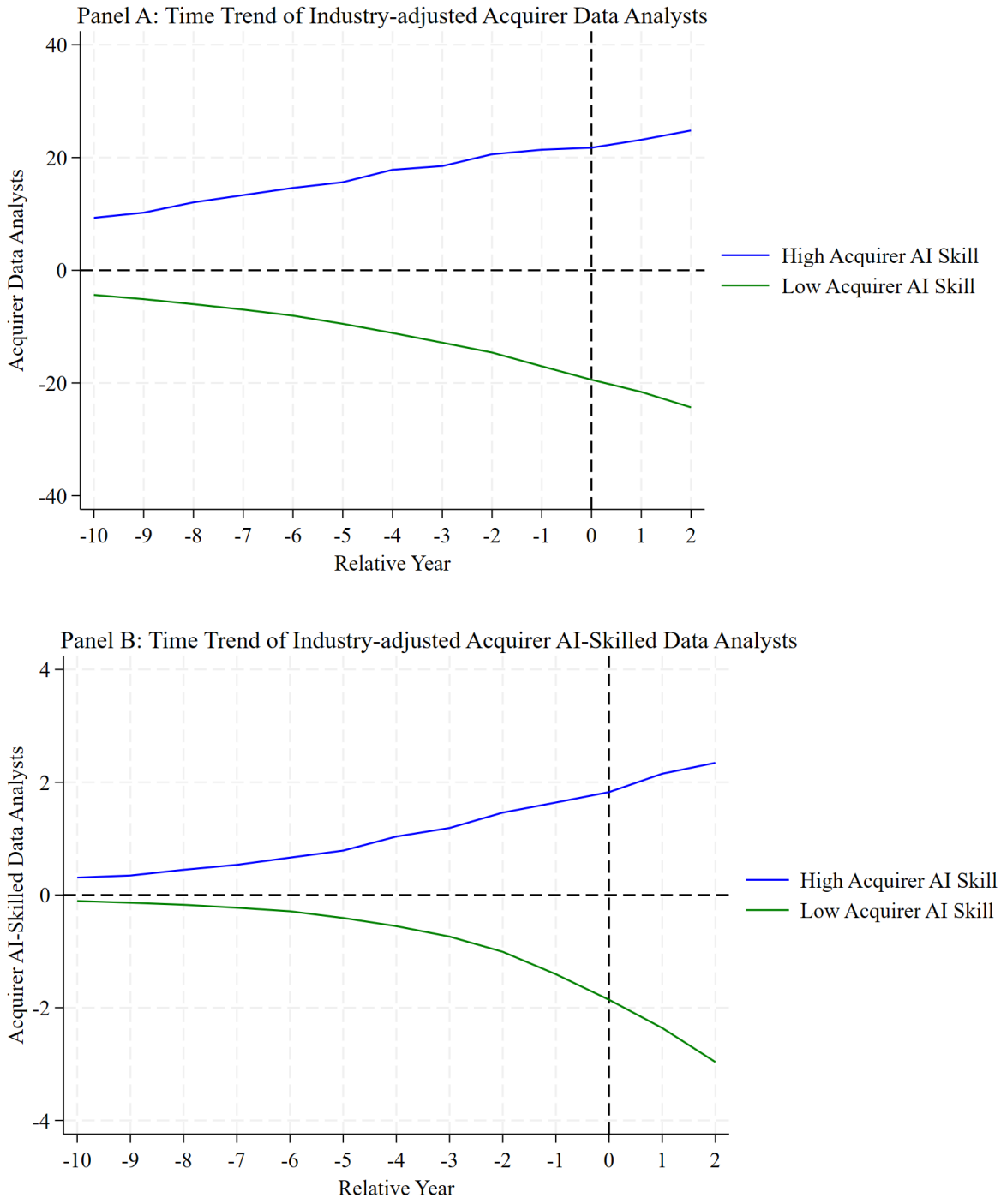


Figure 3: Time Series of Data Analytics Employees
This figure plots the industry-adjusted average number of data analysts in high-AI and low-AI acquirers from 10 years before to 2 years after the M&A announcement year. Acquirers are classified as high or low AI based on the median AI skill level derived from LinkedIn data in the year preceding the transaction. Data analysts are identified job titles containing the keyword "data." The blue line represents high-AI acquirers, and the the green line represents low-AI acquirers; both series are industry-adjusted. Panel A shows trends for all data analysts, while Panel B focuses on data analysts who also report AI-related skills in their LinkedIn profiles.

Table 1:
Summary Statistics

This table reports descriptive statistics for M&A transactions from 2009 to 2019. Panels A–D present the mean, median, standard deviation, minimum, and maximum for acquirer, target, deal, and combined firm characteristics, respectively. The sample is drawn from the Thomson One Platinum SDC M&A database and includes filtered deals with public acquirers worldwide and U.S. public or private targets announced between January 1, 2009, and December 31, 2019. Appendix A provides detailed variable definitions.

	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Panel A: Acquirer Characteristics:					
Acq_Emp	4,471	1,454.88	4,345.06	1.00	80,220.67
Acq_AIEmp	4,471	22.98	148.36	0.00	4,576.92
Acq_AISkill	4,471	31.51	222.20	0.00	7,412.58
LnAcqAIEmp	4,471	1.28	1.49	0.00	7.39
LnAcqAISkill	4,471	1.38	1.59	0.00	7.73
LnAcqEmp	4,471	5.37	2.28	0.00	10.72
Leverage	4,471	0.26	0.22	0.00	0.92
Intanratio	4,471	0.28	0.23	0.00	0.83
RD	4,471	0.03	0.06	0.00	0.29
ROA	4,471	0.02	0.14	-0.68	0.24
Size	4,471	7.24	1.90	2.69	12.76
Tobin's Q	4,471	2.00	1.18	0.75	7.91
Acq_DataIntensity	4,131	318.51	2,084.90	0.00	16,086.00
Panel B: Target Characteristics:					
Tar_Emp	4,471	292.84	1,745.77	0.00	84,096.83
Tar_AIEmp	4,471	3.64	25.93	0.00	793.75
Tar_AISkill	4,471	4.85	36.77	0.00	1,213.08
LnTarAISkill	4,471	0.48	0.96	0.00	4.53
TargetStatus	4,471	0.15	0.36	0.00	1.00
DataIntensity	1,000	649.53	5932.57	0.00	128601.00
DataIntensity_GPT	996	8.06	3.93	0.00	18.00
Cybersecurity_Risk	934	0.26	0.21	0.00	0.93
Panel C: Deal Characteristics:					
AllStock	4,471	0.04	0.20	0.00	1.00
Challenge	4,471	0.01	0.12	0.00	1.00

Table 1 Continued

	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Conglomerate	4,471	0.39	0.49	0.00	1.00
Crossborder	4,471	0.05	0.21	0.00	1.00
Hightechdeal	4,471	0.21	0.40	0.00	1.00
Rel_DealVal	4,471	0.25	0.42	0.00	2.72
CAR[-3,+3]	4,471	0.01	0.08	-0.21	0.25
CAR[-4,+4]	4,471	0.01	0.09	-0.23	0.26
CAR[-5,+5]	4,471	0.01	0.09	-0.25	0.29
TarCAR[-2,+2]	1,048	0.14	0.18	-0.18	0.74
TarCAR[-3,+3]	1,048	0.14	0.18	-0.21	0.75
TarCAR[-41,+3]	587	0.27	0.26	-0.32	1.32
CombinedCAR	587	0.03	0.09	-0.22	0.29
Premium	587	0.04	0.05	-0.11	0.20
Acquirer's Share of Synergy	587	-0.00	0.06	-0.19	0.16
Panel D: Combined Firm Characteristics:					
Patent Application	2,124	99.89	289.94	0.00	1,961.00
Patent Citation	2,124	412.56	1,263.73	0.00	8,079.00
Patent Value	2,124	3.59	3.50	0.00	9.89
AI Patent Application	1,328	34.55	92.18	0.00	534.00
AI Patent Value	1,328	1.09	1.42	0.00	5.04

Table 2:
Likelihood of Acquirer-Target Firm Pairing

This table reports coefficient estimates from conditional logit models using observed acquirer-target pairs and a matched control sample of potential deals. The dependent variable equals one for the observed acquirer-target pairs and zero for control pairs. Columns (1) to (3) report results with all the acquirer/target controls, while Columns (4) to (6) exclude the matching variables used in each respective matching procedure. Variable definitions are provided in Appendix Appendix A. All specifications include deal fixed effects. Robust z-statistics adjusted for deal-level clustering are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	All Controls			Without Matching Variables		
	(1) Ind-Size	(2) Ind-Size-ROA	(3) Ind-Size-ROA-Q	(4) Ind-Size	(5) Ind-Size-ROA	(6) Ind-Size-ROA-Q
HighAcqAI*HighTarData	0.629**	0.723**	0.623**	0.596**	0.779***	0.686**
	(2.14)	(2.40)	(2.09)	(2.09)	(2.66)	(2.28)
HighAcqAI	-1.767***	-1.863***	-1.636***	-1.867***	-1.828***	-1.604***
	(-6.70)	(-6.74)	(-6.00)	(-6.32)	(-6.08)	(-5.25)
HighTarData	-0.198	-0.171	-0.136	-0.190	-0.114	-0.102
	(-1.12)	(-0.85)	(-0.65)	(-1.27)	(-0.68)	(-0.59)
SalesGrowth	0.372***	0.314***	0.334**	0.098	0.118*	0.047
	(2.83)	(3.32)	(2.50)	(0.91)	(1.78)	(0.47)
Tar_SalesGrowth	-0.070	-0.047	-0.118	-0.166**	-0.050	-0.121
	(-0.67)	(-0.55)	(-0.90)	(-1.96)	(-0.74)	(-1.40)
Leverage	0.356	-0.184	0.079	1.068***	0.203	0.227
	(1.23)	(-0.69)	(0.31)	(5.29)	(1.15)	(1.44)
Tar_Leverage	-0.079	-0.355	-0.635**	0.712***	0.057	0.102
	(-0.27)	(-1.11)	(-2.19)	(3.69)	(0.29)	(0.59)
RD	1.006	1.465	0.964	-0.790	-4.413***	-5.294***
	(1.02)	(1.54)	(1.09)	(-1.02)	(-7.11)	(-8.30)
Tar_RD	0.833	1.662**	1.375**	-0.170	-1.943***	-2.740***
	(1.17)	(2.20)	(2.28)	(-0.34)	(-4.05)	(-6.27)

Table 2 Continued

	All Controls			Without Matching Variables		
	(1) Ind-Size	(2) Ind-Size-ROA	(3) Ind-Size-ROA-Q	(4) Ind-Size	(5) Ind-Size-ROA	(6) Ind-Size-ROA-Q
Conglomerate	-1.890***	-1.888***	-2.021***	-2.036***	-2.013***	-2.176***
	(-7.51)	(-7.55)	(-8.35)	(-9.05)	(-9.21)	(-9.73)
SameState	-1.648***	-1.808***	-1.976***	-1.245***	-1.488***	-1.443***
	(-7.35)	(-7.77)	(-8.31)	(-6.91)	(-7.56)	(-7.40)
Tobin's Q	-0.053	-0.056	-0.081**	-0.114***	-0.076***	
	(-1.01)	(-1.48)	(-2.20)	(-3.26)	(-3.11)	
Tar_Tobin's Q	-0.060	-0.095**	-0.161***	-0.101***	-0.105***	
	(-1.32)	(-2.08)	(-3.11)	(-3.46)	(-3.31)	
ROA	1.035**	1.525***	1.489***	2.282***		
	(1.99)	(2.60)	(3.10)	(5.02)		
Tar_ROA	-0.101	0.580*	0.455*	0.977***		
	(-0.27)	(1.82)	(1.85)	(3.71)		
Size	1.363***	0.783***	0.656***			
	(11.24)	(14.75)	(14.93)			
Tar_Size	1.400***	0.886***	0.748***			
	(10.41)	(11.85)	(11.96)			
Observations	4,415	4,167	4,051	4,415	4,167	4,051
Pseudo R^2	0.407	0.362	0.360	0.150	0.173	0.175
No. of Actual Deals	533	515	508	533	515	508
No. of Control Deals	3,882	3,652	3,543	3,882	3,652	3,543
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 3:
Acquirer AI Capabilities and Acquisition Announcement Returns

This table reports results from announcement return regressions. The dependent variables are cumulative abnormal returns (CARs) over the [-3, +3], [-4,+4], and [-5,+5] windows, estimated using the market model with parameters derived from the CRSP value-weighted index over [-240, -11] trading days estimation window. Columns (1)-(3) examine the effect of acquirer AI skill intensity on CARs, while Columns (4)-(6) examine the effect of the number of AI-skilled employees. All specifications include acquirer industry and state × year fixed effects. Robust t-statistics, adjusted for firm-level clustering, are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) CAR[-3,+3]	(2) CAR[-4,+4]	(3) CAR[-5,+5]	(4) CAR[-3,+3]	(5) CAR[-4,+4]	(6) CAR[-5,+5]
LnAcqAISkill	0.005***	0.006***	0.006***			
	(2.77)	(2.91)	(2.73)			
LnAcqAIEmp				0.005**	0.005***	0.005**
				(2.57)	(2.70)	(2.53)
LnAcqEmp	-0.003***	-0.003***	-0.003**	-0.003**	-0.003**	-0.003**
	(-2.68)	(-2.65)	(-2.42)	(-2.52)	(-2.47)	(-2.26)
Size	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
	(-6.61)	(-6.07)	(-6.03)	(-6.54)	(-6.00)	(-5.96)
Tobin's Q	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(-1.47)	(-1.42)	(-1.34)	(-1.45)	(-1.40)	(-1.32)
Leverage	0.026***	0.021**	0.019*	0.026***	0.021**	0.019*
	(2.97)	(2.23)	(1.79)	(2.97)	(2.23)	(1.79)
ROA	0.008	0.003	0.014	0.008	0.003	0.014
	(0.52)	(0.16)	(0.73)	(0.51)	(0.16)	(0.73)
Intanratio	0.002	0.002	-0.002	0.002	0.002	-0.002
	(0.27)	(0.26)	(-0.22)	(0.26)	(0.25)	(-0.23)

Table 3 Continued

	(1) CAR[-3,+3]	(2) CAR[-4,+4]	(3) CAR[-5,+5]	(4) CAR[-3,+3]	(5) CAR[-4,+4]	(6) CAR[-5,+5]
RD	-0.084**	-0.080*	-0.061	-0.083**	-0.079*	-0.060
	(-2.01)	(-1.75)	(-1.29)	(-1.99)	(-1.73)	(-1.27)
TargetStatus	-0.016***	-0.015***	-0.017***	-0.016***	-0.015***	-0.017***
	(-3.35)	(-2.95)	(-3.12)	(-3.35)	(-2.95)	(-3.12)
AllStock	-0.015*	-0.009	-0.012	-0.015*	-0.009	-0.012
	(-1.73)	(-0.99)	(-1.21)	(-1.72)	(-0.98)	(-1.21)
Conglomerate	-0.002	-0.005	-0.008*	-0.002	-0.005	-0.007*
	(-0.73)	(-1.29)	(-1.96)	(-0.72)	(-1.29)	(-1.95)
Rel_DealVal	0.017***	0.015***	0.018***	0.017***	0.015***	0.018***
	(3.40)	(2.85)	(3.03)	(3.39)	(2.84)	(3.03)
Hightechdeal	-0.002	-0.004	-0.007	-0.001	-0.004	-0.007
	(-0.28)	(-0.67)	(-1.09)	(-0.27)	(-0.66)	(-1.09)
Challenge	0.010	0.003	0.002	0.010	0.003	0.002
	(0.94)	(0.21)	(0.17)	(0.95)	(0.22)	(0.18)
Crossborder	-0.011	0.003	0.016	-0.011	0.003	0.016
	(-0.58)	(0.15)	(0.66)	(-0.58)	(0.16)	(0.67)
Observations	4,325	4,325	4,325	4,325	4,325	4,325
R^2	0.218	0.210	0.211	0.217	0.210	0.211
Acq. ind. FE	Yes	Yes	Yes	Yes	Yes	Yes
Acq. State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 4:
Acquirer AI Capabilities and Acquisition Announcement Returns: IV Estimates

This table reports the result from a 2SLS approach adopting an instrumental variable. Columns (1) and (2) report the first stage regression, where *LnAcqAISkill* is regressed on the instrument that measures acquirer firm-level exposure to the local supply of AI talents. Columns (3)-(5) report the second stage results corresponding to the first-stage regression without fixed effects and Columns (6)-(8) report the second stage results corresponding to the first-stage regression with fixed effects. All the Columns include all the baseline control variables. Standard errors are clustered at the firm level. Robust t-statistics are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First stage		Second stage					
	LnAcqAISkill		CAR [-3,+3]	CAR [-4,+4]	CAR [-5,+5]	CAR [-3,+3]	CAR [-4,+4]	CAR [-5,+5]
IV	78.887***	53.654***						
	(7.49)	(4.79)						
LnAcqAISkill			0.017*	0.019**	0.020**	0.032*	0.034*	0.035*
			(1.88)	(2.09)	(2.11)	(1.95)	(1.96)	(1.85)
LnAcqEmp	0.438***	0.440***	-0.007*	-0.008**	-0.008**	-0.015**	-0.015**	-0.016*
	(29.45)	(30.24)	(-1.83)	(-2.04)	(-1.99)	(-2.02)	(-2.01)	(-1.89)
Size	0.148***	0.168***	-0.009***	-0.009***	-0.010***	-0.012***	-0.013***	-0.013***
	(8.68)	(9.37)	(-4.87)	(-4.84)	(-5.04)	(-3.73)	(-3.68)	(-3.65)
Tobin's Q	0.042*	0.018	-0.003*	-0.004**	-0.003*	-0.003	-0.003	-0.003
	(1.87)	(0.83)	(-1.89)	(-2.01)	(-1.79)	(-1.63)	(-1.49)	(-1.39)
Leverage	-0.242*	-0.062	0.017**	0.015*	0.016*	0.024**	0.017*	0.019*
	(-1.89)	(-0.53)	(2.26)	(1.92)	(1.76)	(2.45)	(1.66)	(1.69)
ROA	-0.248*	-0.262*	0.008	0.006	0.017	0.019	0.015	0.027
	(-1.85)	(-1.89)	(0.60)	(0.36)	(0.94)	(1.25)	(0.79)	(1.37)
Intanratio	0.086	-0.240*	0.014**	0.015**	0.014**	0.019*	0.023**	0.019
	(0.73)	(-1.84)	(2.31)	(2.33)	(2.07)	(1.77)	(2.02)	(1.55)
RD	2.674***	1.752***	-0.045	-0.045	-0.034	-0.110**	-0.112**	-0.092
	(6.52)	(3.86)	(-1.10)	(-1.00)	(-0.74)	(-2.18)	(-2.00)	(-1.55)

Table 4 Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First stage		Second stage					
	LnAcqAISkill		CAR [-3,+3]	CAR [-4,+4]	CAR [-5,+5]	CAR [-3,+3]	CAR [-4,+4]	CAR [-5,+5]
TargetStatus	-0.001	0.005	-0.017***	-0.017***	-0.018***	-0.015***	-0.015***	-0.016***
	(-0.03)	(0.12)	(-3.83)	(-3.64)	(-3.75)	(-3.14)	(-2.89)	(-2.85)
AllStock	0.124*	0.092	-0.017**	-0.013	-0.016	-0.018*	-0.012	-0.016
	(1.65)	(1.25)	(-1.98)	(-1.40)	(-1.63)	(-1.90)	(-1.13)	(-1.45)
Conglomerate	0.070*	0.021	-0.004	-0.005*	-0.007**	-0.005	-0.007*	-0.009**
	(1.82)	(0.61)	(-1.60)	(-1.72)	(-2.18)	(-1.35)	(-1.80)	(-2.33)
Rel_DealVal	-0.082*	-0.040	0.022***	0.021***	0.023***	0.020***	0.019***	0.020***
	(-1.88)	(-1.05)	(4.21)	(3.80)	(3.98)	(3.61)	(3.16)	(3.17)
Hightechdeal	0.714***	0.156**	-0.022***	-0.025***	-0.028***	-0.009	-0.013**	-0.016**
	(9.67)	(2.23)	(-2.94)	(-3.24)	(-3.38)	(-1.46)	(-1.97)	(-2.21)
Challenge	0.139	0.015	0.008	0.004	-0.002	0.010	0.003	0.000
	(1.01)	(0.13)	(0.71)	(0.32)	(-0.15)	(0.90)	(0.29)	(0.03)
Crossborder	0.327**	0.381**	-0.007	-0.005	-0.000	-0.022	-0.010	0.004
	(2.36)	(2.14)	(-0.93)	(-0.67)	(-0.02)	(-1.17)	(-0.46)	(0.16)
Observations	4471	4362	4471	4471	4471	4362	4362	4362
R-squared	0.664	0.800	0.008	-0.001	0.002	-0.030	-0.033	-0.028
F Statistic	23.84	41.66	23.84	23.84	23.84	41.66	41.66	41.66
Acq. ind. FE	No	Yes	No	No	No	Yes	Yes	Yes
Acq. State. × Year FE	No	Yes	No	No	No	Yes	Yes	Yes

Table 5:
Evaluation of Combined Wealth Effects and Acquisition Premium

This table reports regressions of *CombinedCAR, Premium,* and the bidder's share of combined synergy gains for deals involving public targets. In Column (1), the dependent variable is *CombinedCAR*, defined as the weighted average of the target's CAR [-41, +3] and the acquirer's CAR [-3, +3], using market capitalizations measured on day -42 for the target and day -4 for the acquirer as weights. Column (2) uses *Premium*, calculated as the target's CAR [-41, +3] multiplied by its market capitalization on day -42, and scaled by the combined market capitalizations of the target (day -42) and the acquirer (day -4). Column (3) examines the *Acquirer's Share of Synergy*,defined as *CombinedCAR* minus *Premium*. Column (4) and (5) report regressions of target CARs. All specifications include acquirer industry fixed effects and acquirer state $\times$ year fixed effects. Robust t-statistics clustered at the firm-level are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) Combined CAR	(2) Premium	(3) Acquirer's Share of Synergy	(4) Target CAR [-2,+2]	(5) Target CAR [-3,+3]
LnAcqAISkill	0.018**	0.004	0.013*	0.046***	0.049***
	(2.20)	(1.00)	(1.96)	(2.76)	(2.90)
LnAcqEmp	-0.012**	-0.003	-0.007*	-0.025**	-0.028**
	(-2.22)	(-1.22)	(-1.76)	(-2.19)	(-2.42)
Size	-0.006	0.001	-0.008*	-0.031**	-0.028**
	(-1.14)	(0.21)	(-1.79)	(-2.48)	(-2.11)
Tobin's Q	-0.011	-0.006	-0.004	0.034**	0.030*
	(-1.44)	(-1.43)	(-0.59)	(2.25)	(1.74)
Leverage	0.020	0.013	0.007	-0.127	-0.129
	(0.39)	(0.48)	(0.17)	(-1.47)	(-1.48)
ROA	0.070	0.027	0.040	0.155	0.124
	(1.07)	(0.73)	(0.78)	(0.99)	(0.79)
Intanratio	-0.022	-0.004	-0.019	-0.005	-0.018
	(-0.52)	(-0.19)	(-0.66)	(-0.07)	(-0.25)

Table 5 Continued

	(1) Combined CAR	(2) Premium	(3) Acquirer's Share of Synergy	(4) Target CAR [-2,+2]	(5) Target CAR [-3,+3]
RD	-0.133	-0.043	-0.132	-0.261	-0.165
	(-0.81)	(-0.48)	(-1.03)	(-0.91)	(-0.53)
AllStock	0.000	-0.013	0.012	-0.113***	-0.111***
	(0.02)	(-1.09)	(0.88)	(-3.28)	(-3.34)
Conglomerate	-0.004	-0.007	0.004	0.021	0.006
	(-0.29)	(-0.99)	(0.37)	(0.73)	(0.20)
Rel_DealVal	0.022*	0.026***	-0.003	-0.080***	-0.087***
	(1.81)	(3.26)	(-0.33)	(-3.50)	(-3.76)
Hightechdeal	-0.006	0.001	-0.002	0.067	0.061
	(-0.23)	(0.05)	(-0.11)	(1.03)	(0.89)
Challenge	0.039*	0.024*	0.014	-0.007	0.005
	(1.83)	(1.76)	(0.91)	(-0.13)	(0.10)
Crossborder	0.055	0.063	-0.015	-0.361***	-0.258**
	(0.76)	(1.24)	(-0.31)	(-2.88)	(-2.25)
Observations	363	363	363	363	363
R^2	0.539	0.563	0.520	0.616	0.603
Acq. ind. FE	Yes	Yes	Yes	Yes	Yes
Acq. State × Year FE	Yes	Yes	Yes	Yes	Yes

Table 6:
Testing the Synergy between Acquirer AI and Target Data

This table reports the effects of different acquirer-target combinations on acquirer announcement returns. In panel A, deals are classified into eight groups based on the acquirer's AI capability, the target's AI capability, and the target's data intensity. Due to collinearity among these indicators, three groups are omitted as reference categories, and the table reports results for the remaining five groups. In panel B, deals are classified into 12 groups based on the acquirer's AI capability and data intensity, as well as the target's AI capability and data intensity, which are high/low acquirer AI + high/low target data (four groups), high/low acquirer data + high/low target AI (four groups) and high/low acquirer AI + high/low target AI (four groups). Due to collinearity among these indicators, five groups are omitted as reference categories, and the table reports results for the remaining seven groups. This table also shows t values of pairwise tests on HighAcqAI × (HighTarData - LowTarData) and (HighAcqAI - LowAcqAI) × HighTarData. All specifications include acquirer industry × year fixed effects and acquirer state × year fixed effects. Columns (4)–(6) further include target industry × year fixed effects. Robust t-statistics adjusted for firm-level clustering are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) CAR [-3,+3]	(2) CAR [-4,+4]	(3) CAR [-5,+5]	(4) CAR [-3,+3]	(5) CAR [-4,+4]	(6) CAR [-5,+5]
Panel A: Acquirer AI - Target Data combinations						
HighAcqAI × HighTarData	0.054	0.067**	0.084**	0.122**	0.138**	0.154***
	(1.65)	(2.05)	(2.53)	(2.50)	(2.52)	(2.73)
HighAcqAI × LowTarData	0.003	0.016	0.025	0.023	0.041	0.055
	(0.12)	(0.67)	(1.06)	(0.52)	(0.91)	(1.12)
HighAcqAI × HighTarAI	0.016	-0.003	-0.011	0.045	0.024	0.002
	(0.67)	(-0.12)	(-0.55)	(1.16)	(0.64)	(0.04)
LowAcqAI × HighTarData	-0.021	-0.022	-0.014	-0.026	-0.025	-0.013
	(-0.69)	(-0.75)	(-0.51)	(-0.66)	(-0.51)	(-0.29)
LowAcqAI × HighTarAI	-0.045*	-0.047*	-0.045	-0.006	-0.017	-0.015
	(-1.72)	(-1.70)	(-1.51)	(-0.21)	(-0.47)	(-0.36)
LnAcqEmp	-0.013**	-0.012**	-0.011**	-0.023***	-0.022***	-0.020**
	(-2.20)	(-2.36)	(-2.14)	(-3.07)	(-3.04)	(-2.58)

Table 6 Continued

	(1) CAR [-3,+3]	(2) CAR [-4,+4]	(3) CAR [-5,+5]	(4) CAR [-3,+3]	(5) CAR [-4,+4]	(6) CAR [-5,+5]
Size	-0.003	-0.002	-0.004	0.009	0.006	0.004
	(-0.54)	(-0.38)	(-0.65)	(1.01)	(0.67)	(0.36)
Tobin's Q	-0.010	-0.006	-0.010	-0.043**	-0.037*	-0.034*
	(-1.13)	(-0.75)	(-1.10)	(-2.46)	(-1.93)	(-1.75)
Leverage	-0.056	-0.064	-0.069	-0.031	-0.020	-0.057
	(-1.21)	(-1.31)	(-1.37)	(-0.28)	(-0.16)	(-0.43)
ROA	-0.084	-0.123	-0.124	-0.109	-0.142	-0.148
	(-1.15)	(-1.55)	(-1.54)	(-1.34)	(-1.48)	(-1.38)
Intanratio	-0.041	-0.036	-0.046	-0.021	-0.017	-0.032
	(-0.93)	(-0.84)	(-1.00)	(-0.30)	(-0.23)	(-0.44)
RD	-0.280	-0.289*	-0.348**	-0.316	-0.247	-0.303
	(-1.65)	(-1.86)	(-2.34)	(-0.94)	(-0.72)	(-0.90)
AllStock	0.006	0.002	-0.003	0.016	0.014	0.019
	(0.19)	(0.07)	(-0.09)	(0.38)	(0.36)	(0.44)
Conglomerate	0.000	-0.003	-0.003	-0.022	-0.022	-0.015
	(0.00)	(-0.19)	(-0.21)	(-0.74)	(-0.69)	(-0.42)
Rel_DealVal	-0.011	-0.007	-0.007	-0.015	-0.017	-0.011
	(-0.88)	(-0.61)	(-0.50)	(-0.93)	(-0.90)	(-0.50)
Hightechdeal	-0.006	-0.021	-0.020	-0.060	-0.097	-0.106
	(-0.17)	(-0.59)	(-0.61)	(-0.93)	(-1.13)	(-1.27)
Challenge	0.029	0.019	0.005	0.042	0.038	0.029
	(1.06)	(0.70)	(0.18)	(1.09)	(0.97)	(0.64)
Crossborder	0.101	0.118	0.072			
	(1.24)	(1.35)	(0.75)			

Table 6 Continued

	(1) CAR [-3,+3]	(2) CAR [-4,+4]	(3) CAR [-5,+5]	(4) CAR [-3,+3]	(5) CAR [-4,+4]	(6) CAR [-5,+5]
Tar_Size	0.007	0.007	0.008*	0.016	0.017*	0.018*
	(1.46)	(1.61)	(1.80)	(1.59)	(1.68)	(1.67)
Tar_Tobinsq	-0.000	-0.003	-0.003	-0.003	-0.003	-0.009
	(-0.06)	(-0.44)	(-0.42)	(-0.20)	(-0.18)	(-0.59)
Tar_Leverage	-0.027	-0.023	-0.025	-0.094	-0.082	-0.042
	(-0.55)	(-0.50)	(-0.59)	(-1.36)	(-1.33)	(-0.67)
Tar_ROA	0.026	0.044	0.080	-0.141	-0.102	-0.043
	(0.38)	(0.73)	(1.41)	(-1.56)	(-1.02)	(-0.42)
Tar_Intanratio	-0.030	-0.046	-0.029	-0.111	-0.116	-0.121
	(-0.65)	(-1.03)	(-0.68)	(-1.60)	(-1.67)	(-1.54)
Tar_RD	0.014	0.030	0.047	-0.174	-0.169	-0.150
	(0.10)	(0.22)	(0.35)	(-0.81)	(-0.68)	(-0.61)
Constant	0.111**	0.098**	0.106**	0.081	0.083	0.094
	(2.18)	(2.04)	(2.02)	(0.90)	(0.90)	(0.91)
HighAcqAI × (HighTarData - LowTarData)	0.051**	0.051**	0.059**	0.099**	0.096**	0.099**
	(2.09)	(2.00)	(2.27)	(2.24)	(2.09)	(2.11)
(HighAcqAI - LowAcqAI) × HighTarData	0.076*	0.089**	0.099**	0.147***	0.163***	0.167***
	(1.66)	(2.07)	(2.32)	(2.93)	(3.09)	(3.17)
Observations	373	373	373	186	186	186
R^2	0.654	0.693	0.698	0.761	0.779	0.772
Acq. ind. × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Acq. State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Tar. ind. × Year FE	No	No	No	Yes	Yes	Yes

Table 6 Continued

	(1) CAR [-3,+3]	(2) CAR [-4,+4]	(3) CAR [-5,+5]	(4) CAR [-3,+3]	(5) CAR [-4,+4]	(6) CAR [-5,+5]
Panel B: Add Acquirer Data - Target AI combinations						
HighAcqAI × HighTarData	0.073**	0.077**	0.090***	0.138***	0.145**	0.168***
	(2.10)	(2.33)	(2.67)	(2.72)	(2.44)	(2.72)
HighAcqAI × LowTarData	0.025	0.031	0.038	0.048	0.061	0.085*
	(0.95)	(1.19)	(1.45)	(1.22)	(1.34)	(1.78)
LowAcqAI × HighTarData	-0.005	-0.008	-0.004	0.021	0.025	0.038
	(-0.17)	(-0.26)	(-0.15)	(0.41)	(0.36)	(0.60)
HighAcqData × HighTarAI	-0.035	-0.034	-0.048	-0.049	-0.044	-0.046
	(-1.20)	(-1.11)	(-1.40)	(-0.78)	(-0.68)	(-0.63)
LowAcqData × HighTarAI	-0.047*	-0.057**	-0.057*	-0.002	-0.012	-0.009
	(-1.81)	(-2.09)	(-1.83)	(-0.07)	(-0.34)	(-0.20)
HighAcqData × LowTarAI	0.038	0.042	0.029	-0.034	0.018	0.040
	(1.16)	(1.29)	(0.87)	(-0.59)	(0.25)	(0.50)
HighAcqAI × HighTarAI	0.072**	0.063*	0.056	0.035	0.039	0.020
	(2.27)	(1.96)	(1.62)	(0.61)	(0.66)	(0.32)
Constant	0.088	0.072	0.087	0.066	0.064	0.086
	(1.47)	(1.27)	(1.45)	(0.59)	(0.56)	(0.71)
Acq. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Deal Controls	Yes	Yes	Yes	Yes	Yes	Yes
Tar. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	315	315	315	149	149	149
R^2	0.663	0.713	0.710	0.792	0.811	0.799
Acq. ind. × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Acq. State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Tar. ind. × Year FE	No	No	No	Yes	Yes	Yes

Table 7:
Alternative Data Measure: Likelihood of Acquirer-Target Firm Pairing

This table provides a robustness check of Table 2 by replacing the target data measure with a GPT-generated measure (Columns (1) to (3)) and cybersecurity risk measure (Columns (4) to (6)). This table reports coefficient estimates from conditional logit models using observed acquirer-target pairs and a matched control sample of potential deals. The dependent variable equals one for the observed acquirer-target pairs and zero for control pairs. Variable definitions are provided in Appendix Appendix A. All specifications include deal fixed effects. Robust z-statistics adjusted for deal-level clustering are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	GPT Measure			Cybersecurity Risk		
	(1) Ind-Size	(2) Ind-Size-ROA	(3) Ind-Size-ROA-Q	(4) Ind-Size	(5) Ind-Size-ROA	(6) Ind-Size-ROA-Q
HighAcqAI × HighTarData	0.930***	1.284***	0.781**	0.156	0.066	0.150
	(2.96)	(3.95)	(2.43)	(0.44)	(0.19)	(0.42)
HighAcqAI	-2.054***	-2.375***	-1.734***	-1.671***	-1.890***	-1.626***
	(-6.90)	(-7.92)	(-5.81)	(-5.24)	(-5.68)	(-4.78)
HighTarData	-0.437**	-0.680***	-0.448**	0.352*	0.218	0.156
	(-2.18)	(-3.29)	(-2.09)	(1.74)	(1.08)	(0.73)
Acq. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Deal Controls	Yes	Yes	Yes	Yes	Yes	Yes
Tar. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,185	3,866	3,745	4,003	3,793	3,777
Pseudo R^2	0.407	0.371	0.354	0.399	0.349	0.342
No. of Actual Deals	506	482	472	480	464	470
No. of Control Deals	3,679	3,384	3,273	3,523	3,329	3,307
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 8:
Alternative Data Measure: Synergy between Acquirer AI and Target Data

This table provides a robustness check of Table 6 Panel A by replacing the target data measure with a GPT-generated measure (Columns (1) to (3)) and cybersecurity risk measure (Columns (4) to (6)). This table reports the effects of different acquirer-target combinations on acquirer announcement returns. Deals are classified into eight groups based on the acquirer's AI capability, the target's AI capability, and the target's data intensity. Due to collinearity among these indicators, three groups are omitted as reference categories, and the table reports results for the remaining five groups. All specifications include acquirer industry × year fixed effects, acquirer state × year fixed effects and target industry × year fixed effects. Robust t-statistics adjusted for firm-level clustering are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	GPT Measure			Cybersecurity Risk		
	(1) CAR [-3,+3]	(2) CAR [-4,+4]	(3) CAR [-5,+5]	(4) CAR [-3,+3]	(5) CAR [-4,+4]	(6) CAR [-5,+5]
HighAcqAI × HighTarData	0.083**	0.095**	0.112***	0.034	0.044	0.060
	(2.34)	(2.46)	(2.87)	(0.79)	(1.15)	(1.52)
HighAcqAI × LowTarData	0.039	0.058	0.066	0.032	0.043	0.055*
	(0.81)	(1.15)	(1.23)	(1.00)	(1.58)	(1.72)
HighAcqAI × HighTarAI	0.056	0.035	0.014	0.022	0.025	-0.002
	(1.46)	(0.93)	(0.37)	(0.56)	(0.55)	(-0.04)
LowAcqAI × HighTarData	0.031	0.024	0.024	0.031	0.027	0.029
	(0.90)	(0.76)	(0.72)	(0.76)	(0.83)	(0.88)
LowAcqAI × HighTarAI	0.001	-0.013	-0.009	-0.060	-0.050	-0.042
	(0.03)	(-0.35)	(-0.22)	(-1.31)	(-1.21)	(-0.93)
Acq. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Deal Controls	Yes	Yes	Yes	Yes	Yes	Yes
Tar. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	186	186	186	175	175	175
R^2	0.735	0.755	0.753	0.710	0.747	0.744
Acq. ind. × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Acq. State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Tar. ind. × Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 9:
Post-Merger Innovation Outcomes: Applications

This table reports estimates of the effect of merger completion on post-merger innovation (combined patent applications of the acquirer and target) using a negative binomial difference-in-differences specification. The sample includes completed mergers (treatment group) and withdrawn bids (control group), estimated over three-, four-and five-year windows centered on the deal announcement year. The even-numbered columns and the the odd-numbered columns show the results with and without controlling for size-weighted acquirer–target characteristics, respectively. Column (7)-(8) present a falsification test using placebo announcement dates set three years after the actual ones. The key coefficient of interest is the AI–data interaction term, $After \times Treat \times HighAcqAI \times HighTarData$, which captures the differential post-merger change in innovation for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations. All specifications include deal and year fixed effects. Coefficients are reported in log incidence rate ratio form, with z-statistics in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Patent Applications						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After×Treat×HighAcqAI×HighTarData	1.558***	1.380***	1.195***	1.135***	1.016***	1.021***	-0.664	-0.779
	(3.73)	(3.21)	(3.08)	(2.82)	(2.74)	(2.66)	(-1.02)	(-1.21)
After×Treat×HighAcqAI	-0.350	-0.188	-0.171	-0.049	-0.097	0.049	0.455	0.626
	(-1.15)	(-0.60)	(-0.61)	(-0.17)	(-0.36)	(0.18)	(0.98)	(1.36)
After×Treat×HighTarData	-1.278***	-1.125***	-1.152***	-1.079***	-1.204***	-1.169***	-0.431	-0.334
	(-3.90)	(-3.31)	(-3.78)	(-3.38)	(-4.16)	(-3.87)	(-0.83)	(-0.65)
After×Treat	0.554**	0.432*	0.529**	0.440**	0.469**	0.357*	0.239	0.102
	(2.37)	(1.80)	(2.50)	(1.99)	(2.34)	(1.72)	(0.77)	(0.32)
After×HighAcqAI×HighTarData	-1.642***	-1.446***	-1.308***	-1.212***	-1.188***	-1.168***	0.394	0.491
	(-4.35)	(-3.74)	(-3.73)	(-3.35)	(-3.58)	(-3.41)	(0.66)	(0.83)
After×HighAcqAI	0.770***	0.577**	0.707***	0.524**	0.749***	0.537**	0.364	0.147
	(2.78)	(2.06)	(2.78)	(2.01)	(3.06)	(2.16)	(0.84)	(0.35)
After×HighTarData	1.339***	1.158***	1.234***	1.109***	1.283***	1.200***	0.448	0.298
	(4.54)	(3.82)	(4.50)	(3.90)	(4.97)	(4.50)	(0.95)	(0.63)

Table 9 Continued

	Patent Applications						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After	-0.909***	-0.727***	-0.994***	-0.822***	-1.074***	-0.877***	-0.660**	-0.521*
	(-4.14)	(-3.27)	(-5.02)	(-4.04)	(-5.76)	(-4.59)	(-2.35)	(-1.86)
Weighted_Size		-0.000		0.027		0.041		0.034
		(-0.01)		(0.87)		(1.50)		(1.24)
Weighted_Q		0.017		0.027		0.002		0.010
		(0.48)		(0.93)		(0.06)		(0.37)
Weighted_Leverage		-0.367*		-0.376**		-0.421**		-0.517***
		(-1.69)		(-2.00)		(-2.47)		(-3.04)
Weighted_Intanratio		-0.479**		-0.692***		-0.832***		-0.859***
		(-2.06)		(-3.66)		(-4.82)		(-4.98)
Weighted_ROA		0.094		-0.012		0.151		0.199
		(0.38)		(-0.05)		(0.72)		(0.96)
Observations	1,052	1,014	1,338	1,266	1,610	1,494	1,610	1,494
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 10:
Post-Merger Innovation Outcomes: Citations

This table reports estimates of the effect of merger completion on post-merger innovation (combined patent citations of the acquirer and target) using a negative binomial difference-in-differences specification. The sample includes completed mergers (treatment group) and withdrawn bids (control group), estimated over three-, four-and five-year windows centered on the deal announcement year. The even-numbered columns and the the odd-numbered columns show the results with and without controlling for size-weighted acquirer–target characteristics, respectively. The key coefficient of interest is the AI–data interaction term, $After \times Treat \times HighAcqAI \times HighTarData$, which captures the differential post-merger change in innovation for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations. All specifications include deal and year fixed effects. Coefficients are reported in log incidence rate ratio form, with z-statistics in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Patent Citations					
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]
After×Treat×HighAcqAI×HighTarData	1.539*** (2.58)	1.267** (2.12)	1.313** (2.40)	1.229** (2.25)	1.555*** (3.01)	1.594*** (3.08)
After×Treat×HighAcqAI	-0.573 (-1.29)	-0.302 (-0.67)	-0.450 (-1.12)	-0.212 (-0.53)	-0.569 (-1.50)	-0.405 (-1.06)
After×Treat×HighTarData	-1.624*** (-3.30)	-1.182** (-2.42)	-1.495*** (-3.32)	-1.201*** (-2.68)	-1.892*** (-4.49)	-1.667*** (-3.95)
After×Treat	0.996*** (2.75)	0.667* (1.82)	0.947*** (2.89)	0.630* (1.91)	1.028*** (3.39)	0.759** (2.47)
After×HighAcqAI×HighTarData	-2.183*** (-4.01)	-2.024*** (-3.72)	-1.983*** (-3.97)	-1.954*** (-3.91)	-2.184*** (-4.67)	-2.221*** (-4.72)
After×HighAcqAI	1.329*** (3.26)	1.088*** (2.62)	1.283*** (3.46)	0.988*** (2.63)	1.488*** (4.23)	1.195*** (3.36)
After×HighTarData	1.846*** (4.07)	1.500*** (3.34)	1.724*** (4.17)	1.491*** (3.62)	2.093*** (5.46)	1.864*** (4.84)

Table 10 Continued

	Patent Citation					
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]
After	-1.918*** (-5.63)	-1.605*** (-4.67)	-1.922*** (-6.23)	-1.565*** (-5.03)	-2.079*** (-7.26)	-1.696*** (-5.85)
Weighted_Size		0.196*** (5.62)		0.209*** (6.99)		0.227*** (8.55)
Weighted_Q		0.052 (1.37)		0.069** (2.10)		0.054* (1.74)
Weighted_Leverage		-1.310*** (-4.33)		-1.323*** (-5.04)		-1.063*** (-4.58)
Weighted_Intanratio		-0.800*** (-2.96)		-0.991*** (-4.41)		-1.361*** (-6.77)
Weighted_ROA		0.168 (0.46)		0.029 (0.10)		-0.059 (-0.21)
Observations	1,011	974	1,277	1,207	1,559	1,436
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 11:
Post-Merger Innovation Outcomes: Value

This table reports estimates of the effect of merger completion on post-merger innovation (combined patent values of the acquirer and target) using an OLS difference-in-differences specification. The sample includes completed mergers (treatment group) and withdrawn bids (control group), estimated over three-, four-and five-year windows centered on the deal announcement year. The even-numbered columns and the the odd-numbered columns show the results with and without controlling for size-weighted acquirer–target characteristics, respectively. Column (7)-(8) present a falsification test using placebo announcement dates set three years after the actual ones. The key coefficient of interest is the AI–data interaction term, $After \times Treat \times HighAcqAI \times HighTarData$, which captures the differential post-merger change in patent values for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations. All specifications include deal and year fixed effects. Robust t-statistics adjusted for deal-level clustering are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Patent Values						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After×Treat×HighAcqAI×HighTarData	2.078***	2.055***	1.785**	1.889**	1.612**	1.762**	-0.150	-0.195
	(2.80)	(2.75)	(2.29)	(2.35)	(2.03)	(2.08)	(-0.14)	(-0.17)
After×Treat×HighAcqAI	-0.713	-0.993*	-0.538	-0.810	-0.427	-0.629	0.318	0.247
	(-1.26)	(-1.81)	(-0.90)	(-1.36)	(-0.71)	(-1.03)	(0.45)	(0.34)
After×Treat×HighTarData	-1.346**	-1.345**	-1.144*	-1.261**	-1.194*	-1.394*	-0.023	-0.116
	(-2.38)	(-2.26)	(-1.95)	(-2.00)	(-1.86)	(-1.96)	(-0.02)	(-0.12)
After×Treat	0.623	0.850*	0.543	0.800	0.542	0.769	0.123	0.228
	(1.25)	(1.73)	(1.03)	(1.52)	(0.99)	(1.39)	(0.20)	(0.37)
After×HighAcqAI×HighTarData	-2.229***	-2.083***	-1.947***	-2.015***	-1.862***	-2.080***	0.074	0.057
	(-3.58)	(-3.28)	(-2.98)	(-2.97)	(-2.81)	(-2.93)	(0.08)	(0.06)
After×HighAcqAI	1.332***	1.445***	1.273**	1.371***	1.300**	1.355**	0.528	0.464
	(2.82)	(3.05)	(2.52)	(2.64)	(2.52)	(2.54)	(0.86)	(0.73)
After×HighTarData	1.680***	1.489***	1.445***	1.462***	1.481***	1.694***	0.025	0.083
	(3.42)	(2.77)	(2.96)	(2.66)	(2.76)	(2.74)	(0.03)	(0.10)

Table 11 Continued

	Patent Values						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After	-1.069**	-1.139***	-1.121**	-1.217***	-1.207**	-1.287***	-0.424	-0.461
	(-2.50)	(-2.63)	(-2.48)	(-2.65)	(-2.58)	(-2.68)	(-0.79)	(-0.87)
Weighted_Size		-0.366**		-0.247		-0.168		-0.232
		(-2.30)		(-1.60)		(-1.24)		(-1.57)
Weighted_Q		0.125		0.176**		0.161**		0.148*
		(1.53)		(2.26)		(2.07)		(1.80)
Weighted_Leverage		0.704		0.343		0.061		0.084
		(1.44)		(0.71)		(0.14)		(0.18)
Weighted_Intanratio		1.178***		0.754*		0.525		0.652
		(2.96)		(1.85)		(1.32)		(1.54)
Weighted_ROA		0.435		0.302		0.255		0.356
		(1.48)		(1.17)		(0.91)		(1.18)
Observations	1,413	1,353	1,776	1,671	2,123	1,961	2,123	1,961
R^2	0.926	0.926	0.918	0.919	0.906	0.906	0.903	0.904
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 12:
Post-Merger Performance of AI-Related Innovation: Application

This table reports estimates of the effect of merger completion on post-merger AI-related innovation using a negative binomial difference-in-differences specification. The sample includes completed mergers (treatment group) and withdrawn bids (control group), estimated over three-, four-and five-year windows centered on the deal announcement year. Column (7)-(8) present a falsification test using placebo announcement dates set three years after the actual ones. The dependent variable is the combined number of AI patent applications filed by the acquirer and target, as classified in the USPTO Artificial Intelligence Patent Dataset (AIPD). The key coefficient of interest is the AI–data interaction term, $After \times Treat \times HighAcqAI \times HighTarData$, which captures the differential post-merger change in AI patenting for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations. Consistent with the development cycle of AI technologies, the post-merger period begins in $t + 2$. All specifications include deal fixed effects. In the even-numbered columns, the table controls for size-weighted acquirer–target characteristics, as well as year fixed effects. In the odd-numbered columns without control variables, year fixed effects are not included to ensure model convergence. Coefficients are reported in log incidence rate ratio form, with z-statistics in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	AI Patent Applications						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After×Treat×HighAcqAI×HighTarData	1.614***	1.172**	1.580***	1.332***	1.109**	1.203***	-0.621	-0.801
	(3.10)	(2.55)	(3.20)	(3.13)	(2.25)	(2.97)	(-0.52)	(-0.73)
After×Treat×HighAcqAI	-0.768*	0.199	-0.785**	0.132	-0.644*	0.162	0.384	1.206
	(-1.94)	(0.53)	(-2.11)	(0.38)	(-1.78)	(0.50)	(0.42)	(1.46)
After×Treat×HighTarData	-0.793**	-0.284	-0.917**	-0.599*	-0.605	-0.715**	0.118	-0.009
	(-1.98)	(-0.86)	(-2.35)	(-1.85)	(-1.55)	(-2.31)	(0.13)	(-0.01)
After×Treat	0.281	-0.520**	0.452	-0.315	0.426	-0.236	0.341	-0.429
	(0.94)	(-1.98)	(1.58)	(-1.24)	(1.59)	(-0.99)	(0.65)	(-1.11)
After×HighAcqAI×HighTarData	-1.401***	-1.265***	-1.215***	-1.211***	-0.657	-1.004***	1.396	1.192
	(-3.14)	(-3.13)	(-2.82)	(-3.22)	(-1.50)	(-2.80)	(1.26)	(1.15)
After×HighAcqAI	0.835**	0.065	0.946***	0.168	0.838**	0.162	-0.174	-0.782
	(2.42)	(0.20)	(2.84)	(0.55)	(2.57)	(0.56)	(-0.20)	(-0.98)

Table 12 Continued

	AI Patent Applications						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After×HighTarData	0.867**	0.533*	0.786**	0.639**	0.366	0.656**	-0.692	-0.306
	(2.53)	(1.90)	(2.30)	(2.28)	(1.06)	(2.44)	(-0.87)	(-0.43)
After	-0.622**	0.356	-0.782***	0.153	-0.793***	0.047	-0.826*	0.281
	(-2.40)	(1.49)	(-3.05)	(0.65)	(-3.32)	(0.21)	(-1.76)	(0.82)
Weighted_Size		0.243***		0.220***		0.184***		0.164***
		(4.83)		(5.13)		(4.81)		(4.35)
Weighted_Q		0.139**		0.178***		0.140***		0.129***
		(2.37)		(3.78)		(3.31)		(3.04)
Weighted_Leverage		-1.028***		-0.953***		-0.792***		-0.655***
		(-3.80)		(-4.02)		(-3.80)		(-3.31)
Weighted_Intanratio		-1.252***		-1.456***		-1.482***		-1.504***
		(-4.50)		(-6.39)		(-7.26)		(-7.41)
Weighted_ROA		-0.115		-0.183		-0.260		-0.356
		(-0.29)		(-0.55)		(-0.84)		(-1.17)
Observations	797	766	1,021	965	1,239	1,142	1,239	1,142
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

Table 13:
Post-Merger Performance of AI-Related Innovation: Value

This table reports estimates of the effect of merger completion on post-merger AI-related patent value using an OLS difference-in-differences specification. The sample includes completed mergers (treatment group) and withdrawn bids (control group), estimated over three-, four-and five-year windows centered on the deal announcement year. The even-numbered columns and the the odd-numbered columns show the results with and without controlling for size-weighted acquirer–target characteristics, respectively. Column (7)-(8) present a falsification test using placebo announcement dates set three years after the actual ones. The dependent variable is the combined number of AI patent value filed by the acquirer and target, as classified in the USPTO Artificial Intelligence Patent Dataset (AIPD). The key coefficient of interest is the AI–data interaction term, $After \times Treat \times HighAcqAI \times HighTarData$, which captures the differential post-merger change in AI patent value for completed mergers that combine high-AI acquirers with data-intensive targets, relative to withdrawn bids and to other acquirer–target combinations. Consistent with the development cycle of AI technologies, the post-merger period begins in $t + 2$. All specifications include deal and year fixed effects. Robust t-statistics adjusted for deal-level clustering are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	AI Patent Values						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After×Treat×HighAcqAI×HighTarData	1.394**	1.425**	1.290*	1.446**	1.431**	1.450**	0.688	0.650
	(2.16)	(2.15)	(1.79)	(2.12)	(2.17)	(2.26)	(0.78)	(0.74)
After×Treat×HighAcqAI	-0.693	-0.788*	-0.633	-0.779	-0.324	-0.410	1.455**	1.366**
	(-1.53)	(-1.67)	(-1.30)	(-1.55)	(-0.77)	(-0.93)	(2.49)	(2.23)
After×Treat×HighTarData	-1.005**	-0.952**	-0.772	-0.898*	-0.860*	-0.854*	-0.633	-0.615
	(-2.27)	(-2.04)	(-1.51)	(-1.76)	(-1.77)	(-1.77)	(-1.22)	(-1.34)
After×Treat	0.518*	0.563*	0.496*	0.599*	0.403	0.479	-0.166	-0.093
	(1.89)	(1.83)	(1.66)	(1.74)	(1.41)	(1.51)	(-0.64)	(-0.37)
After×HighAcqAI×HighTarData	-1.121**	-1.142**	-1.049*	-1.111*	-1.263**	-1.234**	-1.082	-1.077*
	(-2.15)	(-2.12)	(-1.72)	(-1.97)	(-2.40)	(-2.49)	(-1.59)	(-1.66)
After×HighAcqAI	0.515	0.527	0.502	0.491	0.451	0.394	-0.415	-0.411
	(1.53)	(1.42)	(1.26)	(1.17)	(1.41)	(1.15)	(-0.84)	(-0.78)

Table 13 Continued

	AI Patent Values						Placebo Test	
	(1) [-3,+3]	(2) [-3,+3]	(3) [-4,+4]	(4) [-4,+4]	(5) [-5,+5]	(6) [-5,+5]	(7) [-5,+5]	(8) [-5,+5]
After×HighTarData	0.684*	0.587	0.663	0.624	0.809*	0.655	0.579	0.522
	(1.82)	(1.43)	(1.45)	(1.37)	(1.88)	(1.52)	(1.30)	(1.41)
After	-0.311	-0.304	-0.407	-0.366	-0.394	-0.295	-0.160	-0.165
	(-1.29)	(-1.09)	(-1.48)	(-1.14)	(-1.52)	(-0.99)	(-0.71)	(-0.72)
Weighted_Size		-0.329**		-0.415***		-0.361***		-0.339***
		(-2.61)		(-2.97)		(-2.97)		(-3.04)
Weighted_Q		-0.101		-0.051		-0.014		-0.022
		(-0.84)		(-0.48)		(-0.14)		(-0.22)
Weighted_Leverage		0.115		0.162		-0.010		-0.050
		(0.26)		(0.39)		(-0.03)		(-0.13)
Weighted_Intanratio		0.359		0.782*		0.860**		0.775*
		(0.81)		(1.75)		(2.08)		(1.97)
Weighted_ROA		0.537		0.538		0.719**		0.715**
		(0.97)		(1.28)		(2.01)		(2.01)
Observations	871	848	1,102	1,053	1,327	1,244	1,327	1,244
R^2	0.518	0.542	0.473	0.493	0.469	0.483	0.474	0.487
Deal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix A Variable definitions and data sources

Acq_AIEmp Average monthly number of AI-skilled employees at the acquirer firm over the 12 months preceding the deal announcement.

Acq_AISkill Average monthly number of AI-related skills at the acquirer firm over the 12 months preceding the deal announcement.

Acq_Emp Average monthly number of employees at the acquiring firm over the 12 months preceding the deal announcement.

LnAcqAIEmp Natural logarithm of one plus Acq_AIEmp.

LnAcqEmp Natural logarithm of Acq_Emp.

LnAcqAISkill Natural logarithm of one plus $Acq_AISkill$.

HighAcqAI Indicator variable equal to one if the acquirer's $Acq_AISkill$ is above the sample median, and zero otherwise.

Intanratio Ratio of intangible assets to total assets for the acquirer as of the fiscal year-end preceding the deal announcement, computed as `intan/at` using Compustat data.

Leverage Acquirer leverage as of the fiscal year-end preceding the deal announcement, computed as `(dlc+dltt)/at` using Compustat data.

ROA Return on assets of acquirer as of the fiscal year-end preceding the deal announcement, computed as `ni/at` using Compustat data.

RD R&D expense of the acquirer as of the fiscal year-end preceding the deal announcement, computed as `xrd/at` and set to zero if missing, using Compustat data.

Salesgrowth Annual sales growth of the acquirer as of the fiscal year-end preceding the deal announcement, using Compustat data.

Size Natural logarithm of acquirer size, measured as market capitalization 11 days prior to the deal announcement for deal-level regressions, and the natural log of book assets as of the fiscal year-end preceding the deal announcement for firm-level regressions. Market capitalization is computed as `abs(prc*shrout)/1000`, and book assets equal `at`, using Compustat and CRSP data.

Tobin's Q Tobin's Q of the acquirer as of the fiscal year-end preceding the deal announcement, computed as `(at-ceq+prcc_f*csho)/at` using Compustat data.

DataIntensity Frequency of data security–related keywords in Item 1A risk disclosures from firms' 10-K filings.

DataIntensity_GPT Alternative measure of data intensity generated using a large language model applied to firms' 10-K filings.

HighTarData Indicator variable equal to one if the target's *DataIntensity* is above the sample median, and zero otherwise.

HighTarData_GPT Indicator variable equal to one if the target's *DataIntensity_GPT* is above the sample median, and zero otherwise.

Tar_AIEmp Average monthly number of AI-skilled employees at the target firm over the 12 months preceding the deal announcement.

Tar_AISkill Average monthly number of AI-related skills at the target firm over the 12 months preceding the deal announcement.

Tar_Emp Average monthly number of employees at the target firm over the 12 months preceding the deal announcement.

Tar_Intanratio Ratio of intangible assets to total assets for the target as of the fiscal year-end preceding the deal announcement, computed as `intan/at` using Compustat data.

Tar_Leverage Target leverage as of the fiscal year-end preceding the deal announcement, computed as `(dlc+dltt)/at` using Compustat data.

Tar_RD R&D intensity of the target as of the fiscal year-end preceding the deal announcement, computed as `xrd/at` and set to zero if missing, using Compustat data.

Tar_ROA Return on assets of the target as of the fiscal year-end preceding the deal announcement, computed as `ni/at` using Compustat data.

Tar_Salesgrowth Annual sales growth rate of the target as of the fiscal year-end preceding the deal announcement, using Compustat data.

Tar_Size Natural logarithm of target book assets (`at`) as of the fiscal year-end preceding the deal announcement, using Compustat data.

Tar_Tobin's Q Tobin's Q of the target as of the fiscal year-end preceding the deal announcement, computed as `(at-ceq+prcc_f*csho)/at` using Compustat data.

TargetStatus Indicator variable equal to one if the target is public according to SDC, and zero otherwise.

CAR[-3,+3] Market-model cumulative abnormal returns (CARs) over the $[-3, +3]$ event window surrounding the M&A announcement. Market-model parameters are estimated using the CRSP value-weighted index over $[-240, -11]$ trading days relative to the announcement date.

CAR[-4,+4] Market-model cumulative abnormal returns (CARs) over the $[-4, +4]$ event window, estimated using the CRSP value-weighted index over $[-240, -11]$ trading days.

CAR[-5,+5] Market-model cumulative abnormal returns (CARs) over the $[-5, +5]$ event window, estimated using the CRSP value-weighted index over $[-240, -11]$ trading days.

CombinedCAR Weighted average of the target's CAR $[-41, +3]$ and the acquirer's CAR $[-3, +3]$, using market capitalizations on day -42 (target) and day -4 (acquirer) as weights. The target's CAR $[-41, +3]$ is estimated using a market model fitted over $[-240, -42]$.

Premium Acquisition premium, calculated as the target's CAR $[-41, +3]$ multiplied by its market capitalization on day -42, scaled by the combined market capitalizations of the target (day -42) and the acquirer (day -4).

AllStock Indicator variable equal to one if the transaction is paid entirely in stock (SDC variable `ofstock` = 100), and zero otherwise.

Conglomerate Indicator variable equal to one if the acquirer and target operate in different industries—defined as different 4-digit SIC codes in Table 2 and different 2-digit SIC codes elsewhere–and zero otherwise.

Challenge Indicator variable equal to one if the SDC variable `cha` equals "Yes," and zero otherwise.

Crossborder Indicator variable equal to one if the acquirer and target are from different countries, and zero otherwise, based on SDC data.

HighTechdeal Indicator variable equal to one if both the acquirer's and target's 4-digit SIC codes fall within a predefined set of high-technology industries,[9] and zero otherwise.

Rel_DealVal Deal value divided by the acquirer's total assets, using deal value data from SDC and total assets from Compustat.

SameState Indicator variable equal to one if the acquirer and target are located in the same U.S. state, and zero otherwise.

AI Patent Application Total number of AI-related patent applications filed by the acquirer and the target. A patent is classified as AI-related if `predict93_any_ai` equals one in the AIPD.

AI Patent Value Natural logarithm of one plus total value creation of AI-related patents received by the acquirer and the target. Patent value is provided by KPSS (`xi_real`), and is deflated to 1982 (million) dollars using the CPI.

Patent Application Total number of patent applications filed by the acquirer and the target.

[9] 3571, 3572, 3575, 3577, 3578, 3661, 3663, 3669, 3671, 3672, 3674, 3675, 3677, 3678, 3679, 3812, 3823, 3825, 3826, 3827, 3829, 3841, 3845, 4812, 4813, 4899, 7371, 7372, 7373, 7374, 7375, 7378, or 7379.

Patent Citation Total number of forward citations received by patents filed by the acquirer and the target.

Patent Value Natural logarithm of one plus total value creation of patents received by the acquirer and the target. Patent value is provided by KPSS (`xi_real`), and is deflated to 1982 (million) dollars using the CPI.

Weighted_Intanratio Asset-weighted average of acquirer and target intangible asset ratios prior to the acquisition, and the intangible asset ratio of the combined firm after the acquisition.

Weighted_Leverage Asset-weighted average of acquirer and target leverage prior to the acquisition, and leverage of the combined firm after the acquisition.

Weighted_ROA Asset-weighted average of acquirer and target ROA prior to the acquisition, and ROA of the combined firm after the acquisition.

Weighted_Size Asset-weighted average of acquirer and target size prior to the acquisition, and size of the combined firm after the acquisition.

Weighted_Q Asset-weighted average of acquirer and target Tobin's Q prior to the acquisition, and Tobin's Q of the combined firm after the acquisition.

Internet Appendix: When AI Acquires Data: Strategic Complementarities in M&A

Yue Fang
Vidhan Goyal
Yingxuan He
Zilong Zhang

August 9, 2026

Internet Appendix IA: Post-Merger Innovation Treatment Sample Criteria

Internet Appendix IB: Post-Merger AI Innovation Treatment Sample Criteria

IA Post-Merger Innovation Treatment Sample Criteria

Criteria	Treated Deals Left	Control Deals Left
Acquirer-target industry pairs that match those of the withdrawn bids in the control sample	335	68
Announcement year falls within a three-year window centered around the announcement year of the control bids to minimize time-related differences	258	58
For each control bid, select up to 10 matched completed deal with the closest relative-size ratio (target assets divided by acquirer assets)	188	58

IB Post-Merger AI Innovation Treatment Sample Criteria

Criteria	Treated Deals Left	Control Deals Left
Innovative acquirers and target firms	778	40
Acquirer-target industry pairs that match those of the withdrawn bids in the control sample	192	40
Announcement year falls within a three-year window centered around the announcement year of the control bids to minimize time-related differences	142	34
For each control bid, select up to 10 matched completed deal with the closest relative-size ratio (target assets divided by acquirer assets)	115	34