Generative AI – Driven Management Research Methods: Current Developments, Application and Future Directions

Chen Zhao[1] Chen Lin[2]

(1. School of Economics and Management, Beijing University of Posts and Telecommunications;

2. Institute of Quantitative and Technological Economics, Chinese Academy of Social Sciences)

Abstract: Generative AI is reshaping the methodological framework of management research, yet its application still faces challenges such as fragmented methodologies and insufficient technical reliability. While existing literature has validated the potential of generative AI in fields like text mining and sentiment analysis, systematic application frameworks are lacking, and issues such as hallucinated outputs and reproducibility deficiencies hinder its adoption in rigorous academic research. This gap urgently needs addressing, the efficiency, precision, and scalability of generative AI offer significant opportunities for innovation in management research paradigms, whereas the absence of standardized methods may compromise the validity and reliability of research findings. To address this gap, this study systematically examines the current application of generative AI in management research, aiming to distill standardized workflows and propose forward – looking directions for the field.

The study employs bibliometric and content analysis methods to systematically review domestic and international literature that integrates generative AI with management research methods, drawn from the Chinese Social Sciences Citation Index (CSSCI) database and the Web of Science Core Collection. It analyzes publication characteristics and keyword distributions, with a focus on the opportunities and challenges, application models, methodological safeguards, and validation mechanisms of generative AI in management research.

The findings reveal that the application of generative AI in management research exhibits an initial surge followed by steady growth, with applications primarily concentrated in tasks such as text mining and information extraction. It demonstrates broad potential at the macro, meso, and micro levels, spanning strategic planning, tactical decision – making, operational processes, and management information systems. Its advantages encompass efficiency, precision, objectivity, interpretability, and scalability, though reliability and reproducibility remain core challenges. Based on the literature analysis, this study proposes the "12C – PIPE Matrix Model," which covers four phases of generative AI – enabled management research—Preparation, Integration, Processing, and Evaluation—with three operational steps (3C) specified for each phase, forming a 4 × 3 matrix of 12 key procedures (12C) that provides comprehensive methodological guidance for researchers. Additionally, future research directions are outlined in terms of method validation, applicable boundaries, ecosystem development, and ethical governance.

The primary contribution of this study lies in systematically uncovering the methodological applications of generative AI in management research and proposing the 12C – PIPE matrix model as a standardized procedural framework that addresses the fragmentation of existing methods. This provides operational guidelines for the academic community to standardize the use of generative AI and opens new avenues for the application of generative AI tools in management research.

Key Words: artificial intelligence; management research methods; research tools; large language models

AI for Interesting Social Science: A Literature Review on Theoretical Interestingness and Its Implications

Xianyong Zhou[1], Yuxin Liu[2]

(1. School of Public Administration, Southwest Jiaotong University;

2. School of Public Policy and Management, University of Chinese Academy of Sciences)

Abstract: As AI for Social Science (AI4SS) continues to advance, how artificial intelligence (AI) can be used to enhance the theoretical interestingness of social science has become an increasingly important issue. This study systematically reviews the literature on theoretical interestingness and explores its implications in the age of AI. Our findings suggest that the essence of theoretical interestingness lies in denying certain assumptions of the audience. However, this understanding remains largely confined to discursive production, paying insufficient attention to the social construction of theoretical interestingness. Instead, this study argues that theoretical interestingness is a collective consensus reached by knowledge communities through discursive negotiation within intellectual traditions and institutional contexts. In this sense, a knowledge community composed of humans and AI may co-construct an interesting collective consensus through grounding in intellectual traditions, adapting to institutional contexts, and engaging in discursive negotiation. In particular, we identify three major risks, including AI hallucinations, algorithmic convergence, and critical inertia, which need to be addressed with caution. By bridging the theoretical gap between the classical philosophy of social science and the contemporary AI methodology, this study offers practical strategies and implications for using AI to enhance the theoretical interestingness of social science research in China.

Key Words: Artificial Intelligence; AI for Social Science; Theoretical Interestingness; Social Science

The Integration of AI – Generated Synthetic Data and Business Administration Research: Review and Prospect

Zhengyin Peng[1], Yanyu Zhao[1], Jialei Li[1,2], Xinyang Wang[1]

(1. School of Business, Tianjin University of Finance and Economics;

2. School of Economics and Management, Dalian Jiaotong University)

Summary: The artificial intelligence – driven synthetic data technology is constantly breaking through, effectively alleviating the " data shortage" in the digital and intelligent era. Synthetic data optimizes research design and methods and is embedded in the innovative process of various scientific research fields, especially in the field of business administration, providing pre – verification conditions for related research methods and alleviating the data shortage problem in some research. Top international business administration journals have widely utilized synthetic data for research, but domestic attention to synthetic data is very insufficient. This paper comprehensively applies machine learning and knowledge graph methods to conduct systematic literature analysis. First, the paper defines the basic concept of synthetic data, elaborates on its generation method, and analyzes the current research status and evolution process. Secondly, it selects relevant topic literature published in UTD24 and FT50 journals from 1993 to 2025, uses the BERTopic model to conduct text analysis of 133 English documents, combines coding techniques to extract two core issues of synthetic data embedding in business administration: research methods and practical applications. Finally, from the aspects of synthetic data governance, disciplinary applications, and research methodology innovation, it looks forward to the research direction of synthetic data in the field of business administration. This study provides an application framework for domestic teams to join the international dialogue in this field, provides a methodological basis for further deepening the application of synthetic data in business administration research, and has important theoretical and practical significance.

Key Words: Business Administration; Research Methods; Synthetic Data; Systematic Review; Topic Model

Purpose

With the advancement of deep learning techniques such as generative adversarial networks and Transformer models, AI – driven synthetic data technology has gradually gained attention from leading international journals in the business administration field, whereas domestic attention to synthetic data remains notably insufficient. Given this gap, this study integrates machine learning and knowledge graph methods to conduct a systematic literature analysis, offering an application framework for domestic research teams to engage in global academic dialogue and providing a methodological foundation for deepening the application of synthetic data in business administration research.

Design

This study employs the BERTopic model and CiteSpace to conduct a systematic literature analysis of synthetic data research. First, CiteSpace is used to perform knowledge graph analysis on 254 English – language synthetic data articles and their cited references retrieved from the Web of Science database in fields such as management, business, and economics, thereby capturing the overall research landscape of synthetic data within the business administration domain. Second, 133 relevant articles published between 1993 and 2025 in UTD24 and FT50 journals are selected, and the BERTopic model is applied for textual analysis.

Combined with manual coding, two core themes for embedding synthetic data into business administration research are identified: research methods and practical applications. The research methods theme mainly comprises robust optimization and decision making, causal inference and experimental design, and market space segmentation; the practical applications theme primarily includes consumer choice and decision making, dynamic pricing, and other research areas.

Findings

This paper's systematic review of the synthetic data literature reveals the following findings: (1) The application scenarios of synthetic data are becoming increasingly diverse. With technological advances, synthetic data research and applications have expanded from a single discipline to multidisciplinary domains, and its connotations have been progressively enriched. (2) Current research themes on synthetic data generally exhibit a logical progression from methods to practical applications. (3) Synthetic data enriches and extends causal inference theory. On the one hand, synthetic data provides a manipulable testing environment that can be used for simulation validation of various identification strategies; by examining estimation bias, the validity and robustness of methods can be assessed. It can also assist research design in settings with confounders, testing the applicability conditions and effects of control strategies, and thereby deepening the verification foundation of relevant research methods. On the other hand, synthetic data can be applied to theoretical model construction, traditional statistical method validation, and simulation of management practices, providing strong support for the feasibility of methods and the practical relevance of research findings.

Value

The potential marginal contributions of this paper are as follows. First, it clarifies the conceptual characteristics, generation methods, evolutionary stages, and development trajectory of synthetic data, thereby constructing a current analytical framework for synthetic data within the domestic context. Second, by integrating machine learning and knowledge graph analysis methods, combining the BERTopic model with manual verification, and grounding the analysis in the field of business administration, it identifies two core themes for integrating synthetic data into business administration research. Third, it discusses the specific application modes of synthetic data in business administration, attempts to propose usage procedures and evaluation metrics for synthetic data within business administration research methods, and thereby provides an application framework for domestic teams to engage in the international dialogue on this topic, as well as a methodological foundation for further deepening the application of synthetic data in business administration research.

Suggestions for future research

In terms of synthetic data governance, risk identification and mitigation during the data generation process remain a critical issue, and discussions on governance mechanisms and strategies for synthetic data should be embedded in practical managerial application contexts. In terms of disciplinary application, the cross - fertilization of synthetic data with management and other research fields should be systematically advanced, and the application scenarios and value mechanisms of the synthetic data - driven paradigm in diverse management settings should be further explored. In terms of research method innovation, future research could focus on its practical use in real - world scenarios such as scale development and validity assessment, and data imputation.

Complementarity and Boundaries of Multimodal Information Integration: Implications for Empirical Management Research

Zhuang Liu, Yiming Chang

(School of FinTech, Dongbei University of Finance and Economics)

Abstract: Against the backdrop of the rapid development of artificial intelligence, especially LLMs and multimodal technologies, researchers can more easily extract latent features from unstructured data such as text, images, and audio, and use them to construct research indicators. Multimodal fusion has therefore gradually become an important information representation strategy in management research. Existing studies generally suggest that the application and integration of multimodal data can improve information representation and prediction accuracy; however, this is not always the case in practice. From an information-processing perspective, this study proposes that multimodality does not necessarily lead to better information representation; its effectiveness depends on the complementary structure among modalities, not on the simple accumulation of modalities. Taking BYD as the empirical research object, this study constructs a multi-source dataset covering stock market data, text, images, and audio, and compares the effects of different modality fusion approaches under the same model framework. The findings show that, in this case, multimodal data fail to form effective complementarity, leading to the dilution of core feature weights. Even after changing modality combinations and fusion methods, the newly added modalities still do not provide task-relevant incremental information. This study verifies the existence of boundary effects and provides a formal definition of complementarity, offering methodological insights for applying multimodal approaches to information representation, indicator construction, and research design in management studies.

Key Words: Multimodal information integration; Boundary effects; Modality complementarity; Information representation; Management research

Purpose

With the rapid development of artificial intelligence, especially multimodal technologies, management researchers can more easily extract relevant features from unstructured data such as text, images, and audio, and use them to construct new research indicators. Existing studies generally assume that adding more modalities can provide additional information and therefore improve information representation, prediction accuracy, and decision quality. However, this is not always the case. The introduction of additional modalities may also lead to information redundancy and increased data noise, resulting in higher information-processing costs and weaker interpretability. To address this gap, this study empirically examines whether multimodal integration necessarily leads to better information representation and under what conditions it may fail. The core proposition of this paper is that multimodality does not necessarily produce better information representation; rather, its effectiveness depends on the complementarity among modalities, instead of the simple accumulation of modalities in number.

Design/methodology/approach

This study takes BYD as the empirical research object and constructs a multisource dataset covering stock market data, news text, image data, and audio information. News reports related to BYD are collected from *People's Daily*, while image and audio data are obtained from public videos such as BYD product launches, speeches, and related events. Text sentiment is analyzed using a specially trained news sentiment analysis model. Image sentiment is extracted through video frame sampling and face recognition. Audio sentiment is obtained through speech transcription, text summarization, and sentiment scoring using a pretrained model.

To test the core proposition of this paper, this study first provides a formal definition of complementarity among modalities, namely the three conditions of complementarity. Based on this framework, the study compares the performance of the multimodal model with that of the unimodal baseline model under the same dual – channel LSTM framework. In this framework, one channel processes market – related variables, while the other channel processes sentiment – related variables. In addition, this study uses the Integrated Gradients method to conduct feature contribution analysis, and applies strategy backtesting to evaluate the practical decision – making value of the models. Robustness checks, ablation experiments, weight – grid searches, and comparisons of different fusion methods are also conducted to examine the reliability of the findings.

Findings

The results show that, in this empirical setting, multimodal integration does not generate effective informational gains; instead, it causes the dilution of useful features. Specifically, the Integrated Gradients feature contribution analysis shows that the contribution of sentiment interaction variables declines significantly in the multimodal model, and some interaction terms provide almost no effective contribution. At the same time, in the strategy backtesting experiment, the unimodal model outperforms the multimodal model in almost all aspects, including cumulative returns, win rate, and risk – adjusted returns.

Further tests show that changing modality combinations, adjusting modality weights, or replacing the fusion method still cannot make the multimodal model outperform the best unimodal model. This indicates that the decline in model performance is not caused by a single parameter setting or random fluctuation, but by the lack of effective complementarity among modalities. To further verify this point, this study examines the complementary structure among modalities based on the proposed three conditions of complementarity: information non – redundancy, task – relevant information gain, and fusion usability. The results show that the three modalities only satisfy the condition of information non – redundancy, but do not meet the other two conditions. Therefore, they do not form effective complementarity. Overall, this study argues that multimodality does not necessarily lead to better information representation. Modality complementarity, rather than the number of modalities, is the key mechanism determining whether multimodal integration can generate effective incremental value.

Originality/value

This study makes four main contributions. First, it revises the implicit assumption that expanding information modalities can improve information representation. Second, it proposes a formal definition of modality complementarity and uses three criteria—information non – redundancy, task – relevant incremental information, and fusion usability—as the basis for judging complementarity. This clarifies the boundary conditions of multisource information integration. Third, this study constructs a comprehensive multimodal stock market dataset and empirically verifies the existence of boundary effects in information integration

through systematic comparative experiments, interpretability analysis, robustness checks, and ablation experiments. It also reveals the mechanism through which complementarity affects information representation. Fourth, this study extends the discussion of multimodal methods from a purely technical performance issue to a broader issue of information representation, indicator construction, and applicability boundaries in management research.

Implications/research limitations/suggestions for future research

The findings suggest that management researchers and practitioners should not introduce multimodal information merely because such data are available, or because a certain modality appears to provide additional information on the surface. Before introducing a new modality, researchers should use small – scale data or diagnostic analysis to examine whether it provides non – redundant information, whether it is relevant to the research task, and whether the existing fusion method can truly make use of this information. This can help avoid unnecessary data collection costs and reduce the introduction of noise. This study also has several limitations. For example, it focuses only on BYD as a single firm. Future research can further extend this framework to cross – firm samples, different management decision – making tasks, richer multimodal fusion architectures, and causal research designs. More importantly, future studies can develop end – to – end multimodal diagnostic tools to help management researchers more easily judge whether a newly introduced modality can provide effective informational gains and improve information representation.

Artificial Intelligence Patent Identification Using Generative Large Language Models: A Methodological Framework and Empirical Study

Xincheng Wang[1]; Yuming Guo[2]; Yuan Li[2]; Xiaoyu Yu[3]

(1. School of Economics and Management, Tongji University;

1. Laboratory of High Quality Urban Development and Strategic Decision, Tongji University;

2. Antai College of Economics and Management, Shanghai Jiao Tong University;

3. School of Management, Shanghai University)

Abstract: Artificial intelligence patents represent crucial artifacts of technological innovation, encoding essential technical trajectories and application development pathways. However, the exponential growth in AI patent volumes, coupled with increasing terminological complexity and continuously shifting innovation boundaries, has exposed significant limitations in conventional identification methodologies regarding temporal adaptability, coverage breadth, and classification consistency. This research conducts a systematic examination of theoretical foundations and application contexts governing existing patent identification paradigms, subsequently proposing an innovative framework for AI patent identification leveraging generative Large Language Models (LLMs), substantiated through rigorous empirical validation. Employing standardized datasets and controlled experimental protocols, we perform comparative analyses contrasting LLM – based approaches against both traditional machine learning and deep learning benchmarks. Empirical findings demonstrate the superior performance of our LLM – driven methodology in out – of – sample identification tasks, particularly excelling in processing complex patent documents characterized by nebulous technical boundaries and rapidly evolving terminological frameworks.

The verifiable and replicable AI patent identification model advanced in this study provides not only robust methodological support for patent corpus construction and innovation policy evaluation but also establishes a reliable measurement instrument for empirical AI research, embodying substantial theoretical contributions and practical implications.

Key Words: AI patent; Patent classification; Large language models; Machine learning

Amplification and Mitigation of Gender Gap in the Academic Reputation of Supervisors: Experimental Evidence from Large Language Models

Zhidong Zhao[1], Matthew Tingchi Liu[2], Zhenghao Zhu[3]

(1. School of Business, Jiangsu Second Normal University;

2. Faculty of Business Administration, University of Macau;

3. School of Business and Trade, Nanjing University of Industry Technology)

Abstract: The rapid advancement of Large Language Models (LLMs) has raised concerns regarding their potential to exacerbate the spread of social biases. This study investigates the extent of discrimination exhibited by LLMs, based on interactions among gender, age, and academic qualifications, in the context of evaluating academic reputations of scholarly supervisors. Our experimental results indicate that LLMs assign lower academic reputation ratings to middle-aged female scholars compared to their male counterparts, and anticipate higher costs for these women when competing for the same graduate student. This highlights an amplifying effect of age on gender disparities in perceived academic reputation. Furthermore, our findings show that female scholars with doctoral degrees are rated by LLMs as having higher academic reputations and lower expected costs compared to their male peers, suggesting a mitigating effect of academic qualifications on gender disparities. Auxiliary experiments confirmed the presence of derivative effects arising from these two primary effects and demonstrated their out-of-sample significance, thereby validating their robustness. These findings challenge the objectivity and neutrality of LLM outputs and highlight the need for corrective strategies and mechanisms in the application of AI in educational and academic fields.

Key Words: Supervisors; Academic Reputation; Gender Gap; Interplay; Large Language Models

Purpose

Large language models (LLMs) are increasingly embedded in academic evaluation and educational decision-making contexts. Existing studies have shown that LLMs may reproduce and even amplify the social biases inherent in their training data, including stereotypes such as gender-occupation links and ethnicity-performance associations. However, extant literature tends to examine the biases in isolation or to treat multiple biases as operating independently, thus overlooks the possibility that they may interact in complex and nonlinear ways. In higher-educational settings where LLMs may be used to support evaluation of academic supervisors, little is known about whether and how these models replicate real-world social biases and discrimination patterns. To address these gaps, this study adopted an intersectional theoretical perspective and hypothesized that LLMs would generate differential academic reputation evaluations based on the supervisor's gender, age, and academic credentials, and that these biases would exhibit interactive rather than merely additive effects.

Design

This study employed a randomized controlled experiment to measure the causal influence of identity attributes on LLM-generatedsupervisor evaluations. Real-world textual evaluations of STEM supervisors were collected from an online platform. The

linguistic content, provided anonymously by students describing their opinions on supervisors' mentoring qualities, was held constant. While supervisors' identity cues—including gender (male/female), age status (middle – aged), and educational attainment (a Ph. D. degree) —were systematically manipulated across experimental conditions. These modified profiles were then submitted to the SKEP model from Baidu and the GPT – 4o model from OpenAI: the former was used to generate an academic reputation score for the supervisors, and the latter was prompted to provide outputs about the expected financial support the supervisors would need to offer to recruit a graduate student. Statistical comparisons across controlled variations enabled direct inference on whether and how identity traits shaped the model's evaluative output.

Findings

The results demonstrated that the LLMs produced systematically biased evaluations based on gender, age, and academic credentials. Middle – aged supervisors, regardless of gender, were assigned higher expected stipends than supervisors without disclosed identity attributes ($p < 0.01$), indicating an algorithmic embedding of age bias. Moreover, middle – aged women were expected to offer substantially higher support (6.3% above baseline) for a same graduate student compared to their male counterparts (3.6% above baseline), suggesting a reinforcing interaction between gender and age biases. However, possession of a doctoral degree reduced the stipends assigned to women more than to men ($p < 0.05$), indicating that academic credentials may function as a bias – attenuating factor. These patterns were further confirmed by another two derivative effects: the cancelling effect of gender gap when both 'middle – age' and 'Ph. D degree' identity attributes were included, and the augmented – mitigation effect of gender gap when a harder – to – achieve academic credential was assigned. These patterns were robust to replication using evaluations from economics and management disciplines as well.

Contribution

This study contributes to the literature by moving beyond the binary question of whether LLMs are biased and showing instead that bias can operate through interaction effects, thereby contributing to theory by demonstrating that LLM biases are intersectional and structurally patterned, rather than independent or mechanically additive. It shows that algorithmic evaluations in educational and scholarly contexts can subtly reproduce existing social biases, and thus challenges the presumed neutrality of AI – assisted academic assessment. Methodologically, the study introduces a replicable experimental framework for detecting multidimensional bias in LLM outputs. In practical terms, the findings highlight the necessity for corrective strategies and mechanisms in the application of AI in educational and academic fields.

Implication

The experiments suggest that LLMs may not simply reflect historical inequalities but also intensify them under specific combinations of social and credential – related cues. This has important practical implications for the use of LLMs in academic decision – making, such as recruitment, evaluation, and screening, where unchecked model outputs may reproduce and legitimize gendered disadvantage, distort talent allocation, and ultimately undermine fairness, diversity, and long – term institutional efficiency. These findings suggest that organizations should treat LLMs as tools requiring active governance, including algorithmic auditing, debiasing interventions, and human oversight.

Research Limitations

This study has several limitations, though. The experimental materials are based on anonymous student evaluations in a

relatively specific disciplinary and cultural setting, which may limit generalizability to other contexts, evidence types, and academic fields. In addition, although the research design helps isolate causal effects, the internal mechanisms through which LLMs generate biased evaluations remain insufficiently understood. Future research should therefore test whether similar patterns hold across more objective academic materials and more diverse institutional settings, while also using explainable AI and related methods to uncover the deeper representational mechanisms through which bias is formed, activated, and potentially amplified.

Artificial Intelligence Capability Gap and Opportunism in Strategic Alliances: A Technological Invasion Perspective

Jinlei Yu[1], Rui Chen[1], Qi Wu[2]

(1. Business School, Nankai University; 2. School of Management, Xiamen University)

Abstract: Strategic alliances expose firms to technological invasion, whereby one firm leverages collaboratively acquired knowledge to file patents in its partner's core technology domains. This study introduces AI capability gaps as a novel source of information asymmetry that amplifies this risk. Drawing on transaction cost economics, we argue that a larger AI capability gap between alliance partners increases the probability of technological invasion by the focal firm, and examine how equity – based alliances and relational embeddedness moderate this relationship. Using a panel of 10531 partner – focal firm – year observations from Chinese A – share listed firms (2010—2023), fixed – effects regressions confirm that AI capability gaps significantly increase technological invasion, while both equity – based governance and relational embeddedness attenuate this effect. A Wald test reveals that formal governance exerts a stronger moderating effect than relational governance. Results are robust across multiple identification strategies. Additionally, technological invasion significantly reduces the partner's market value while increasing the focal firm's market value, confirming its opportunistic nature.

Key Words: strategic alliances; AI capability gap; technological invasion; transaction cost economics; governance mechanisms

Purpose

Strategic alliances allow firms to pool resources, explore new technologies, and strengthen competitive positions. However, alliances also expose partners to opportunistic behavior, particularly technological invasion, a process in which one firm leverages knowledge acquired through collaboration to file patents in its partner's core technology domains, thereby eroding the partner's competitive advantage. Prior research has examined how traditional inter – firm asymmetries, such as differences in firm size, network centrality, and organizational status, contribute to opportunistic behavior in alliances. Yet an increasingly important source of asymmetry remains largely unexplored: the gap in artificial intelligence (AI) capabilities between alliance partners. AI technologies are fundamentally changing how firms identify, acquire, and process information, potentially reshaping the information structure within alliances and creating new governance challenges. Although a few studies have theoretically suggested that superior AI capabilities could be misused to appropriate partners' interests, systematic empirical evidence on whether and how AI capability gaps affect opportunistic behavior in interorganizational relationships is still lacking. Addressing this gap is critical because, as AI adoption accelerates unevenly across firms, understanding the governance risks arising from such capability disparities is essential for both alliance theory and managerial practice.

Research Design

This study draws on transaction cost economics (TCE) to build its theoretical framework. TCE argues that information asymmetry between transacting parties increases the likelihood of opportunistic behavior, as the better – informed party can exploit its advantage for self – interest. We define AI capability gap as the relative difference between a focal firm and its alliance partner

in using AI technologies to identify, acquire, and integrate external information, and argue that this gap constitutes a novel source of information asymmetry that heightens the risk of technological invasion. We further examine two governance mechanisms that may mitigate this risk: alliance type (equity – based versus non – equity – based), representing formal governance, and relational embeddedness (measured by the frequency of repeated cooperation), representing relational governance. We construct a comprehensive alliance database of Chinese A – share listed firms from 2010 to 2023. Alliance data are manually collected from public announcements on the CNINFO website and cross – verified with the Wind database. AI capability is measured using firms' independently filed AI – related patents, classified according to the system issued by China's National Intellectual Property Administration. Technological invasion is operationalized as whether the focal firm files patents for the first time in the partner's pre – alliance dominant technology domains, excluding joint patents. The final sample includes 10531 partner – focal firm – year observations. We employ fixed – effects panel regression models controlling for industry and year effects, with standard errors clustered at the firm level.

Findings

Our results confirm all three hypotheses. First, a larger AI capability gap between partners significantly increases the probability of technological invasion by the focal firm. Second, equity – based alliances significantly weaken this positive relationship through two complementary mechanisms: an information mechanism that enhances transparency and monitoring, and a cost mechanism that amplifies penalties for detected opportunism. Third, higher relational embeddedness also attenuates the effect of AI capability gaps on technological invasion by increasing behavioral predictability and reinforcing trust – based constraints. A comparative analysis using the Wald test further reveals that equity – based alliances exert a significantly stronger moderating effect than relational embeddedness, indicating that formal governance is more effective than informal governance in curbing AI – driven opportunism. These findings are robust across multiple specifications, including instrumental variable estimation, propensity score matching, alternative variable measures, exclusion of AI – intensive industries, Poisson pseudo – maximum likelihood estimation, and alternative alliance duration windows. Additionally, technological invasion is found to significantly reduce the partner's market value while increasing the focal firm's market value, confirming its opportunistic nature.

Value

This study makes three contributions. First, it introduces AI capability gaps as a novel form of inter – firm asymmetry into the alliance opportunism literature, revealing a distinctive information processing asymmetry that differs fundamentally from traditional resource or size disparities. Second, it advances understanding of alliance governance effectiveness by empirically comparing formal and relational governance mechanisms under AI – driven information asymmetry, responding to ongoing debates about their situational effectiveness. Third, it enriches AI governance research by shifting the focus from macro – level policy discussions to micro – level interorganizational dynamics.

Implications and Future Research

For practitioners, our findings suggest that firms should incorporate AI capability matching into partner selection and favor equity – based structures when capability gaps are substantial. Cultivating relational embeddedness through repeated collaboration serves as a valuable complementary safeguard. For policymakers, promoting shared digital infrastructure and neutral collaboration platforms can reduce structural risks arising from AI capability disparities. This study has limitations. First, patent – based

measures may not fully capture AI capabilities; future studies could incorporate indicators such as AI talent or computational infrastructure. Second, the sample is limited to Chinese listed firms; cross – national comparisons would enhance generalizability. Third, while we emphasize the trust – constraining role of relational embeddedness, its potential dark side, whereby deeper collaboration increases technological exposure, merits further exploration under conditions of alliance termination or intensified competition.