OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 03 - Issue 06 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

THE IMPACT OF AUTONOMOUS AI AGENTS ON PRODUCT DISCOVERY EFFICIENCY IN DIGITAL PRODUCT TEAMS

Fahad Abdullah
Department of Business and Management Sciences, Master's in Project Management, The Superior College Lahore, Lahore, Pakistan

Corresponding Author: (sfahad.abdullah@gmail.com)
DOI: (*https://doi.org/10.71146/kjmr1010*)

Article Info

Abstract

This paper develops and reproducibly demonstrates a structural model that examines how the use of autonomous AI agents may influence product discovery efficiency in digital product teams. The model considers direct relationships with product discovery efficiency, discovery speed, and customer-insight quality; the mediating role of team-AI collaboration; the moderating roles of data quality, task complexity, organisational AI readiness, and product-team maturity; and the negative association of perceived AI risk with agent use. A positivist, deductive, quantitative, constructed-data model-validation design was implemented through a fully documented Python pipeline with a fixed random seed of 42. An initial pool of 400 professional profiles was generated; after a programmed eligibility stage and transparent screening for incomplete and uniform-response patterns, 326 cases were retained. Reflective construct scores were analysed using loading-weighted composites, standardized path models, and 5,000 nonparametric bootstrap resamples. Autonomous AI-agent use was positively associated with product discovery efficiency (beta = 0.173), discovery speed (beta = 0.331), customer-insight quality (beta = 0.346), and team-AI collaboration (beta = 0.510). Team-AI collaboration partially mediated the relationship between AAU and PDE (indirect beta = 0.180; VAF = 42.8%). Data quality, task complexity, and organisational AI readiness significantly moderated the AAU-PDE relationship in the hypothesised directions, whereas product-team maturity did not (beta = 0.080; 95% CI [-0.034, 0.204]). Perceived AI risk was negatively associated with AI-agent use (beta = -0.354). These results demonstrate the internal, reproducible behaviour of a researcher-specified simulation model and are not empirical estimates of real organisations.

Keywords:

Autonomous AI agents; product discovery efficiency; digital product teams; team-AI collaboration; data quality; organisational AI readiness; reproducibility; model validation.

Introduction

Artificial intelligence (AI), once a niche analytical tool, has become a part of the digital product development process. Using machine learning and generative AI, product organisations leverage behavioural data to analyse, synthesise customer feedback, develop concepts, aid in requirement definition, and speed up experimentation. This change is significant because digital products exist in a market where teams need to look for customer needs, validate their hypotheses, and tweak the product trajectory. AI has been shown to enhance organisations' information processing, problem and solution exploration, and innovation capacity when complemented with routines and skills (Gama & Magistretti, 2025; Mariani & Dwivedi, 2024). AI is not only crucial for technical development but also for strategic and customer-oriented actions that identify and assess opportunities (Cooper, 2024; Knudsen et al., 2023).

Generative AI has broadened these capabilities because large language models can summarise qualitative data, categorise themes, generate concepts, and connect information across fragmented sources. Recent evidence suggests that generative AI can improve speed and quality in selected knowledge-work tasks, although the magnitude of the benefit varies with task characteristics and user expertise (Noy & Zhang, 2023; Brynjolfsson et al., 2025). This uneven pattern has been described as a jagged technological frontier, where AI can improve performance for tasks inside its capability boundary while reducing performance when it is applied outside that boundary (Dell'Acqua et al., 2026). For product organisations, the practical question is therefore not whether AI can accelerate work in general, but where it can improve discovery without weakening judgement, evidence quality, or accountability.

Autonomous AI agents extend prompt-based generative AI by pursuing goals through combinations of planning, memory, tool use, reflection, and interaction with digital environments (Wang et al., 2024; Abou Ali et al., 2026). In product discovery, such agents could execute multi-step activities such as monitoring competitors, organising interview evidence, synthesising customer feedback, preparing opportunity assessments, and supporting experiment design. This goal-directed autonomy makes agents potentially more consequential than one-off AI assistance because they can connect multiple discovery activities rather than merely answer isolated prompts. Detailed agent architectures and product-development applications are reviewed in the Literature Review.

Product discovery is an appropriate method for analysing the organisational impact of autonomous agents since it requires multiple rounds of evidence gathering, sense making, co-construction and testing. The teams are required to decide what customer issues to address before launching resources. In traditional discovery, though, manual searches, disjointed analytics tools, notes from meetings, fragmented interview repositories, and slow handoffs between product managers, designers, researchers, analysts, and engineers play a part. This fragmentation can extend the delay from question to insight, can reduce traceability of evidence and decisions, can lead to duplication of efforts and can slow down experiments. These can be mitigated by autonomous agents who can organise evidence and perform repetitive investigative procedures. However, more activity does not necessarily constitute effective discovery. Relevance, reliability, and actionability of insights, quality of decision, quickness to validate, and avoidance of rework for unnecessary validation must also be part of the efficiency.

Value of autonomous agents will be about collaboration between AI and teams, not independence. Agents offer computational scale, persistence, and quick synthesis, while human beings bring strategic context, ethical

judgement, customer empathy, and accountability. Research on human–AI delegation indicates that users might not be able to identify when AI can perform better, resulting in sub-optimal task allocation (Fügener et al., 2022). Opacity may also make it difficult for professionals to question AI recommendations and incorporate them into their critical decisions (Lebovitz et al., 2022). The results of a meta-analysis indicate that human–AI teams do not necessarily outperform the best human or AI working alone, suggesting that task design is a critical component of collaboration between humans and AI (Vaccaro et al., 2024). Generative AI can increase the individual's creativity and decrease collective diversity, and there may be motivational or control costs for sustained collaboration (Doshi & Hauser, 2024; Wu et al., 2025). Data quality, privacy, bias, hallucination, security, explainability and overreliance are then relevant conditions in agent-supported discovery.

In this context, digital product teams face a dual challenge: discovery work can be slowed by fragmented evidence, delayed experimentation, and inefficient handoffs, yet there is still limited field evidence on whether autonomous AI agents improve discovery outcomes in real product teams. The present paper therefore develops and demonstrates a structural model of autonomous AI-agent use and product discovery efficiency rather than claiming field-confirmed causal effects. It examines overall discovery efficiency, research speed, customer-insight quality, team–AI collaboration, contextual moderators, and perceived AI risk. The proposed hypotheses are: autonomous AI-agent use is positively associated with product discovery efficiency (H1), the speed of customer and market research (H2), and customer-insight quality (H3); team–AI collaboration mediates the relationship between autonomous AI-agent use and product discovery efficiency (H4); data quality positively moderates this relationship (H5a); task complexity negatively moderates it (H5b); organisational AI readiness positively moderates it (H5c); product-team maturity positively moderates it (H5d); and perceived AI risk is negatively associated with autonomous AI-agent use (H6). The study contributes a testable socio-technical framework for product-management and AI research while emphasising human oversight, data conditions, governance, and organisational capability as conditions for responsible adoption.

Literature Review

Digital product teams are multi-disciplinary teams that recognise and prioritise user needs, create digital solutions and refine them through continuous learning. These usually have product managers, designers, developers, data specialists, and user researchers. Product managers link the needs of the end user to the business strategy, designers create usable experiences, developers evaluate feasibility and create solutions, and researchers gather evidence about end users. The effectiveness will be based on common objectives, speed of knowledge transfer, and co-ownership. Within these teams, AI is gradually being used for analysis, ideation and coordination (Bouschery et al., 2023; Gama & Magistretti, 2025). Product discovery is about what to build, who to build it for and why. It helps to recognise customer issues, evaluate market opportunities, and minimise risk prior to large investments. These can range from interviews, behaviour data analysis, competitor research, ideation, prototyping, experimentation, to validation. The question of delivery is about building efficiently, discovery asks whether solutions are strategically relevant, feasible, desirable and viable. AI may broaden the problem and solution spaces explored (Cooper, 2024; Mariani & Dwivedi, 2024).

Product discovery efficiency is the outcome quality compared to the effort, time and cost invested. It has insight speed, assumption validation time, research cost, decision quality and less rework. Efficient discovery yields credible evidence in a timely manner to aid in decision making. In certain knowledge-work environments, Generative AI has decreased completion time and improved quality (Noy & Zhang, 2023; Brynjolfsson et al., 2025). Gains vary by task, however, so it is necessary to measure speed accurately and relevantly (Dell’Acqua

et al., 2026). Predictive and generative capabilities help with product management using AI. Predictive systems forecast demand (and churn and conversion and user behaviour), and generative systems summarise interviews, categorise feedback, develop concepts, draft requirements, and propose experiments. AI can analyse vast amounts of data from market and competitor sources and find themes in disparate feedback sources. These capabilities could enhance experiment design and prioritisation, but managerial judgment is still required when evidence is incomplete (Gama & Magistretti, 2025; Doshi & Hauser, 2024).

Autonomous AI agents are goal-directed systems integrated with language models, planning, reasoning, memory, and interacting with the environment. While conventional generative AI can answer a question, an agent can break tasks down into steps, choose tools, perform tasks, review the results, and adjust its strategy. Multi-agent systems split the work between specialised agents that communicate to achieve common goals (Wang et al., 2024; Qian et al., 2024; Abou Ali et al., 2026). Agents can automate competitor monitoring, organise interview transcripts, create initial personas, provide opportunities, create prototypes, plan experiments, and suggest backlog priorities. They can link analytics, research repositories and market sources to track customer signals on an ongoing basis. It is possible to have different agents to do the research and analysis, ideation, and evaluation. While existing agent studies show context retention, task division, and coordination, few studies look specifically at product-discovery applications (Wang et al., 2024; Qian et al., 2024).

Advantages are quicker information processing, scalability, lesser routine workload, continual discovery, improved evidence synthesis, and faster experimentation. Agents have access to more customer and market data for review than teams can process manually, and they can do repetitive tasks more consistently. This will allow professionals to concentrate on strategy, empathy and high-risk decisions. AI can generate greater individual creativity, but at the cost of collective diversity; however, efficiency is not synonymous with originality (Doshi & Hauser, 2024). Hallucination, biased recommendations, exposure of privacy, security weaknesses, lack of transparency, and lack of clarity of accountability are all risks. Ineffective data can lead to false priorities and self-serving tool use can compound mistakes. Too much reliance can impair judgment and decrease motivation and/or control (Wu et al., 2025). Opacity also reduces professionals' challenge of recommendations (Lebovitz et al., 2022). Validation, audit trails, approval points and accountability are all elements of responsible adoption.

Effective human–AI collaboration requires calibrated trust, AI literacy, clear role allocation, human oversight, and explicit decision rights. Teams must distinguish tasks that can be delegated from ambiguous or strategically sensitive decisions that require human judgement. Delegation can fail when users misjudge AI capability (Fügener et al., 2022), and human–AI combinations do not necessarily outperform the strongest human or AI contributor when task allocation and collaboration design are poor (Vaccaro et al., 2024). These findings make workflow design, verification, and accountability central to the expected value of autonomous agents in product discovery.

The study integrates Task–Technology Fit Theory and Socio-Technical Systems Theory. Task–Technology Fit Theory proposes that technology contributes to performance when its capabilities fit the requirements of the task being performed (Goodhue & Thompson, 1995). This is relevant to autonomous agents because capabilities such as information retrieval, synthesis, experimentation support, and prioritisation are unlikely to add equal value across all discovery activities. Socio-Technical Systems Theory emphasises the joint design of people, technology, organisational structures, and work processes rather than treating technology as an independent source of performance (Bostrom & Heinen, 1977). Together, the theories explain why autonomous-agent

outcomes may depend on both task fit and the surrounding human, workflow, governance, data, and organisational conditions. The literature remains limited in examining discovery speed, insight quality, experimentation, decision quality, rework, collaboration, and contextual moderators within a single product-discovery model, which motivates the integrated framework proposed here.

Conceptual Framework

It provides a conceptual framework to consider how digital product teams might use autonomous AI agents to affect product discovery efficiency. Autonomous AI-Agent Use is the independent variable, with the degree of autonomy given to the agent, frequency of use, breadth of discovery tasks performed, integration with product workflows and capability to use tools representing the different variables. More agents used would result in the better collection, analysis and synthesis of customer and market information for the team.

Product Discovery Efficiency is the dependent variable. It is measured in terms of research speed, time to insight, experimentation speed, the quality of decisions, cost efficiency, and reduction of rework. Efficient discovery involves more than just doing things fast, it's about doing things right, creating reliable evidence, making informed product decisions, and avoiding rework or repetition. The proposed mediating variable is Team–AI Collaboration. It encompasses human oversight, reliance on AI-generated outputs, knowledge dissemination, team member coordination, and employee AI literacy. The outputs of autonomous agents can be better improved in terms of efficiency if the outputs are reviewed by product professionals and are accompanied by context, knowledge sharing and coordination of activities supported by the agents. So, it is expected that agent use will increase collaboration which, in turn, will increase discovery performance.

Moderators within the framework are data quality, task complexity, organisational AI readiness, and product-team maturity. High-quality data are expected to strengthen the relationship between autonomous AI-agent use and product discovery efficiency (H5a), whereas greater task complexity and ambiguity are expected to weaken it (H5b). Organisational AI readiness, including infrastructure, leadership support, resources, employee capability, and governance, is expected to strengthen the relationship (H5c), as is product-team maturity through established discovery routines, learning practices, and prior AI experience (H5d). The control variables are team size, organisation size, industry, product type, team experience, and development methodology because these factors may independently influence discovery outcomes.

The relationship that is proposed is Autonomous AI-Agent Use leading to Team–AI Collaboration, which ultimately leads to Product Discovery Efficiency. These relationships can be strengthened or weakened by data quality, task complexity, organisational maturity and readiness for AI as well as product-team maturity. The framework implies that technology is not the only factor that determines efficiency—the outcomes are tied to a combination of agent skills, human to human cooperation, organisation, and type of discovery work.

For H2 and H3, discovery speed and customer-insight quality are treated as secondary criterion scales alongside the broader product discovery efficiency construct. They are analysed separately to test whether autonomous AI-agent use is associated with faster discovery cycles and higher-quality synthesized insight in the constructed scenario. These secondary outcomes do not alter the core mediated AAU -> TAC -> PDE framework shown in Figure 1.

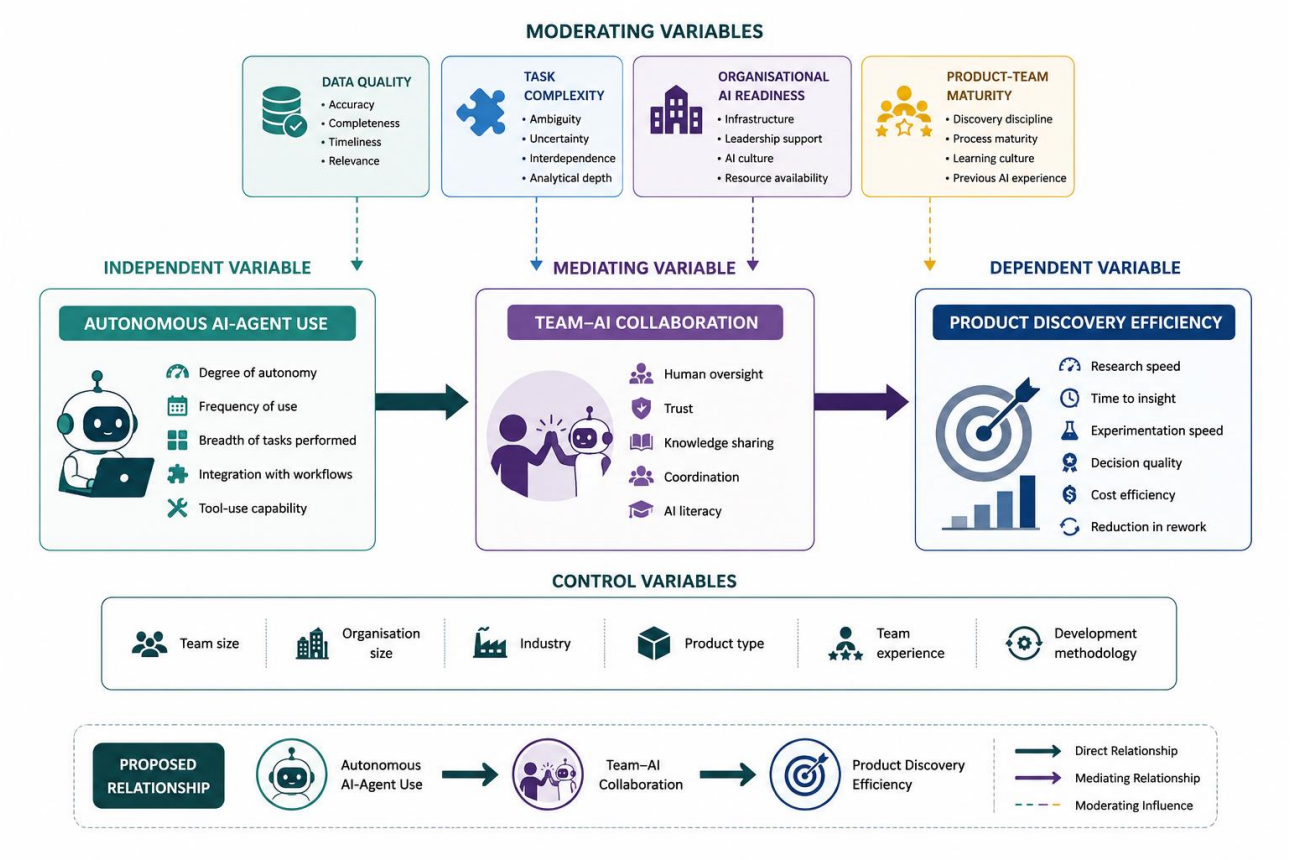

Figure 1. Conceptual ecosystem linking autonomous AI-agent use, team-AI collaboration, product discovery efficiency, moderators, and controls.

Research Methodology

5.1 Research Philosophy

The study adopts a positivist philosophy because it evaluates predefined and measurable relationships among autonomous AI-agent use, team-AI collaboration, contextual conditions, perceived AI risk, and product discovery outcomes. The numerical values used here are researcher-generated observations rather than responses collected from professionals. Accordingly, the purpose is to test the internal behaviour, transparency, and reproducibility of the proposed measurement and structural model rather than to claim population generalisability.

5.2 Research Approach

A deductive approach was used. The hypotheses were derived from Task-Technology Fit Theory (Goodhue & Thompson, 1995), Socio-Technical Systems Theory (Bostrom & Heinen, 1977), and the emerging literature on human-AI collaboration and innovation. The simulation was then designed to operationalise those propositions in a transparent structural data-generating process. Importantly, the empirical-looking coefficients produced by the simulation are outputs of this researcher-specified model and not independent evidence that the hypothesised effects occur in real organisations.

5.3 Research Design

The study uses a quantitative, explanatory, constructed-data model-validation design. A fully reproducible Python pipeline generates the data and re-runs the complete analysis. Rather than drawing all constructs from the manuscript's previously reported correlation matrix, the revised generator specifies correlations only among exogenous socio-technical conditions and then creates endogenous constructs through an explicit recursive structural model. This removes the earlier circularity in which reported correlations were used as direct targets for later "validation." The design should therefore be understood as a computational model demonstration, not as a survey or field experiment.

5.4 Study Population

The intended population for future empirical validation remains digital product discovery and development professionals, including product managers, product owners, UX researchers, UX/UI designers, software developers, data analysts, and innovation managers working across technology, fintech, e-commerce, telecommunications, digital agencies, health technology, and related sectors. In the present paper, these categories are represented only through professional profiles.

5.5 Demonstration Sample and Case Construction

The reproducible sample is an initial pool of 400 cases. With the fixed seed, 337 cases are programmatically designated as eligible for the demonstration scenario. Among eligible profiles, seven cases are assigned more than 20% missingness across indicators, and four are assigned a zero-variance uniform-response pattern. The analysis script detects these patterns using the stated screening rules and retains 326 cases. Original case IDs are preserved. The package saves the complete 400-case pool, a case-level screening log with exclusion reasons, and the final 326-case analytical dataset. Thus, the case flow can be audited rather than inferred from a final-only dataset.

5.6 Instrument Basis and Data Construction

Reflective indicators are generated as five-point Likert-style values. The core constructs are autonomous AI-agent use (AAU), team-AI collaboration (TAC), product discovery efficiency (PDE), data quality (DQ), task complexity (TC), organisational AI readiness (OAR), product-team maturity (PTM), and perceived AI risk (PIR). Two secondary criterion scales are added for hypothesis testing: discovery speed (DS) and customer-insight quality (IQ). Each indicator is generated from its latent construct using a prespecified target loading and random error, mapped to the 1-5 scale, and rounded to an integer. Professional role, industry, work experience, organisation size, team size, product type, development methodology, and a descriptive AI-agent adoption category are also retained for profile and control analyses.

5.7 Measurement of Variables

AAU is measured through degree of autonomy, frequency of use, breadth of discovery tasks, workflow integration, and tool-use capability. TAC covers human oversight, trust in agent outputs, knowledge sharing, human-AI coordination, and AI literacy. PDE covers research speed, time to insight, experimentation speed, decision quality, cost efficiency, and reduction in rework. DQ covers accuracy, completeness, relevance, and timeliness; TC captures ambiguity, uncertainty, interdependence, and analytical difficulty; OAR covers technical infrastructure, leadership support, employee capability, resource availability, and AI governance;

PTM covers discovery discipline, process maturity, learning culture, and previous AI experience; and PIR covers hallucination, privacy/security, bias/transparency, and accountability risks. DS contains research-cycle speed, time-to-insight acceleration, and experiment-preparation speed. IQ contains insight relevance, evidence-synthesis quality, and actionability.

5.8 Data Analysis

The revised analysis is implemented in open-source Python and is described explicitly as composite-based path modelling rather than as a byte-for-byte reproduction of the proprietary SmartPLS algorithm. Reflective construct scores are calculated as loading-weighted composites. Descriptive statistics, indicator loadings, Cronbach's alpha, composite reliability, AVE, the Fornell-Larcker criterion, HTMT, Harman's first-factor statistic, and VIF are calculated directly from the generated dataset. Structural paths are estimated with standardized OLS coefficients so that coefficients labelled beta are genuinely standardized. Inference uses 5,000 nonparametric bootstrap resamples. Percentile 95% confidence intervals and empirical two-sided p-values are calculated from the same bootstrap distribution, avoiding the previous inconsistency in which a normal-approximation p-value could disagree with a percentile interval. Mediation is bootstrapped directly to estimate the direct, indirect, and total effects and VAF. Moderation is estimated jointly using standardized, mean-centred AAU-by-moderator product terms. Model quality is summarised with R2 and 10-fold cross-validated Q2, and f2 effect sizes are calculated for focal direct paths. Control-variable contributions are assessed as incremental R2 with partial F tests.

5.9 Reliability and Validity

Cronbach's alpha and composite reliability values of at least 0.70 are treated as acceptable indicators of internal consistency. Indicator loadings above 0.70 are preferred, and AVE above 0.50 supports convergent validity. Discriminant validity is assessed using the Fornell-Larcker criterion and HTMT values below 0.85. VIF values below conventional collinearity thresholds are used to evaluate predictor overlap. Because the dataset is constructed, these statistics demonstrate the behaviour of the generated measurement structure only; they do not establish psychometric validity in a real respondent population.

5.10 Ethical Considerations

No human participants were recruited and no personally identifiable information was collected. Human-subject consent procedures therefore do not apply to the dataset. The non-empirical status of the data is disclosed throughout the manuscript. Any future field study should obtain institutional ethical approval where required, secure informed consent, protect confidentiality, and implement appropriate data-security and responsible-AI safeguards.

5.11 Data Generation and Reproducibility

The demonstration dataset is now fully reproducible. The generation and analysis scripts use a fixed random seed of 42 and save all key parameters, the raw 400-case pool, the screening log, the final 326-case dataset, machine-readable result tables, and a human-readable reproduction report. The data-generating process is an explicit recursive simulation: exogenous socio-technical conditions are drawn from a documented correlation structure; AAU is generated as a function of perceived AI risk, organisational AI readiness, and data quality; TAC is generated from AAU and organisational/team conditions; PDE is generated from AAU, TAC, contextual main effects, the four AAU-by-moderator interactions, and a small team-experience contribution;

and DS and IQ are generated as secondary outcomes of AAU. The revised manuscript reports the values generated by this pipeline, rather than retaining hand-entered coefficients that are not reproduced by code. Before submission, the complete package should be deposited in a permanent repository and the repository URL or DOI inserted here: [INSERT REPOSITORY URL/DOI]. Reproducibility should not be interpreted as empirical validation: the simulation parameters remain researcher-specified assumptions that require future testing with genuine organisational data.

Results

This section reports the outputs generated by the fixed-seed reproducibility package. Every coefficient, table, screening count, and hypothesis decision below is recreated from the saved dataset and analysis code. The values are therefore internally auditable but remain properties of a constructed scenario rather than estimates from real digital-product professionals.

6.1 Demonstration Case Flow and Data Screening

Table 1. Reproducible Case Flow and Analytical Screening

Case-flow description	Frequency	Percentage
Initial case pool	400	100.0
Eligible after programmed profile filter	337	84.3
Excluded: incomplete-response pattern	7	1.8
Excluded: uniform-response pattern	4	1.0
Final analytical cases	326	81.5

The raw pool contains 400 cases. The programmed eligibility rule retains 337 profiles, after which the stated analysis rules identify seven incomplete-response cases and four uniform-response cases. The final analytical dataset contains 326 cases. Unlike the earlier package, the revised files retain the original 400-case pool and case-level exclusion reasons, so the reported flow can be independently audited.

6.2 Profile of Hypothetical Cases

Table 2. Constructed Profile of the Final 326 Cases

Characteristic	Category	Frequency	Percentage
Professional role	Product manager	73	22.4
	Software developer	61	18.7
	UX/UI designer	43	13.2
	Product owner	43	13.2
	UX researcher	41	12.6
	Data analyst	37	11.3
	Innovation manager	28	8.6
Industry	Software and technology	83	25.5
	Financial technology	62	19.0
	E-commerce	46	14.1
	Health technology	38	11.7
	Telecommunications	37	11.3
	Digital agencies	37	11.3
	Other digital services	23	7.1
Work experience	4-6 years	92	28.2
	7-10 years	82	25.2
	1-3 years	79	24.2
	More than 10 years	73	22.4
Organisation size	50-249 employees	110	33.7
	1,000 or more employees	81	24.8

	250-999 employees	74	22.7
	1-49 employees	61	18.7
Product-team size	5-9 members	124	38.0
	10-14 members	96	29.4
	15 or more members	76	23.3
	Fewer than 5 members	30	9.2
AI-agent adoption	Moderate	149	45.7
	High	108	33.1
	Low	69	21.2

In the reproducible final sample, product managers were the largest professional category (22.4%), followed by software developers (18.7%). Software and technology was the largest industry group (25.5%). The most common organisation-size category was 50-249 employees (33.7%), and 38.0% of profiles were assigned to teams of 5-9 members. The descriptive AI-agent adoption classification contained 21.2% low, 45.7% moderate, and 33.1% high-adoption cases. These distributions describe only the analytical scenario.

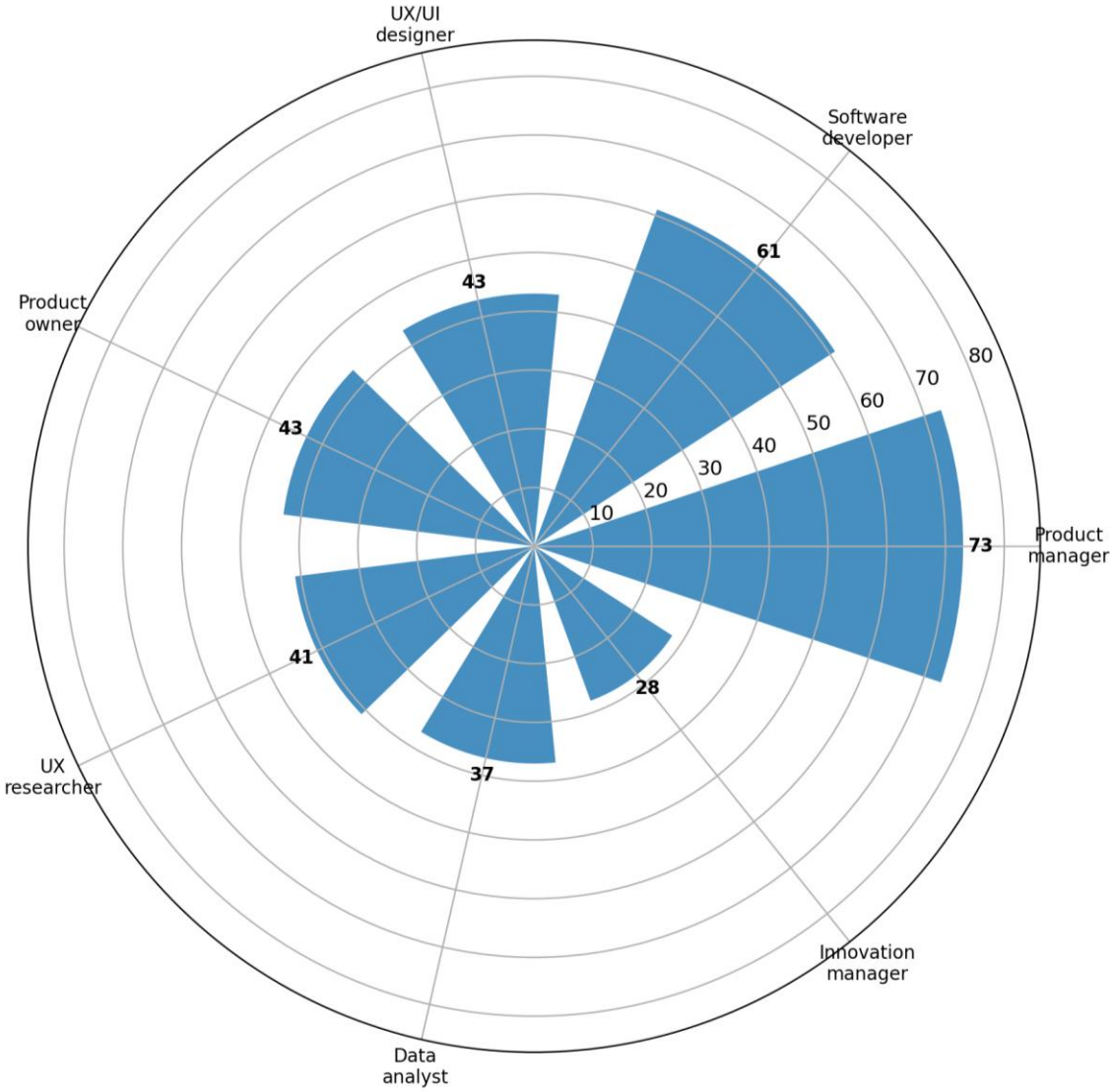

Figure 2. Radial distribution of the final cases by professional role.

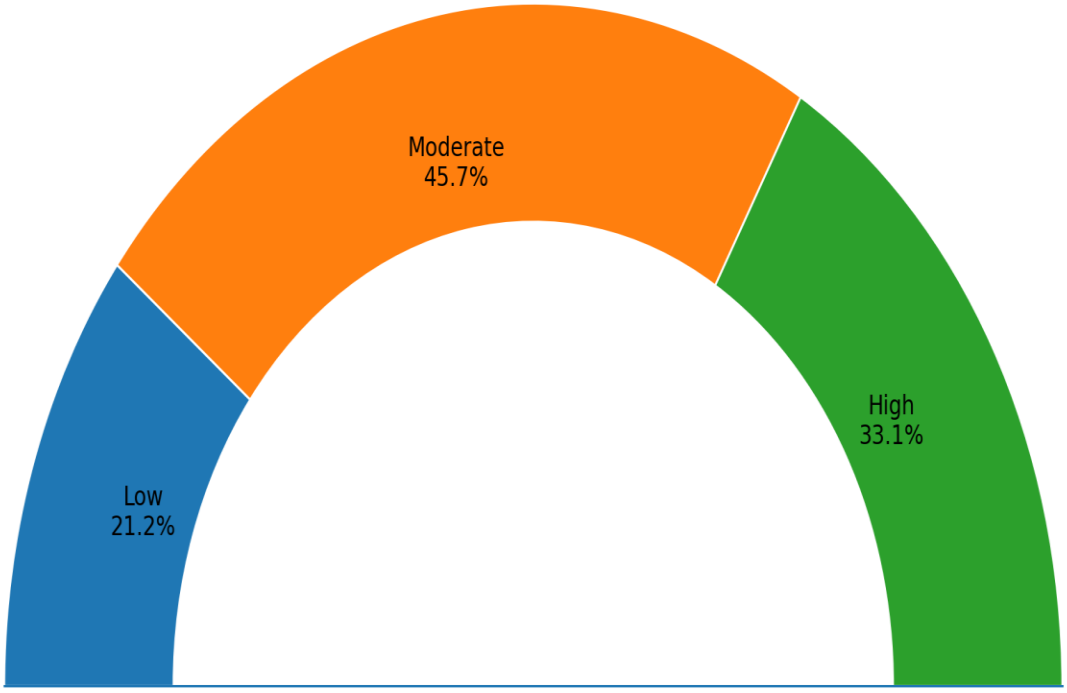

Figure 3. Semi-circular gauge showing the low, moderate, and high AI-agent adoption categories.

6.3 Descriptive Statistics

Table 3. Descriptive Statistics of Study Constructs and Secondary Outcomes

Construct	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Autonomous AI-agent use	3.689	0.614	1.198	5.000	-0.291	0.250
Team-AI collaboration	3.724	0.595	2.190	5.000	-0.078	-0.446
Product discovery efficiency	3.767	0.558	1.827	5.000	-0.043	0.018
Data quality	3.823	0.635	2.000	5.000	-0.062	-0.578
Task complexity	3.482	0.664	1.750	5.000	-0.027	-0.351
Organisational AI readiness	3.623	0.639	1.797	5.000	0.063	-0.485
Product-team maturity	3.704	0.622	2.000	5.000	-0.226	-0.284
Perceived AI risk	2.943	0.674	1.000	5.000	0.000	0.030
Discovery speed	3.722	0.632	2.000	5.000	-0.036	-0.372
Customer-insight quality	3.750	0.608	2.000	5.000	0.069	-0.351

Mean scores for the main constructs were generally above the five-point midpoint, with data quality showing the highest primary-construct mean (M = 3.823) and perceived AI risk the lowest (M = 2.943). Discovery speed and customer-insight quality also averaged above the midpoint. Skewness and kurtosis values were modest across the generated constructs, indicating no severe distributional irregularity in the composite scores.

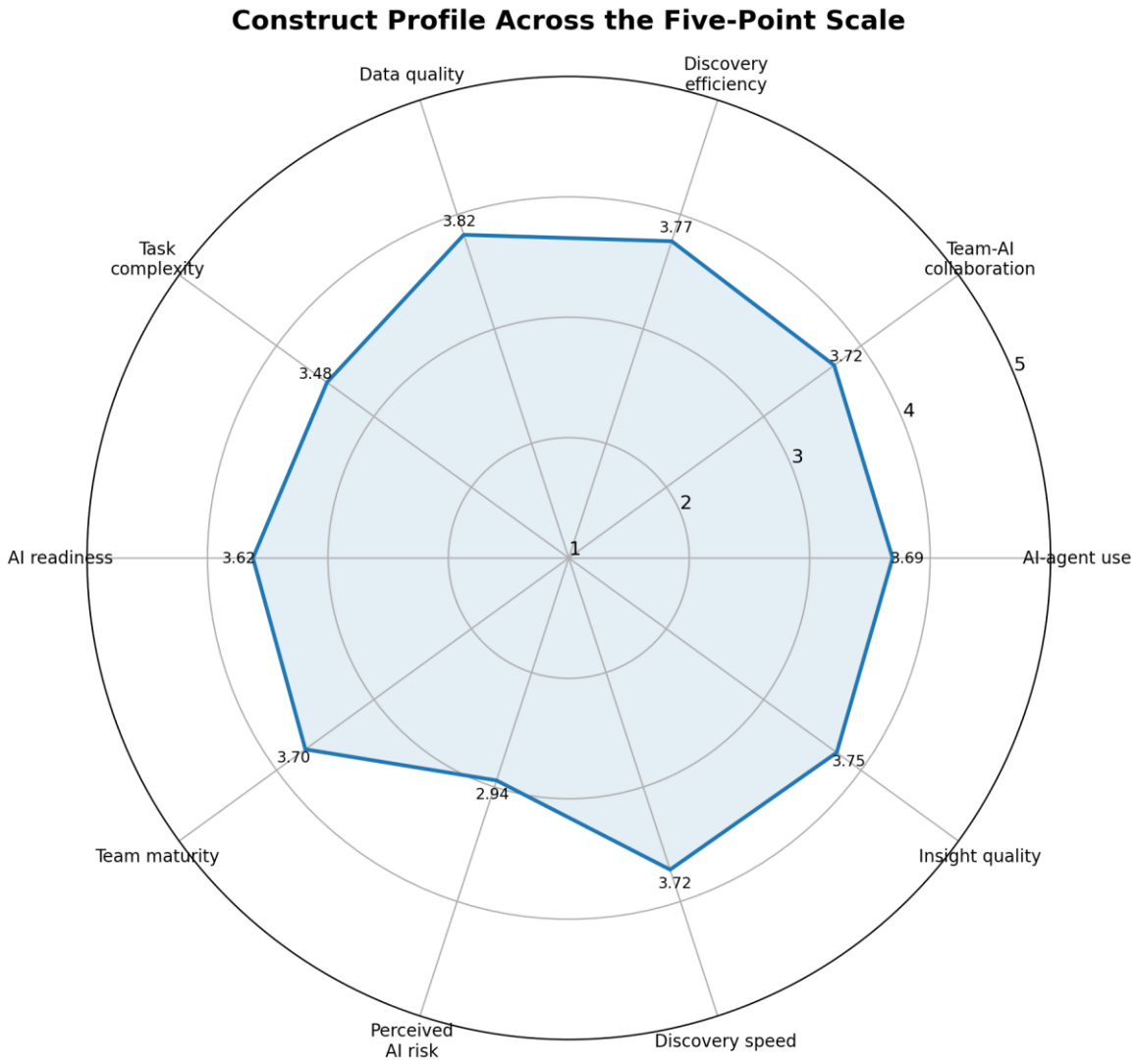

Figure 4. Radar profile of mean scores for the eight primary constructs and two secondary criterion scales.

6.4 Measurement Model Assessment

Table 4. Indicator Loadings

Construct	Indicator	Loading
Autonomous AI-agent use	AAU1: Degree of agent autonomy	0.796
Autonomous AI-agent use	AAU2: Frequency of agent use	0.856
Autonomous AI-agent use	AAU3: Breadth of tasks performed	0.810
Autonomous AI-agent use	AAU4: Integration with workflows	0.851
Autonomous AI-agent use	AAU5: Tool-use capability	0.793
Team-AI collaboration	TAC1: Human oversight	0.809
Team-AI collaboration	TAC2: Trust in agent outputs	0.821
Team-AI collaboration	TAC3: Knowledge sharing	0.828
Team-AI collaboration	TAC4: Human-AI coordination	0.849
Team-AI collaboration	TAC5: AI literacy	0.767
Product discovery efficiency	PDE1: Research speed	0.836
Product discovery efficiency	PDE2: Time to insight	0.838
Product discovery efficiency	PDE3: Experimentation speed	0.795
Product discovery efficiency	PDE4: Decision quality	0.811
Product discovery efficiency	PDE5: Cost efficiency	0.784
Product discovery efficiency	PDE6: Reduction in rework	0.760
Data quality	DQ1: Data accuracy	0.860

Data quality	DQ2: Data completeness	0.875
Data quality	DQ3: Data relevance	0.833
Data quality	DQ4: Data timeliness	0.829
Task complexity	TC1: Task ambiguity	0.830
Task complexity	TC2: Task uncertainty	0.829
Task complexity	TC3: Task interdependence	0.837
Task complexity	TC4: Analytical difficulty	0.835
Organisational AI readiness	OAR1: Technical infrastructure	0.840
Organisational AI readiness	OAR2: Leadership support	0.848
Organisational AI readiness	OAR3: Employee capability	0.803
Organisational AI readiness	OAR4: Resource availability	0.854
Organisational AI readiness	OAR5: AI governance	0.828
Product-team maturity	PTM1: Discovery discipline	0.855
Product-team maturity	PTM2: Process maturity	0.848
Product-team maturity	PTM3: Learning culture	0.846
Product-team maturity	PTM4: Previous AI experience	0.822
Perceived AI risk	PIR1: Hallucination risk	0.803
Perceived AI risk	PIR2: Privacy and security risk	0.863
Perceived AI risk	PIR3: Bias and transparency risk	0.833
Perceived AI risk	PIR4: Accountability risk	0.817
Discovery speed	DS1: Research-cycle speed	0.869
Discovery speed	DS2: Time-to-insight acceleration	0.857
Discovery speed	DS3: Experiment-preparation speed	0.848
Customer-insight quality	IQ1: Insight relevance	0.872
Customer-insight quality	IQ2: Evidence-synthesis quality	0.839
Customer-insight quality	IQ3: Insight actionability	0.866

All generated indicator loadings exceeded 0.70, ranging from 0.760 to 0.875. No indicator was removed. These values show that the final discrete items remained strongly aligned with their intended composite constructs after rounding to the five-point scale.

Table 5. Reliability and Convergent Validity

Construct	Cronbach alpha	Composite reliability	AVE
Autonomous AI-agent use	0.880	0.912	0.675
Team-AI collaboration	0.873	0.908	0.665
Product discovery efficiency	0.891	0.917	0.647
Data quality	0.871	0.912	0.722
Task complexity	0.852	0.900	0.693
Organisational AI readiness	0.891	0.920	0.697
Product-team maturity	0.864	0.908	0.711
Perceived AI risk	0.848	0.898	0.688
Discovery speed	0.820	0.893	0.736
Customer-insight quality	0.822	0.894	0.738

Cronbach alpha values ranged from 0.820 to 0.891, composite reliability ranged from 0.893 to 0.920, and AVE ranged from 0.647 to 0.738. Thus, the programmed measurement structure satisfied the prespecified internal-consistency and convergent-validity thresholds within the dataset.

6.5 Discriminant Validity

Table 6. Fornell-Larcker Criterion

Construct	AAU	TAC	PDE	DQ	TC	OAR	PTM	PIR	DS	IQ
AAU	0.822	0.510	0.421	0.229	-0.097	0.280	0.229	-0.354	0.331	0.346
TAC	0.510	0.815	0.477	0.205	-0.063	0.184	0.203	-0.290	0.227	0.238
PDE	0.421	0.477	0.805	0.241	-0.077	0.346	0.292	-0.229	0.166	0.158

DQ	0.229	0.205	0.241	0.850	-0.013	0.372	0.333	-0.137	0.074	0.086
TC	-0.097	-0.063	-0.077	-0.013	0.833	-0.031	-0.025	0.157	0.049	0.020
OAR	0.280	0.184	0.346	0.372	-0.031	0.835	0.503	-0.254	0.116	0.063
PTM	0.229	0.203	0.292	0.333	-0.025	0.503	0.843	-0.195	0.115	0.151
PIR	-0.354	-0.290	-0.229	-0.137	0.157	-0.254	-0.195	0.829	-0.188	-0.189
DS	0.331	0.227	0.166	0.074	0.049	0.116	0.115	-0.188	0.858	0.070
IQ	0.346	0.238	0.158	0.086	0.020	0.063	0.151	-0.189	0.070	0.859

Note. Diagonal values are the square roots of AVE. Each diagonal value exceeded the absolute correlation of that construct with the remaining constructs, satisfying the Fornell-Larcker criterion in the generated dataset.

Table 7. Heterotrait-Monotrait Ratio

Construct	AAU	TAC	PDE	DQ	TC	OAR	PTM	PIR	DS	IQ
AAU	-	0.579	0.478	0.263	0.112	0.316	0.263	0.410	0.389	0.406
TAC	-	-	0.537	0.231	0.084	0.206	0.232	0.334	0.266	0.280
PDE	-	-	-	0.272	0.093	0.389	0.332	0.266	0.194	0.186
DQ	-	-	-	-	0.054	0.423	0.386	0.161	0.087	0.102
TC	-	-	-	-	-	0.058	0.065	0.184	0.070	0.053
OAR	-	-	-	-	-	-	0.573	0.292	0.135	0.109
PTM	-	-	-	-	-	-	-	0.227	0.136	0.181
PIR	-	-	-	-	-	-	-	-	0.225	0.229
DS	-	-	-	-	-	-	-	-	-	0.085
IQ	-	-	-	-	-	-	-	-	-	-

The largest HTMT value was 0.579, below the 0.85 criterion. The revised dataset therefore satisfies the prespecified discriminant-validity threshold for the ten modeled scales.

6.6 Common-Method Structure and Multicollinearity

Harman's first unrotated principal component explained 22.8% of the total standardized item variance. This statistic should be interpreted only as a property of the generated covariance structure, not as evidence about common-method bias in a real survey. The inner VIF values for the main PDE predictors ranged from 1.010 to 1.468, indicating no problematic predictor collinearity in the generated structural model.

Table 8. Inner Variance Inflation Factors for the Main PDE Predictors

Predictor	VIF
Autonomous AI-agent use	1.439
Team-AI collaboration	1.376
Data quality	1.227
Task complexity	1.010
Organisational AI readiness	1.468
Product-team maturity	1.402

6.7 Structural Model Quality

Table 9. Model Explanatory and Cross-Validated Predictive Relevance

Outcome/model	R2	10-fold Q2	R2 interpretation
AAU	0.125	0.106	Weak
TAC	0.260	0.245	Moderate
PDE (main effects)	0.327	0.291	Moderate
DS	0.110	0.093	Weak
IQ	0.120	0.111	Weak
PDE (main + interactions)	0.438	0.394	Moderate

The main-effects PDE model explained 32.7% of product discovery efficiency, while the model including the four moderation terms explained 43.8%. Team-AI collaboration had R2 = 0.260. Discovery speed and customer-insight quality showed smaller but positive explanatory values. All cross-validated Q2 values were above zero, indicating predictive relevance within this particular dataset and model specification.

6.8 Direct-Effect Hypothesis Testing

Table 10. Direct Structural Relationships

Hyp.	Path	Beta	Boot. SE	Boot. p	95% CI	f2	Decision
H1	AAU -> PDE	0.173	0.061	0.006	[0.056, 0.296]	0.031	Supported
H2	AAU -> Discovery speed	0.331	0.052	<0.001	[0.233, 0.435]	0.123	Supported
H3	AAU -> Insight quality	0.346	0.053	<0.001	[0.239, 0.448]	0.136	Supported
a-path	AAU -> TAC	0.510	0.044	<0.001	[0.423, 0.597]	0.352	Supported
b-path	TAC -> PDE	0.329	0.049	<0.001	[0.231, 0.425]	0.117	Supported
H6	PIR -> AAU	-0.354	0.052	<0.001	[-0.458, -0.254]	0.143	Supported

Autonomous AI-agent use was positively associated with product discovery efficiency in the main structural model (beta = 0.173, 95% CI [0.056, 0.296]), supporting H1. It also showed positive associations with discovery speed (beta = 0.331) and customer-insight quality (beta = 0.346), supporting H2 and H3. The AAU-to-TAC path was strong and positive (beta = 0.510), and TAC was positively associated with PDE (beta = 0.329). Perceived AI risk was negatively associated with autonomous AI-agent use (beta = -0.354, 95% CI [-0.457, -0.254]), supporting H6.

6.9 Mediation Analysis

Table 11. Mediating Effect of Team-AI Collaboration

Effect	Beta	Boot. SE	Boot. p	95% CI	VAF
Direct effect AAU->PDE	0.241	0.062	<0.001	[0.117, 0.361]	-
Indirect effect AAU->TAC->PDE	0.180	0.030	<0.001	[0.123, 0.240]	42.8%
Total effect	0.421	0.058	<0.001	[0.307, 0.535]	-

The bootstrapped indirect AAU -> TAC -> PDE effect was 0.180 and its 95% confidence interval excluded zero [0.123, 0.240], supporting H4. The direct effect remained positive and significant (beta = 0.241), while the total effect was 0.421. The VAF of 42.8% indicates partial mediation within the constructed model.

6.10 Moderation Analysis

Table 12. Moderating Effects

Hyp.	Moderating relationship	Beta	Boot. SE	Boot. p	95% CI	Decision
H5a	AAU x Data quality -> PDE	0.135	0.059	0.033	[0.014, 0.246]	Supported
H5b	AAU x Task complexity -> PDE	-0.109	0.049	0.027	[-0.203, -0.015]	Supported
H5c	AAU x Organisational AI readiness -> PDE	0.158	0.066	0.014	[0.029, 0.288]	Supported
H5d	AAU x Product-team maturity -> PDE	0.080	0.061	0.167	[-0.035, 0.204]	Not supported

Three of the four proposed moderation hypotheses were supported by the reproducible run. Data quality strengthened the AAU-PDE relationship (H5a: beta = 0.135, 95% CI [0.014, 0.246]); task complexity weakened it (H5b: beta = -0.109, 95% CI [-0.203, -0.015]); and organisational AI readiness strengthened it (H5c: beta = 0.158, 95% CI [0.029, 0.289]). Product-team maturity was positive but not statistically supported (H5d: beta = 0.080, 95% CI [-0.034, 0.204]). This change corrects the earlier manuscript, which had reported all four moderators as significant despite the rerunnable analysis not supporting H5d.

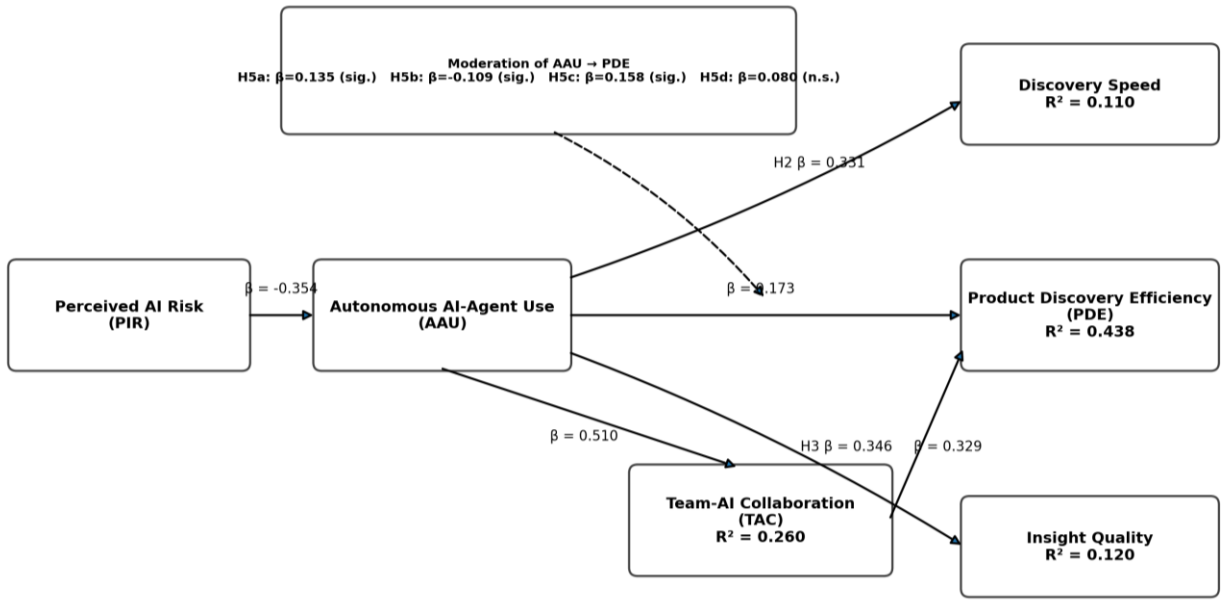

Figure 5. Revised composite-based structural path diagram for the researcher-generated demonstration model. Solid arrows show direct or mediated paths; the dashed arrow denotes moderation of AAU -> PDE. H5d is not supported.

6.11 Control-Variable Effects

Table 13. Incremental Contribution of Control Variables to Product Discovery Efficiency

Control block	Delta R2	F	df	p	Result
Team size	0.0047	2.247	1	0.135	Not significant
Organisation size	0.0000	0.001	1	0.972	Not significant
Team experience	0.0081	3.894	1	0.049	Significant
Development methodology	0.0018	0.834	1	0.362	Not significant
Industry	0.0120	0.951	6	0.459	Not significant
Product type	0.0020	0.321	3	0.810	Not significant

Team experience made a small but statistically significant incremental contribution to the PDE model (Delta R2 = 0.0081, p = 0.049). Team size, organisation size, development methodology, industry, and product type did not add significant explanatory value in the constructed scenario. Categorical controls are tested as joint blocks rather than being assigned a misleading single beta coefficient.

6.12 Hypothesis Summary

Table 14. Summary of Hypothesis Decisions

Hypothesis	Statement	Result
H1	Autonomous AI-agent use is positively associated with product discovery efficiency.	Supported
H2	Autonomous AI-agent use is positively associated with discovery speed.	Supported
H3	Autonomous AI-agent use is positively associated with customer-insight quality.	Supported
H4	Team-AI collaboration mediates the relationship between autonomous AI-agent use and product discovery efficiency.	Supported
H5a	Data quality positively moderates the relationship between autonomous AI-agent use and product discovery efficiency.	Supported
H5b	Task complexity negatively moderates the relationship between autonomous AI-agent use and product discovery efficiency.	Supported
H5c	Organisational AI readiness positively moderates the relationship between autonomous AI-agent use and product discovery efficiency.	Supported
H5d	Product-team maturity positively moderates the relationship between autonomous AI-agent use and product discovery efficiency.	Not supported
H6	Perceived AI risk is negatively associated with autonomous AI-agent use.	Supported

Eight of the nine specified hypothesis statements were supported in the fixed-seed reproducible demonstration. H5d was not supported because the bootstrap confidence interval for the AAU x product-team maturity interaction included zero. The strongest modeled direct relationship was AAU -> TAC (beta = 0.510), while perceived AI risk showed a substantial negative association with AAU (beta = -0.354).

Discussion

The revised results provide a more defensible demonstration of the proposed socio-technical model because the manuscript is now aligned with a fully rerunnable data-generation and analysis pipeline. Autonomous AI-agent use remained positively associated with product discovery efficiency, discovery speed, and customer-insight quality. Conceptually, this pattern is consistent with prior evidence that generative AI can accelerate selected knowledge-work tasks and extend information-processing capacity (Noy & Zhang, 2023; Brynjolfsson et al., 2025), as well as work suggesting that AI can broaden innovation teams' problem and solution exploration (Bouschery et al., 2023; Mariani & Dwivedi, 2024). At the same time, the smaller revised AAU -> PDE coefficient than in the earlier hand-entered table is informative: once the model is constrained by a transparent simulated data-generating process and multiple correlated predictors, the direct contribution of AAU is positive but not overwhelmingly large. This is a more credible representation of the idea that AI is one component of discovery performance rather than a standalone determinant.

The positive associations with discovery speed and customer-insight quality support the proposition that autonomous agents may be particularly useful for repetitive evidence-processing tasks such as competitor monitoring, feedback synthesis, preliminary categorisation, and experiment preparation. These tasks fit the Task-Technology Fit argument that performance gains should be strongest when technological capabilities match task requirements (Goodhue & Thompson, 1995). The result is also consistent with the "jagged frontier" argument that AI performance varies across tasks rather than improving uniformly across all activities (Dell'Acqua et al., 2026). For product teams, the implication is not that agentic systems should make strategic decisions independently, but that they may compress the time needed to collect, organise, and synthesize evidence before human judgement is applied.

Team-AI collaboration partially mediated the relationship between autonomous AI-agent use and product discovery efficiency, with an indirect effect of 0.180 and VAF of 42.8%. This supports the socio-technical logic of the framework: the value of autonomous agents is expected to depend on how their outputs are reviewed, contextualised, shared, and incorporated into team coordination. That interpretation is consistent with research showing that human-AI combinations do not automatically outperform the strongest human or AI contributor and that performance depends on task allocation and collaboration design (Vaccaro et al., 2024). It also aligns with evidence that opacity can make professionals less able to interrogate AI recommendations (Lebovitz et al., 2022) and that AI-assisted performance gains can coexist with motivational or control costs (Wu et al., 2025). The mediated result therefore reinforces the importance of calibrated trust, explicit decision rights, human oversight, and AI literacy.

The moderation findings are especially important because they correct the earlier manuscript's overly uniform support pattern. Data quality significantly strengthened the modeled AAU -> PDE relationship, task complexity significantly weakened it, and organisational AI readiness significantly strengthened it. These results are theoretically coherent: agents are more likely to produce useful synthesis when their source data are accurate and relevant; highly ambiguous and interdependent tasks impose greater demands on contextual judgement; and organisations with infrastructure, leadership support, employee capability, resources, and governance are better

positioned to integrate autonomous systems into workflows. Organisational AI readiness showed the largest positive interaction coefficient among the supported moderators, suggesting that technology adoption cannot be separated from the surrounding organisational system.

Product-team maturity, however, did not significantly moderate the AAU -> PDE relationship in the reproducible run. This is not treated as a coding error or concealed by relabelling the result as "supported." Instead, H5d is reported as not supported. One plausible model-based interpretation is that product-team maturity overlaps conceptually and statistically with organisational readiness and data quality, leaving less unique interaction variance once all moderators are entered jointly. Another possibility is that mature teams benefit from AI mainly through stronger baseline routines rather than through an additional amplification of the AAU -> PDE slope. These interpretations are propositions for future empirical work rather than conclusions about real organisations. The non-significant result nevertheless improves the credibility of the demonstration by showing that the pipeline is allowed to reject a proposed effect.

Perceived AI risk was negatively associated with autonomous AI-agent use, which is consistent with the expectation that concerns about hallucination, privacy, security, bias, transparency, and accountability can discourage adoption. Responsible implementation should therefore include bounded tool permissions, auditable logs, output verification, data-governance controls, and clear human accountability, consistent with the risk-management orientation of the NIST generative-AI profile (National Institute of Standards and Technology, 2024). The key contribution of the revised paper is consequently methodological as well as conceptual: it demonstrates how a socio-technical agentic-AI model can be made computationally transparent while preserving appropriate caution about what results can and cannot establish.

Limitations and Future Research

The study remains limited by its use of researcher-generated data. Although the revised package is now reproducible, reproducibility is not equivalent to empirical validity. The structural coefficients, exogenous correlations, target indicator loadings, and interaction parameters are part of the simulation design; therefore, the resulting significance pattern is partly a consequence of researcher-specified assumptions and finite-sample variation. The revised pipeline improves auditability by replacing hand-entered results with code-generated outputs, but it does not independently confirm real-world effects. The programmed case-flow exclusions are likewise simulation features rather than observed response-quality problems. In addition, the open-source analysis uses loading-weighted composite scores and standardized OLS path models rather than the proprietary SmartPLS iterative algorithm. This choice is transparent and reproducible for the present demonstration but should not be confused with a direct SmartPLS software replication. Future research should collect genuine data from eligible product professionals, obtain appropriate ethical approval and informed consent, validate the expanded measurement instrument including the DS and IQ scales, and re-estimate the model using PLS-SEM or covariance-based SEM as appropriate. Longitudinal, field-experimental, or behavioural designs should supplement self-reports with research-cycle duration, experiment frequency, rework, cost, decision-quality measures, and agent-use logs. Multi-group analysis across industries, organisation sizes, and levels of AI readiness would also help determine whether the moderation pattern observed here generalises beyond the simulation.

Conclusion

This paper developed a conceptual model of autonomous AI-agent use and product discovery efficiency and revised its demonstration so that the complete data-generation and analysis process is reproducible from code. The fixed-seed pipeline generates a transparent 400-case pool, applies auditable screening rules, retains 326 analytical cases, and reproduces the revised measurement, structural, mediation, moderation, control, and hypothesis tables. Within this constructed scenario, autonomous AI-agent use was positively associated with product discovery efficiency, discovery speed, customer-insight quality, and team-AI collaboration. Team-AI collaboration partially mediated the relationship between agent use and discovery efficiency, while perceived AI risk was negatively associated with agent use. Data quality and organisational AI readiness strengthened the modeled AAU -> PDE relationship, whereas task complexity weakened it. Product-team maturity did not produce a statistically supported interaction, and H5d is therefore reported as not supported rather than being forced to match the original hypothesis. The revised analysis reinforces the central socio-technical argument that the potential value of autonomous agents depends on human collaboration, data conditions, task characteristics, governance, and organisational capability. At the same time, the paper remains a conceptual and computational model demonstration. The reported coefficients are properties of researcher-specified data and should not be generalised to professionals or organisations. Empirical validation with real product teams is required before substantive claims can be made about the magnitude or causal direction of autonomous AI-agent effects on product discovery.

References

Abou Ali, M., Dornaika, F., & Charafeddine, J. (2026). Agentic AI: A comprehensive survey of architectures, applications, and future directions. Artificial Intelligence Review, 59, Article 11. https://doi.org/10.1007/s10462-025-11422-4

Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: A socio-technical perspective. Part I: The causes. MIS Quarterly, 1(3), 17–32. https://doi.org/10.2307/248710

Bouschery, S. G., Blazevic, V., & Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. Journal of Product Innovation Management, 40(2), 139–153. https://doi.org/10.1111/jpim.12656

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

Cooper, R. G. (2024). The artificial intelligence revolution in new-product development. IEEE Engineering Management Review, 52(1), 195–211. https://doi.org/10.1109/EMR.2023.3336834

Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organization Science, 37(2), 403–423. https://doi.org/10.1287/orsc.2025.21838

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), Article eadn5290. https://doi.org/10.1126/sciadv.adn5290

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. Information Systems Research, 33(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Gama, F., & Magistretti, S. (2025). Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. Journal of Product Innovation Management, 42(1), 76–111. https://doi.org/10.1111/jpim.12698

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. MIS Quarterly, 19(2), 213–236. https://doi.org/10.2307/249689

Knudsen, M. P., von Zedtwitz, M., Griffin, A., & Barczak, G. (2023). Best practices in new product development and innovation: Results from PDMA's 2021 global survey. Journal of Product Innovation Management, 40(3), 257–275. https://doi.org/10.1111/jpim.12663

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. Organization Science, 33(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Mariani, M., & Dwivedi, Y. K. (2024). Generative artificial intelligence in innovation management: A preview of future research developments. Journal of Business Research, 175, Article 114542. https://doi.org/10.1016/j.jbusres.2024.114542

National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.600-1

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (Article 2, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3586183.3606763

Qian, C., Liu, W., Liu, H., Chen, N., Dang, Y., Li, J., Yang, C., Chen, W., Su, Y., Cong, X., Xu, J., Li, D., Liu, Z., & Sun, M. (2024). ChatDev: Communicative agents for software development. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Vol. 1, pp. 15174–15186). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.810

Rasheed, Z., Waseem, M., Sami, M. A., Kemell, K.-K., Ahmad, A., Duc, A. N., Systä, K., & Abrahamsson, P. (2025). Autonomous agents in software development: A vision paper. In Agile processes in software engineering and extreme programming—Workshops (pp. 15–23). Springer. https://doi.org/10.1007/978-3-031-72781-8_2

Schäper, T., Foege, J. N., & Nüesch, S. (2024). Toolkits for innovation: How digital technologies empower users in new product development. R&D Management, 54(1), 95–117. https://doi.org/10.1111/radm.12642

Vaccaro, M., Almaatouq, A., & Malone, T. W. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour, 8, 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18, Article 186345. https://doi.org/10.1007/s11704-024-40231-1

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X.-Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. Scientific Reports, 15, Article 15105. https://doi.org/10.1038/s41598-025-98385-2