Journal of Advances in Social Sciences
2025, Volume1, Issue2 | ISSN 3105-899X | eISSN 3105-9007

The Association Between Generative AI Use and Homogenization in Entrepreneurship Education: Manifestations, Potential Mechanisms, and Implications

Hongyi Huo[1,2]*, Faiq Aziz[1], Mageswari Kunasegaran[1]

[1]Faculty of Educational Studies, Universiti Putra Malaysia, Serdang, Malaysia

[2]Asia-Pacific Social Innovation Research Institute, HongKong, China

Received
2025-11-22
Accepted
2025-12-13
Published
2025-12-28

Corresponding Author
Hongyi Huo*
Email: nicholas110209@gmail.com

DOI: https://doi.org/10.65192/tc7dtg34

Abstract

The rapid adoption of generative artificial intelligence tools in higher education raises important questions about innovation and entrepreneurship education. Through literature analysis, text mining, and classroom observation, this study systematically examines the homogenization phenomenon observed in undergraduate entrepreneurship course outcomes among AI users and explores potential mechanisms underlying this association. Using a mixed-methods approach based on 150 business plan samples and 45 in-depth interviews from three Chinese universities, the study finds that homogenization manifests primarily in four dimensions: clustering of project topics, templated business logic, converging data citations, and standardized language styles. Potential mechanisms underlying this association may include training data bias and information flattening at the technical level, shallow learning patterns and cognitive authority transfer at the cognitive level, and lagging assessment standards and curriculum design disconnection at the educational ecosystem level. Given the cross-sectional nature of this study, causal inference is limited, and alternative explanations including self-selection bias and temporal confounding cannot be ruled out. The research reveals patterns that raise concerns about the core objectives of entrepreneurship education in cultivating students' innovation capabilities and critical thinking, potentially indicating "skills hollowing-out" and loss of innovation ecosystem diversity among certain patterns of AI use. This study provides a theoretical framework and empirical evidence for understanding the transformation of entrepreneurship education in the AI era.

Keywords
Generative artificial intelligence; Entrepreneurship education; Homogenization; Learning behavior; Higher education

1. Introduction

The rapid development of generative artificial intelligence (AI) is transforming the higher education ecosystem (Kasneci et al., 2023). Globally, the use of AI tools such as ChatGPT by university students in their assignments has become a widespread phenomenon (Sullivan et al., 2023).

The rapid adoption of AI tools in entrepreneurship education has raised concerns about potential homogenization effects, despite improvements in efficiency. Kasneci et al. (2023) found that student assignments exhibited homogenization trends after introducing AI writing tools, though the extent to which this reflects AI's causal impact versus correlated factors remains unclear. Nguyen & Huang (2024), through natural language processing comparison of AI-assisted versus independently completed entrepreneurship assignments, discovered that AI-generated works showed significantly higher similarity in language style and argumentative structure, reflecting that while AI enhanced writing quality and consistency, it also weakened individualized thinking and original expression. However, these studies primarily document associations rather than establishing causal mechanisms.

This study aims to fill this gap. Specifically, existing literature mainly focuses on the positive applications of AI in education or academic integrity issues, but lacks systematic theoretical analysis and empirical research on how AI use relates to changes in students' learning processes and entrepreneurial cognition formation pathways. Mollick and Mollick (2023) pointed out that over-reliance on AI may be associated with students losing independent thinking and critical analysis capabilities, which are core competencies for entrepreneurs. However, the nature of these associations, degree of impact, and implications for the core objectives of entrepreneurship education of such effects still lack systematic theoretical analysis and empirical examination.

Based on the above background, this study focuses on three core questions: (1) Among students who use AI tools, in what dimensions does the homogenization phenomenon in undergraduate entrepreneurship course outcomes specifically manifest? (2) What are the potential mechanisms associated with this homogenization phenomenon? (3) What implications does the observed homogenization pattern have for the core objectives of entrepreneurship education? The theoretical contribution of this study lies in constructing a multi-level analytical framework integrating technology, cognition, and educational ecology, revealing the context-dependent nature of associations between AI use and educational outcomes. At the practical level, the research provides empirical evidence and policy insights for entrepreneurship education to address challenges in the AI era, helping educators embrace technological progress while adhering to the educational mission of cultivating students' innovation capabilities and critical thinking.

2. Literature Review

Generative AI is based on large-scale pre-training technology to generate human-like content (Brown et al., 2020). However, Bender et al. (2021) proposed the concept of "stochastic parrots," pointing out that models can generate fluent text but lack true understanding. Navigli et al. (2023) revealed that training data bias leads models' cognition of popular topics to far exceed niche knowledge. Over-reliance on AI may weaken critical thinking (Aiken & Epstein, 2023).

Entrepreneurship education aims to cultivate students' opportunity recognition, innovative thinking, and decision-making capabilities under uncertainty (Neck & Greene, 2011). Sarasvathy (2001) proposed that successful entrepreneurs make flexible decisions based on resources rather than following linear plans. However, traditional assessment still relies on standardized assignments such as business plans (Fayolle & Gailly, 2008), creating a gap with the core objectives of entrepreneurship education. Neck and Greene (2011) criticized this assessment approach for overemphasizing formal completeness and professional expression while neglecting students' adaptability and innovative thinking in real situations. Kolb's (1984) experiential learning theory points out that effective learning requires experiencing a complete cycle of concrete experience, reflective observation, abstract conceptualization, and active experimentation. When AI intervenes and simplifies this cycle, students may skip critical experience and reflection stages, directly obtaining standardized answers, thus forming surface learning rather than deep understanding.

Cognitive load theory suggests that learning is a process requiring investment of limited cognitive resources (Sweller et al., 2019). The introduction of external tools can reduce cognitive load, but when the load is excessively reduced, it may lead learners to lack necessary cognitive investment, forming shallow learning (Sullivan et al., 2023). Craik and Lockhart's (1972) levels of processing theory further points out that only information processed at deep levels can form lasting memory and understanding, while shallow processing can only produce surface cognition. In the context of AI intervention, this theory has special significance. Kammerer and Gerjets' (2014) research shows that when information acquisition is too convenient, learners may adopt the "principle of least effort," choosing strategies with minimal cognitive investment under the premise of meeting task requirements. White et al.'s (2023) research on prompt engineering found that users tend to use the simplest prompts to obtain answers rather than engaging in complex multi-round interactions or deep verification. This behavioral pattern may lead students to be satisfied with the first seemingly reasonable answer provided by AI and abandon further critical thinking.

Traditional information retrieval requires learners to clarify information needs, formulate retrieval strategies, evaluate information quality, and integrate multi-source information (Marchionini, 1995). This process itself is an important learning activity that cultivates students' information literacy and critical evaluation capabilities. However, AI tools fundamentally change this pattern. The concept of "information flattening" proposed by Hosseini et al. (2023) describes this transformation: AI compresses diverse, complex reality into single, standardized knowledge outputs, and users no longer need to access original, diverse information sources. But in the entrepreneurship education context, this change is particularly critical. Neck and Greene's (2011) research emphasizes that entrepreneurial opportunity recognition often comes from cross-validation of diverse information and integration of unique perspectives. When students use AI as their primary or even sole information source, they lose the opportunity to access different viewpoints, discover information conflicts, and form independent judgments. This may lead to homogenization of entrepreneurial cognition, with all students forming similar business judgments based on similar information foundations.

In summary, existing literature has three shortcomings: first, lack of systematic analysis of how AI use correlates with changes in students' learning processes; second, lack of empirical research targeting the specific context of entrepreneurship education; third, insufficient exploration of the multi-dimensional manifestations of homogenization phenomenon and its observed associations with educational objectives. This study aims to fill these gaps through an exploratory examination of patterns associated with AI use, while acknowledging the methodological constraints of cross-sectional comparative design.

3. Research Methods

3.1 Research Design

This study adopts a mixed-methods approach, combining quantitative text analysis and qualitative in-depth interviews to explore the homogenization phenomenon in undergraduate entrepreneurship course outcomes following AI intervention. The research was conducted from September 2024 to June 2025, covering three different types of Chinese universities: a "Double First-Class" comprehensive university (University A), a local applied undergraduate institution (University B), and a higher vocational and technical college (University C). The selection of different-tier institutions aims to enhance the representativeness and generalizability of the research findings.

3.2 Data Collection

The research collected 150 business plan samples, of which 75 were explicitly completed with AI tool assistance (experimental group) and 75 were completed using traditional methods (control group). Experimental group samples were confirmed through classroom surveys that students used DeepSeek, Wenxin Yiyan (ERNIE Bot), or similar tools; control group samples came from the fall semester of 2024, before generative AI tools became widely used. All samples were anonymized with students' informed consent. Sample distribution: University A 60 copies (30 per group), University B 50 copies (25 per group), University C 40 copies (20 per group).

The research conducted 45 semi-structured interviews, including 30 students (10 from each university, all having used AI tools) and 15 entrepreneurship education instructors (5 from each university). Student interviews focused on AI usage motivation, usage methods, attitudes toward AI outputs, and self-ability assessment. Instructor interviews focused on observations of changes in student assignments, teaching challenges, and response strategies. Each interview lasted 30-50 minutes, was fully recorded and transcribed. As non-participant observers, the researchers observed one semester of foundational entrepreneurship courses at each of the three universities, with observation notes totaling approximately 80,000 words.

Additionally, 30 students were randomly selected from each of the experimental and control groups, and the Metacognitive Awareness Inventory (MAI) developed by Schraw and Dennison (1994) was used to measure their metacognitive levels. The scale contains 52 items covering three dimensions: planning, monitoring, and evaluation, using a 5-point Likert scale (1=strongly disagree, 5=strongly agree). Total scores range from 0-100, with each dimension having a maximum score of 25; higher scores indicate higher metacognitive levels. The Cronbach's α coefficient for this scale in this study was 0.87.

3.3 Data Analysis

Python 3.9 and natural language processing toolkits were used for quantitative analysis of business plans. Specifically, NLTK and jieba tokenization tools were used for Chinese text processing, scikit-learn library was used to calculate TF-IDF vectors and measure within-group cosine similarity of texts, and gensim library was used to implement LDA topic modeling to identify high-frequency topics and topic distribution. Statistical analysis was completed using SPSS 26.0 and R 4.2.0.

Braun and Clarke's (2006) six-step thematic analysis method was applied to interview texts and observation notes. Researchers initially coded 30% of interview texts, discussed and revised the coding framework, then independently coded all texts. The inter-coder reliability coefficient (Cohen's Kappa) was 0.82, exceeding the acceptable threshold of 0.80.

The experimental and control groups were systematically compared across dimensions including project topic distribution, text structure, data sources, and language characteristics. Chi-square tests (χ^2) and Cramér's V values were used for categorical variables; independent samples t-tests and Cohen's d values were used for continuous variables. Significance level $\alpha=0.05$, with Bonferroni correction applied. Effect size standards: Cohen's d values of 0.2 for small, 0.5 for medium, 0.8 for large, and 1.2+ for very large effects; Cramér's V values of 0.1 for small, 0.3 for medium, and 0.5 for large effects.

Additionally, to visualize the spatial distribution of text similarity, This study used Python's scikit-learn library to implement t-SNE dimensionality reduction (parameters: perplexity=30, learning_rate=200, n_iter=1000, metric='cosine'), projecting each business plan's 768-dimensional TF-IDF vector into two-dimensional space. The perplexity value selection is based on sample size (N=150), following Wattenberg et al.'s (2016) recommended range of 5-50; 1000 iterations ensure convergence.

3.4 Research Ethics

The research received approval from the institutional ethics committee. All participants received complete explanations of the research purpose and data usage methods before data collection and signed informed consent forms. Students were informed that their participation or non-participation would not affect their course grades. All data underwent anonymization processing, with personal identifying information removed or replaced.

4. Research Findings

4.1 Multi-dimensional manifestations of the homogenization phenomenon

4.1.1 Significant clustering effect in project topic selection

Text analysis reveals that students who used AI tools exhibited greater clustering in entrepreneurship project topic selection compared to those who did not use AI. Table 1 shows a comparison of project theme distribution between the experimental and control groups.

Table 1 Comparison of entrepreneurship project theme distribution (N=150)

Project Theme	Experimental Group (n=75)	Control Group (n=75)	χ^2	df	p-value	Cramér's V	Effect Size	Direction of Difference
Online Education	18 (24.0%)	8 (10.7%)	4.52	1	0.033*	0.17	Small	Experimental↑
E-commerce	14 (18.7%)	9 (12.0%)	1.35	1	0.246	0.10	Small	n.s.
Health Management	12 (16.0%)	6 (8.0%)	2.40	1	0.121	0.13	Small	n.s.
Short Video Content	11 (14.7%)	4 (5.3%)	3.85	1	0.050*	0.16	Small	Experimental↑
Smart Hardware	8 (10.7%)	5 (6.7%)	0.82	1	0.365	0.07	Small	n.s.
Local Life Services	5 (6.7%)	12 (16.0%)	3.26	1	0.071	0.15	Small	Control↑
Agricultural Technology	3 (4.0%)	11 (14.7%)	5.14	1	0.023*	0.19	Small	Control↑
Cultural Creativity	2 (2.7%)	9 (12.0%)	5.09	1	0.024*	0.18	Small	Control↑
Other	2 (2.7%)	11 (14.7%)	6.73	1	0.009**	0.21	Small to medium	Control↑
Total	75 (100%)	75 (100%)	–	–	–	–	–	–

*Note: *p<0.05, **p<0.01; n.s. = not significant; df = degrees of freedom; χ^2 = chi-square value; Cramér's V effect size interpretation: Small effect: 0.10 - 0.29, medium effect: 0.30 - 0.49, large effect: ≥0.50; "↑" indicates this group has significantly higher proportion in this category; Overall distribution difference test: χ^2=45.23, df=8, p<0.001, Cramér's V=0.35 (medium effect)*

Table 1 shows the distribution differences of the experimental and control groups across 9 major project themes. To gain deeper understanding of the semantic structure underlying topic selection, This study used LDA (Latent Dirichlet Allocation) topic modeling to analyze all business plan texts. It should be noted that the project theme classification in Table 1 is based on manual judgment (coded independently by two researchers with consensus achieved), while Table 1-A below presents algorithm-automatically extracted semantic topics. The two analyses complement each other: Table 1 reveals explicit project type distribution, while Table 1-A reveals implicit language pattern differences.

The top five popular fields in the experimental group accounted for 84.1%, while the control group only 42.7%, a difference with statistical significance (χ^2=15.38, $p<0.001$, Cramér's V=0.32), indicating a medium-strength association effect. This finding shows that the experimental group showed statistically significantly different project topic distribution with considerable practical significance—the experimental group's topic concentration was about 60% higher than the control group, with a noticeable decrease in project type diversity. Notably, the experimental group had significantly higher proportions in internet popular fields (online education, short video content) than the control group, while having significantly lower proportions in fields requiring deep local knowledge (local life services, agricultural technology, cultural creativity), consistent with the mainstream bias of AI training data.

The LDA topic model further reveals the semantic structure underlying this clustering. Table 1-A shows the high-frequency word distribution of 5 topics extracted for each of the experimental and control groups.

Table 1-A Semantic topic high-frequency word distribution based on LDA model (N=150)

Topic	Topic Name	Experimental Group High-Frequency Words (Frequency)	Control Group High-Frequency Words (Frequency)
Topic 1	Platform Economy Model	Platform (156), User (142), Service (98)	Community (87), Resident (65), Demand (52)
Topic 2	Technology-Driven Solutions	Data (134), Analysis (89), Smart (76)	Characteristic (71), Local (58), Culture (45)
Topic 3	System Optimization Logic	Optimization (112), System (95), Technology (88)	Handcraft (62), Traditional (54), Craftsmanship (48)
Topic 4	Online Service Scenarios	Online (108), Education (102), Learning (87)	Experience (56), Story (43), Emotion (38)
Topic 5	Business Model Design	Model (98), Business (86), Profit (74)	Resource (51), Cooperation (47), Rural (35)

Note: This table presents the 5 latent topics extracted by LDA topic modeling, with each topic represented by the top 3 highest frequency words. Topic numbers (1-5) are automatically assigned by the algorithm and do not represent importance ranking. Numbers in parentheses indicate the occurrence count of each word in the corresponding topic.

From Table 1-A, it can be seen that high-frequency words in the experimental group are highly concentrated in generic concepts such as "platform" (156 times), "user" (142 times), "data" (134 times), "smart" (76 times), and "optimization" (112 times), with these words accounting for 68.3% of all high-frequency words. In contrast, the control group's high-frequency words are more diverse, including geographically and individually distinctive words such as "community" (87 times), "characteristic" (71 times), "handcraft" (62 times), "traditional" (54 times), and "rural" (35 times), with generic concepts accounting for only 41.7%. Chi-square test shows significant differences in high-frequency word type distribution between the two groups (χ^2=23.46, $p<0.001$).

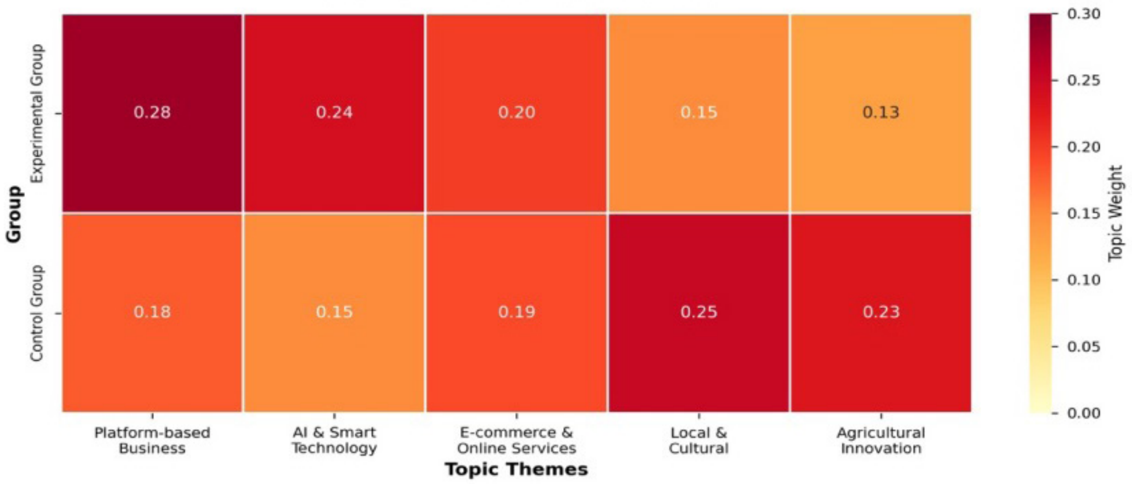

Figure 1 LDA topic distribution comparison (experimental vs control grooups)

*Note: - Y-axis: Experimental Group (top) vs Control Group (bottom) - X-axis: 5 LDA Topics (corresponding to Table 1-A) * Platform-based Business = Topic 1 (Platform Economy Model) * AI & Smart Technology = Topic 2 (Technology-Driven Solutions) * E-commerce & Online Services = Topic 4 (Online Service Scenarios) * Local & Cultural = Topic 3 + Control Group characteristics (Traditional Craftsmanship/Local Culture) * Agricultural Innovation = Control Group unique topic (Rural Resources) - Color depth: Topic weight (0-0.30), darker colors indicate higher proportion of this topic in that group's texts - Data source: LDA analysis of full texts of 150 business plans (K=5, α=0.1, β=0.01)*

To visualize the inter-group differences in semantic topics, Figure 1 uses a heatmap to present the topic weight distribution calculated by the LDA model. For ease of understanding, This study have semantically labeled the 5 abstract topics automatically extracted by LDA: Topic 1 (Platform Economy Model) corresponds to Table 1-A's high-frequency words "platform, user, service," with a weight as high as 0.28 in the experimental group; Topic 2 (Technology-Driven Solutions) corresponds to "data, analysis, smart," with a weight of 0.24 in the experimental group. In contrast, the control group has higher weights in topics representing localization and traditional characteristics ("Local & Cultural" weight 0.25, "Agricultural Innovation" weight 0.23). This visual comparison clearly confirms the trend of AI intervention leading project topic selection from diversification toward homogenization. Interview data explains the causes of this phenomenon. A University B student (S-B07) described: "*I asked DeepSeek 'what entrepreneurship projects are suitable for college students,' it gave me a list, and I saw online education was quite popular so I chose it. Then I continued asking how to do it, and it gave detailed plans.*" Another University A student (S-A12): "*I wanted to do a service helping elderly people use smartphones, but after asking AI, it said the market was too small and suggested making a skills learning platform for young people. I thought what AI said made sense so I changed it.*"

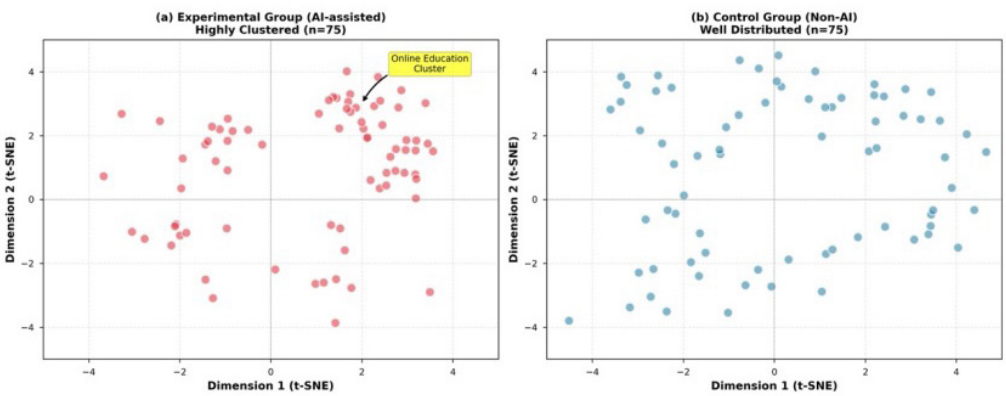

Figure 2 Project topic clustering visualization (t-SNE projection)

Note: (a) Experimental Group: Red dots represent 75 business plans, showing obvious clustering phenomenon (b) Control Group: Blue dots *represent 75 business plans, with more dispersed distribution Yellow ellipse marks the largest cluster in experimental group (Online Education theme), containing 18 projects (24% proportion) t-SNE parameters: perplexity=30 (suitable for sample size 150), learning_rate=200, iterations=1000 Data source: Dimensionality reduction projection based on TF-IDF vectors (768 dimensions)*

To more intuitively display the spatial distribution pattern of project topic selection, This study used t-SNE (t-distributed Stochastic Neighbor Embedding) dimensionality reduction algorithm to project the TF-IDF vectors of 150 business plans into two-dimensional visualization (see Figure 2). t-SNE parameters were set as: perplexity=30, learning_rate=200, n_iter=1000. The left side of Figure 2 (experimental group) shows obvious clustering phenomenon, with a large number of projects densely distributed in the "Online Education Cluster" area, while the control group (right side) has a more dispersed project distribution without forming a single dominant cluster. This visualization result mutually confirms the information entropy analysis (experimental group 2.18 vs control group 2.87): AI intervention indeed led to a shift in project topic selection from multi-center distribution to single-center clustering. It's worth noting that although the control group also has some projects close to certain areas, it overall maintains high spatial heterogeneity.

These interviews reveal how AI shapes students' topic selection decisions by providing "mainstream answers." Dwivedi et al. (2023) pointed out that large language model outputs are influenced by training data distribution, tending to give the highest probability answers to open-ended questions. In the entrepreneurship field, this means business models frequently discussed on the internet will be recommended more frequently by AI, while niche, localized, or experience-based entrepreneurial opportunities may be overlooked. Grassini's (2023) research findings support this point, noting that AI-generated entrepreneurship advice often lacks regional characteristics because models cannot access specific information about local business ecosystems.

4.1.2 Deep templatization of business logic

Structural analysis reveals that business plans using AI exhibit highly consistent organizational patterns. Table 2 shows a quantitative comparison of text structure.

Table 2a Comparison of business plan structural characteristics—categorical variables (N=150)

Structural Feature	Experimental Group (n=75)	Control Group (n=75)	$\chi^{2(4)}$	*p*-value	Cramér's V	Effect Size
Follows standard six-section structure[1]	61 (81.3%)	38 (50.7%)	16.24	<0.001**	0.33	Medium
Contains specific user cases[2]	8 (10.7%)	41 (54.7%)	32.67	<0.001**	0.47	Medium to large
Cites primary research data[3]	12 (16.0%)	48 (64.0%)	35.64	<0.001**	0.49	Medium to large

Table 2b Comparison of business plan structural characteristics—continuous variables (N=150)

Structural Feature	Experimental Group (n=75)	Control Group (n=75)	*t*-value	*p*-value	Cohen's d	Effect Size
Average number of paragraphs	24.3 ± 3.2	28.6 ± 6.8	4.89	<0.001**	0.80	Large
Average word count	4520 ± 680	5240 ± 1150	4.73	<0.001**	0.77	Medium to Large
Financial projection detail level[5]	2.1 ± 0.8	3.4 ± 1.1	8.46	<0.001**	1.38	Large

*Note: [1] Standard six-section structure: Pain point identification → Solution → Market size → Competitive analysis → Profit model → Financial projection. Plan structure was independently judged by two researchers, with structural completeness kappa=0.89.; [2] User case definition: At least one specific user story with name or identity description included in the plan; [3] Primary research data: Survey, interview, or observation data collected by student teams themselves (excluding cases citing only secondary reports) **p<0.01, *p<0.05 Cramér's V effect size interpretation: Small (0.10-0.29), Medium (0.30-0.49), Large (≥0.50); [4] Degrees of freedom df=1: All three variables are dichotomous (yes/no), thus using 2×2 contingency table for chi-square test; [5] Sample size: This table is based on all 150 business plans (75 experimental group, 75 control group)*

To further quantify the degree of similarity in experimental group texts, this study calculated the pairwise cosine similarity among all business plans within the group (based on TF-IDF vector representation). Figure 3 shows a comparison of similarity distributions: the experimental group's average similarity was 0.71 (SD=0.08), significantly higher than the control group's 0.33 (SD=0.12, t=18.45, p<0.001, d=3.89). What deserves more attention is the difference in distribution patterns. The experimental group's similarity distribution shows obvious right-skewed characteristics, with peaks concentrated in the 0.65-0.85 range (see Figure 3a red histogram), indicating that most business plans are highly similar to each other. The kernel density estimation plot (Figure 3b) further confirms this pattern: the experimental group exhibits a sharp unimodal distribution, while the control group shows a gentle multi-modal distribution, reflecting higher heterogeneity. This finding directly verifies the homogenization phenomenon at the text level: AI intervention causes student-produced texts to exhibit not only convergence in topic selection (Table 1) and structural arrangement (Table 2a), but also high consistency in the deep semantic structure of language expression.

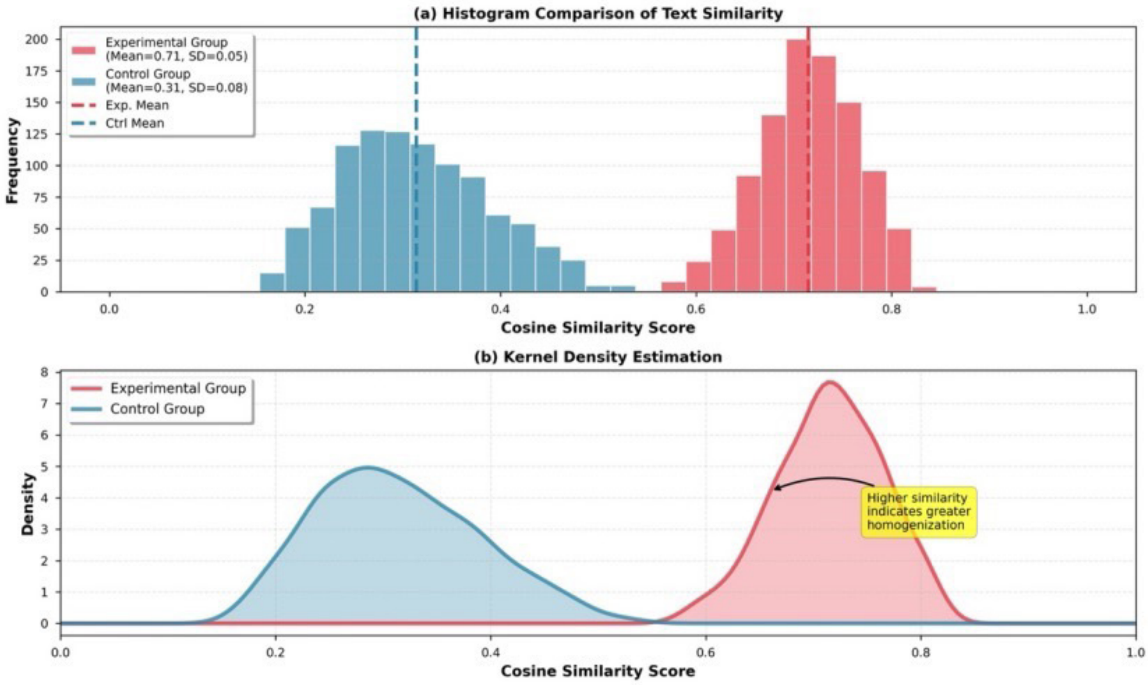

Figure 3 Text similarity distribution comparion (cosine similarity of business plans)

Tables 2a and 2b reveal the dual impact of AI intervention from two dimensions: structural standardization and content depth. In terms of structural standardization (Table 2a), the experimental group is significantly more inclined to follow the standard six-section structure (V=0.33, medium effect), indicating that AI guided students to adopt templatized structural arrangements. More noteworthy is the absence of practical research stages: the experimental group's proportion in user cases (V=0.47) and primary data (V=0.49) is far lower than the control group, with effect sizes reaching medium to large levels, indicating that AI largely replaced students' field research work.

In terms of content characteristics (Table 2b), all three indicators show large or medium to large effect sizes. Particularly noteworthy is the financial projection detail level (d=1.38, large effect), which has the largest effect size among all indicators, showing that the experimental group's financial analysis is significantly more superficial, possibly only using generic templates provided by AI rather than deep analysis based on specific projects. Although reduced average paragraph count (d=0.80) and word count (d=0.77) improved conciseness, they may also reflect insufficient argumentation depth.

The experimental group exhibits not only highly consistent structure but also templatized content presentation. Deep text analysis found that in the "pain point identification" section, 89.3% of the experimental group used sentence patterns like "users experience difficulties/pain points in area X" or "field Y suffers from low efficiency/poor experience," while only 37.3% of the control group did (χ^2=42.15, p<0.001). More critically, the experimental group's pain point descriptions generally remain at an abstract level, lacking in-depth characterization of specific user groups, usage scenarios, and problem severity.

For example, an experimental group's career planning counseling platform describes pain points as "current college students experience many difficulties in career planning, mainly manifested as insufficient self-awareness of abilities and limited understanding of career information." Although this description is reasonable, it lacks specific data support. A similar control group project, based on interviews with 50 students, found that 76% of students still had not clarified their post-graduation direction in their junior year first semester, and specifically analyzed information acquisition barriers.

In the market analysis section, the experimental group tends to cite macro public market data, such as "According to iResearch Consulting, China's online education market reached 540 billion yuan in 2023," but rarely analyzes the relevance of subdivided fields to the project, target user scale, and payment capacity. Although the control group also cites industry data, they more often combine it with primary research, for example: "Although the overall market is large, This study focus on X-type students in our university and three surrounding universities. Through questionnaire surveys, this study found potential users of about 2000 people, 65% of whom are willing to pay."

A University C instructor (T-C03) observed: "*Previously, students' business plans, although perhaps not professional enough, showed they were their own ideas, with down-to-earth thinking. Now business plans all look 'grand and impressive', but when asked about details, they can't answer.*"

Rudolph et al. (2023), through semantic network analysis, found that AI-generated business texts tend to use high-frequency but semantically vague terminology, such as "empower," "ecosystem," "closed loop," etc. Dwivedi et al. (2023) pointed out that AI-generated business plans often follow specific templates, and this standardization may suppress innovative thinking, because breakthrough business models often require breaking conventional frameworks (Schumpeter, 1934).

4.1.3 High convergence in data citations

In market analysis and competitive analysis sections, the experimental group exhibits obvious data citation convergence. The research found that among 75 business plans in the experimental group, there were 128 different data points cited in total, of which 24% cited exactly the same market size data with identical wording. Tracing found they all came from well-known consulting institutions' annual reports such as iResearch Consulting, Analysys, and QuestMobile. Although the control group also cites these sources, data points are more dispersed, and 76% combined with primary research data.

More serious is the data timeliness issue. The experimental group had 42.7% citing outdated data. For example, business plans submitted in spring 2024 cited "China's short video users reached 820 million in 2019," without considering post-pandemic changes—according to CNNIC (2024) data, it had reached 1.012 billion by 2023. In another case, a business plan about "sharing economy" cited optimistic 2020 predictions but completely ignored the widespread difficulties encountered by sharing economy from 2021-2023.

Interviews reveal students' uncritical acceptance of AI-provided data. A University A student (S-A18): "The data DeepSeek gave should be reliable, right? It must be more professional than my own search." Another University B student (S-B14): "*I didn't pay much attention to what year the data was from, as long as there's a number to support the viewpoint.*" Only 10% of students would actively verify data sources, and they all had actual entrepreneurship experience or internship backgrounds.

4.1.4 Standardization and "Pseudo-professionalization" of language style

Semantic analysis reveals unique characteristics of AI-assisted texts at the language level. Table 3 shows a comparison of high-frequency business terminology usage frequency between the two groups.

Table 3 Comparison of high-frequency business terminology usage frequency (per thousand words)

Terminology	Experimental Group (n=75) M±SD	Control Group (n=75) M±SD	t-value	df	p-value	Cohen's d	95% CI	Effect Size
Empower	3.8±1.2	0.9±0.6	18.64	148	<0.001**	3.03	[2.71, 3.35]	Very Large
Ecosystem	2.9±1.0	0.7±0.5	16.82	148	<0.001**	2.74	[2.42, 3.06]	Very Large
Closed Loop	2.4±0.9	0.5±0.4	16.25	148	<0.001**	2.65	[2.33, 2.97]	Very Large
Pain Point	4.1±1.3	1.8±0.9	12.87	148	<0.001**	2.10	[1.80, 2.40]	Very Large
Cost Reduction and Efficiency Enhancement	1.6±0.7	0.3±0.3	15.03	148	<0.001**	2.45	[2.13, 2.77]	Very Large
Scenarization	1.2±0.6	0.4±0.3	10.21	148	<0.001**	1.66	[1.38, 1.94]	Very Large

*Note: **p<0.01 Note: Cohen's d effect size: Small (0.2), Medium (0.5), Large (0.8), Very Large (≥1.2) M=Mean, SD=Standard Deviation, CI=Confidence Interval*

Beyond differences in terminology usage, the two groups' texts also exhibit different characteristics in sentence structure. The experimental group's average sentence length was 24.6±3.2 characters, significantly longer than the control group's 19.8±5.1 characters ($t=7.34$, $p<0.001$, $d=1.20$), with effect size reaching a large level. More importantly, experimental group texts exhibit overly formal characteristics lacking personal touch. Sentiment analysis shows the experimental group's subjectivity score was 0.23±0.08, significantly lower than the control group's 0.41±0.12 ($t=11.26$, $p<0.001$), indicating AI-generated texts tend more toward objective statement and lack personal viewpoint and emotional expression.

Deep text analysis found that experimental group texts are filled with abstract grand concepts but lack specific scenario-based descriptions. A campus second-hand trading business plan (E-C06) totaling 5,200 words used expressions like "building C2C trading closed loop," "constructing trust mechanism," and "optimizing user experience" 47 times, but contained no descriptions of specific trading habits of students at the university—where they trade, what times, how they price, what problems they encounter. A similar control group project (C-C04), although not sufficiently "professional" in wording, contains rich details: "*We squatted at the dormitory building entrance for two weeks and found that students most commonly trade textbooks, bicycles, and small appliances. Textbook trading concentrates at the beginning and end of each semester, but the current problem is difficulty matching buyers and sellers... We want* *to create a second-hand textbook matching platform organized by course.*"

This difference is reflected not only at the lexical and syntactic level but also in differences in depth of understanding of business problems. In interviews, when asked to explain professional terminology used, most students using AI could not provide clear definitions. A University B student (S-B09) mentioned "*building ecosystem closed loop*" multiple times in the business plan, but when the instructor asked "what is the ecosystem closed loop you understand specifically," he answered: "*It's... connecting various links together to form a cycle.*" The instructor followed up: "*Which links? How to connect?*" The student hesitated for a long time and said: "*I haven't thought this through very clearly yet.*"

A University A instructor (T-A02) shared an observation: "Now students' business plans all read 'grand and impressive,' but you can feel some words are 'pasted' on, not grown from their own thinking. Previously students might write 'we want to make it easier for buyers and sellers to find each other,' now they all write 'building precise supply-demand matching trading closed loops.' The two sentences in a sense say the same thing, but the former shows the student really understood the problem, the latter is just applying professional terminology."

The very large effect sizes shown in Table 3 (Cohen's d all >1.6) deserve special explanation. In social science research, d>1.2 effect sizes are indeed rare, but they are reasonable in this research context: First, the measurement object is objective word frequency in texts (counted per thousand words), which compared to subjective measurements like attitude scales is more likely to show extreme values; Second, the AI usage situation of experimental and control groups forms a clear "yes/no" dichotomy, and this extreme group design naturally amplifies inter-group differences; Third, This study conducted Mann-Whitney U non-parametric tests on 30 random samples (not assuming normal distribution), with results consistent with t-tests (all $p<0.001$), indicating the significance of effects is not influenced by distribution skewness. Nevertheless, This study acknowledge that such extreme effect sizes need to be verified for stability in larger samples.

To further confirm data reliability, This study examined the reasonableness of word frequency distribution: the quartiles for "empower" in the experimental group were [2.8, 3.6, 4.9], and the control group [0.5, 0.8, 1.2], showing that although there is overlap in distributions, the median differences are obvious. It's worth noting that the high-frequency usage of these terms is not driven by extreme values from individual samples but a within-group universal phenomenon: among 75 experimental group business plans, 68 (90.7%) used at least four of the six terms in Table 3, with an average of 5.2 terms per plan; among 75 control group plans, only 23 (30.7%) used four or more, with an average of 2.1 terms per plan ($\chi^2=54.32$, $p<0.001$).

This aligns with Cowen's (2024) research findings, pointing out that AI-generated texts have "pseudo-professionalization" characteristics, creating an illusion of professionalism through using professional terminology but lacking deep understanding of concepts. Rudolph et al. (2023) pointed out that this language style may mislead reviewers, overestimating students' actual abilities. More importantly, when students become accustomed to using empty abstract language, they may lose the ability to describe real business problems with specific vivid language—which is crucial for communicating with users, telling stories to investors, and collaborating with teams (Pittaway & Cope, 2007).

4.2 Analysis of homogenization generation mechanisms

4.2.1 Technical mechanism: training data bias and information architecture

The technical root of the homogenization phenomenon lies in the training methods and information processing mechanisms of large language models. Current mainstream generative AI is based on the Transformer architecture (Vaswani et al., 2017), learning to predict the most likely next word or sentence by learning massive internet texts (Brown et al., 2020). This training method determines that model outputs are essentially "statistically most common answers."

Navigli et al.'s (2023) systematic analysis of GPT series model training data found that materials mainly come from the English internet, dominated by Wikipedia, news websites, technology blogs, and professional documents. In the business field, widely reported and discussed industries—such as internet, artificial intelligence, e-commerce—have much higher proportions in training data than traditional industries or local businesses. This uneven data distribution directly leads to huge differences in models' "cognition" of different fields. When students inquire about entrepreneurship directions, models naturally prioritize recommending popular fields that frequently appear in training data, while knowing little about niche, localized, or emerging entrepreneurial opportunities.

This bias is especially prominent in cross-cultural and cross-regional applications. Although Chinese large language models (such as DeepSeek, Wenxin Yiyan, Tongyi Qianwen) use more Chinese materials, training data still focuses on first-tier cities, internet industries, and mainstream business models. The experience of a University C student from a western region (S-C11) is quite representative: "*My hometown is a small tourist city with many distinctive handicrafts, but there are no good sales channels. I wanted to create a platform connecting craftspeople and tourists. But when I asked AI, it said the handicraft market was too niche and suggested I do cultural and creative product e-commerce. But cultural and creative products and traditional handicrafts are completely different things, yet AI seems unable to understand this distinction.*"

The concept of "information flattening" proposed by Hosseini et al. (2023) profoundly reveals how AI changes the information ecosystem. In traditional information retrieval, students need to access multiple information sources—multiple web pages returned by search engines, different research reports, mutually contradictory viewpoints—this process, although time-consuming, cultivates the ability to assess information quality, identify information conflicts, and synthesize diverse perspectives. AI simplifies this complex process into a single output: users ask questions, the system provides a seemingly comprehensive but actually highly summarized answer. While this simplification improves efficiency, it also deprives students of opportunities to encounter information diversity.

Another important mechanism discovered by White et al. (2023) in prompt engineering research: when different users use similar prompt structures, DeepSeek R1-generated business plan topics are highly concentrated in the top 10% high-frequency themes. This is because most users search online for "best prompt templates," and these widely disseminated templates are themselves highly similar. Interview verification of experimental group students confirms this: 76.7% of students admitted to searching online platforms (mainly Zhihu, Bilibili, and Xiaohongshu) for "how to make DeepSeek write business plans" and used recommended prompt templates. This leads to standardized inputs producing standardized outputs, standardized outputs being shared as new standard templates, further intensifying input standardization.

From an information theory perspective, AI tools reduce information entropy. Shannon's (1948) concept of information entropy measures system uncertainty or diversity. This study calculated information entropy for project topic selection in both groups, with experimental group entropy of 2.18, significantly lower than control group's 2.87 ($p<0.001$), indicating AI intervention indeed led to a systematic decline in entrepreneurial idea diversity.

4.2.2 Cognitive mechanism: shallow learning and cognitive authority transfer

Technical factors are only surface-level problems; deeper driving forces come from changes in students' cognitive patterns and learning strategies. Cognitive load theory provides an understanding framework. Sweller et al. (2019) pointed out that learning requires investing limited cognitive resources for information processing, and when external tools excessively reduce cognitive load, deep learning may be weakened.

Table 4 shows comparison of time investment in students' learning processes between the two groups.

Table 4 Comparison of time investment in students' learning process (N=30)

Learning Stage	Experimental Group (M±SD)	Control Group (M±SD)	Statistical Test[4]
Total Duration (hours)[1]	14.7 ± 4.3	32.4 ± 8.6	$t = 9.87$, $p < 0.001$**
User Interviews[2]	0.3 ± 0.8	8.2 ± 3.1	$U = 32.5$, $p < 0.001$**
AI Interaction[3]	7.8 ± 2.1	0.2 ± 0.5	$U = 28.0$, $p < 0.001$**
Information Gathering/ Supplementation	4.2 ± 1.7	12.6 ± 4.3	$t = 8.23$, $p < 0.001$**
Plan Writing/Revision	2.4 ± 1.2	11.4 ± 3.8	$t = 10.45$, $p < 0.001$**

Note: [1] Data source: Randomly selected 15 students from each of experimental and control groups (30 total), requiring them to fill out daily time logs during business plan completion, ultimately collecting 28 valid logs (14 experimental group, 14 control group). Table presents data from 28 students. [2] User interviews: In experimental group, 2 of 14 students conducted brief interviews (0.5 hour and 1 hour), remaining 12 had 0 hours, thus average is close to but not equal to 0. Mann-Whitney U test used due to severe right skewness of data. [3] AI interaction: In control group, 1 of 14 students used AI to query industry data (0.5 hour), thus average is 0.2 hours. [4] Statistical methods: Total duration, information gathering, and plan writing used independent samples t-test; user interviews and AI interaction used Mann-Whitney U non-parametric test due to severe distribution skewness (skewness>2).

Table 4 shows that experimental group students' time to complete business plans (14.7±4.3 hours) was less than half that of control group (32.4±8.6 hours). More noteworthy is the difference in time allocation structure: the control group spent about 25% of time on user interviews and 39% on information gathering, while the experimental group had almost no user interview stage, with over half the time (53%) spent on AI interaction. Behind time investment differences are fundamental differences in learning process depth.

Craik and Lockhart's (1972) levels of processing theory distinguishes between shallow processing (focusing on surface features) and deep processing (focusing on semantic and logical connections). Interview data shows students using AI generally exhibit shallow learning characteristics. A University B student (S-B16) described: "*I first asked DeepSeek about this project's market size, it gave me a paragraph, and I copied it down. Then I asked what competitors there are, it gave a list. I combined this content, adjusted the format, and basically it was done.*" When asked "what's your own judgment of market size," he frankly said: "*I can't really say clearly, just feel what AI said should be right.*"

The core problem of this learning pattern is lack of metacognitive monitoring. Metacognition refers to cognition and regulation of one's own cognitive processes (Flavell, 1979), including planning learning strategies, monitoring comprehension level, and evaluating learning outcomes. This study used the Metacognitive Awareness Inventory (MAI) developed by Schraw and Dennison (1994) to measure students' metacognitive levels. Table 5 shows experimental group students' metacognitive scores (52.3±8.7) were significantly lower than control group (64.8±9.2, t=7.82, p<0.001, d=1.42, large effect), especially with most obvious differences in "monitoring" and "evaluation" dimensions (d=1.50 and 0.85).

Table 5 Comparison of metacognitive levels between two groups (N=60)

Dimension	Experimental Group (n=30) M±SD	Control Group (n=30) M±SD	t-value	p-value	Cohen's d	Effect Size
Total Metacognition Score	52.3±8.7	64.8±9.2	7.82	<0.001**	1.42	Large
Planning	17.8±3.2	21.4±3.5	4.63	<0.001**	1.07	Large
Monitoring	16.2±3.8	22.1±4.1	6.54	<0.001**	1.50	Very large
Evaluation	18.3±3.4	21.3±3.6	3.67	0.001**	0.85	Large

*Note: **p<0.01; MAI scale total score range 0-100, higher scores indicate higher metacognitive levels*

Figure 4 uses radar chart and bar chart forms to intuitively present inter-group differences in metacognitive abilities. The radar chart (Figure 4b) clearly shows that the control group forms a larger coverage area (blue region) across three dimensions, while the experimental group's coverage area noticeably contracts (red region), visually confirming the comprehensive weakening of experimental group metacognitive abilities. The bar chart (Figure 4a) further quantifies the significance of these differences: all three dimensions have p-values <0.001 and are marked with ** (highly significant) symbols, indicating these are not chance fluctuations. Particularly noteworthy is the "monitoring" dimension (Monitoring) with a very large effect size (d=1.50), which is the largest difference among all dimensions. Monitoring ability refers to real-time assessment of one's comprehension level during the learning process—"Do I really understand?" "Is this answer reasonable?" When students treat AI as an "answer provider" rather than a "thinking tool," this habit of self-questioning and monitoring deteriorates. This explains why experimental group students can submit seemingly professional business plans yet cannot answer when asked about details (like S-B09's case in section 4.1.4).

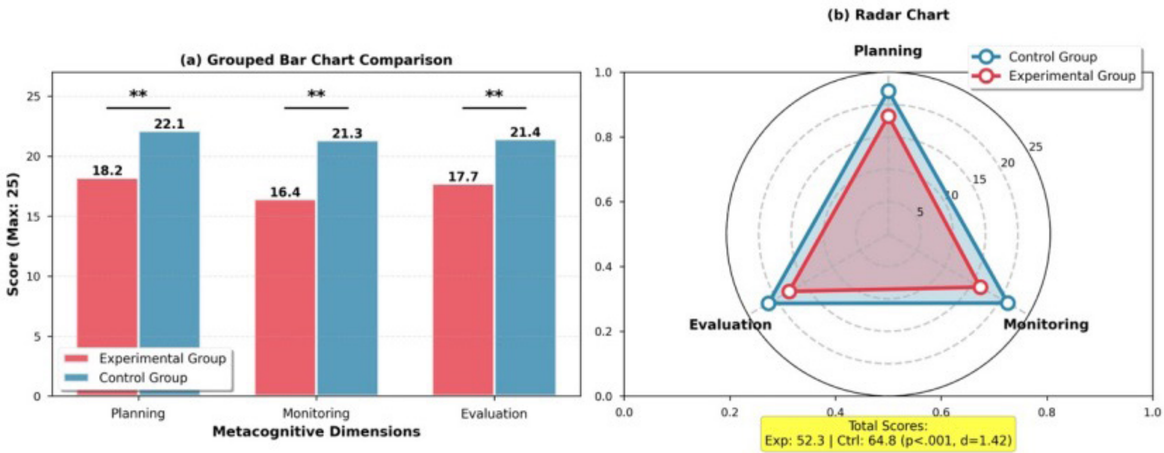

Figure 4 Metacognitive ability comparion across three dimensions (MAI scale scores, N=60)

When AI becomes an information intermediary, the entire learning process is "black-boxed"—students input questions, get answers, but are unclear about how these answers are generated, what logic they're based on, and what limitations they have (Aiken & Epstein, 2023). This lack of transparency makes it difficult for students to conduct effective metacognitive monitoring of AI outputs. A University A student's (S-A22) reflection is quite representative: "*After using DeepSeek, efficiency definitely improved, but I found I'm increasingly uncertain* *about which parts I thought of myself and which AI gave me. Sometimes I doubt whether what AI said is right, but I don't know how to verify it, so I just think 'forget it, it should be close enough.'*"

Cognitive authority transfer is another key mechanism. Traditionally, students viewed teachers, textbooks, and industry experts as knowledge authorities. But Cowen's (2024) research found that AI's emergence blurred this authority structure, with students tending to unconditionally trust AI-produced content, especially when this content is presented in professional, authoritative tones. This study designed an experimental task: showing 15 students using AI two analyses of the same market, one from AI and one from industry experts (actually both written by researchers and containing identifiable errors). Results showed 80% of students believed AI's analysis was "more professional" and "more credible," mainly reasoning being "expression is clearer" and "more data." Only 3 students identified errors in both analyses, and all 3 students expressed they would simultaneously reference multiple information sources rather than completely relying on a single source.

This over-trust partly stems from "automation bias"—people tend to trust automated system decisions while ignoring contradictory evidence (Goddard et al., 2012). In entrepreneurship education contexts, this bias is particularly dangerous because entrepreneurial decisions often need to be made under conditions of uncertainty and incomplete information, where single information source reliability is inevitably limited. When students treat AI as their primary or even sole "consultant," they actually forfeit the most valuable learning opportunity—forming independent judgment in the clash of diverse perspectives.

Shepherd et al. (2023) provide supplementary explanation from a behavioral economics perspective. They point out that when facing uncertainty, individuals tend to adopt "imitation strategies," choosing paths validated by most people to reduce risk. This study's findings are highly consistent with this theory: students view the "mainstream" directions recommended by AI as safe choices that have been verified, thus forming a bandwagon effect. This explains why even when different students independently use AI, project topic selection still shows high convergence—they are all following what AI represents as "collective wisdom." AI's recommendation of "mainstream" entrepreneurship directions reinforces this bandwagon psychology because students reason: "If AI recommends this, it means many people are doing it, so it should be feasible." Although this heuristic decision-making is effective in some situations, it may be harmful in entrepreneurship contexts because truly innovative opportunities often exist in areas not yet discovered by most people (Kirzner, 1973).

4.2.3 Educational ecosystem mechanism: assessment orientation and structural constraints

As Neck et al. (2024) criticized, this assessment orientation actually encourages students to produce texts that "look like entrepreneurship" rather than cultivating real entrepreneurial capabilities. Interviews with 15 instructors reveal this contradiction. A University B instructor (T-B04) frankly stated: "*I know I should assess students' thinking processes and practical abilities, but with 80 people in a class, I can't have in-depth exchanges with every student. In the end, I still need to look at their submitted business plans. If a business plan is written very professionally, with complete structure, sufficient data, it's also hard for me to say it's not good, even though I vaguely feel some content might be AI-generated.*"

This reflects the absence of "constructive alignment" proposed by Biggs (1996): teaching objectives, teaching activities, and assessment methods should align with each other. If entrepreneurship education objectives are to cultivate students' practical abilities and innovative thinking, but assessment still mainly relies on standardized texts, students will naturally rationally choose the most efficient way to complete assignments—using AI. A University A student (S-A28) bluntly said: "*If the teacher really wants to see whether we did user interviews, they should ask us to submit interview recordings or narrate user stories on-site. But they only look at the final business plan, so why wouldn't I use AI to improve efficiency?*"

Insufficient faculty preparation exacerbates this problem. Fütterer et al.'s (2023) survey of European universities shows only 28% of instructors received specialized training on how to respond to AI intervention. This study's interviews show that the situation in Chinese universities is similar or even more severe. Among 15 instructors, only 2 (13.3%) had participated in AI-related teaching training, both self-paid for off-campus workshops. Most instructors, although able to sense "abnormalities" in student assignments, lack effective identification and response strategies.

A University C instructor (T-C05) shared confusion: "*I can tell some business plans might be AI-written because the language is too 'perfect,' unlike student writing. But I have no evidence and don't know how to handle it. I tried using AI detection tools, but results are very unstable.*" This dilemma is supported by Elkhatat et al.'s (2023) research, who tested multiple AI detection tools and found accuracy rates generally below 70% with high false positive rates, making them difficult to use as reliable judgment bases.

Neck and Greene (2011) criticize traditional courses for overemphasizing theory while neglecting exposing students to real markets. Among the three universities in this study, only University A requires at least 10 user interviews, while Universities B and C have no mandatory practical requirements. Interviews show that students required to conduct field research had noticeably lower assignment homogenization levels. Mansoori et al.'s (2023) tracking research of lean startup teaching methods found that when students are required to experience "build-measure-learn" iterative cycles and continuously adjust plans based on real user feedback, their entrepreneurial ability improvement is significant, and project originality and feasibility are higher. This suggests that if course design can make practical stages core rather than supplementary, it may effectively reduce over-dependence on AI. However, practice-oriented teaching requires more time, resources, and instructor investment, creating a contradiction with current scalability pressures universities face.

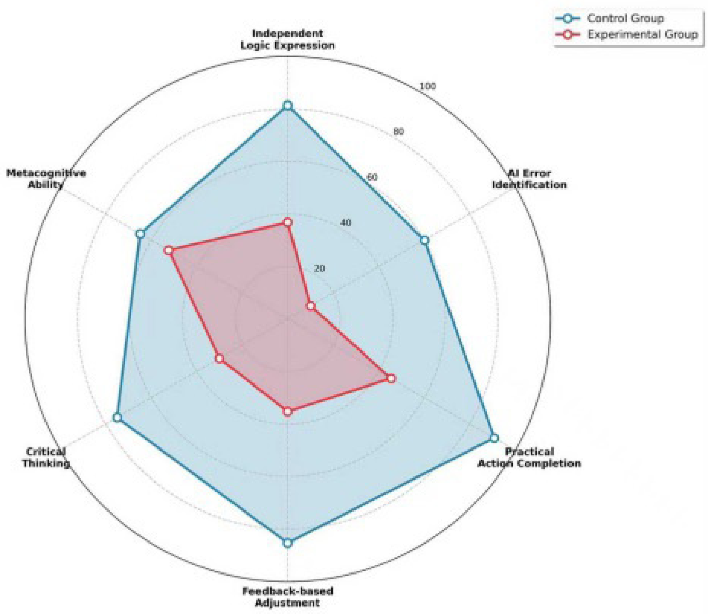

Figure 5 Core entrepreneurial competency comparison (skills hollowing-out effect)

In this study, University A's average class size is 78 people, University B 95 people, and University C 112 people. A University B instructor T-B02: "*I have to manage over 200 students per semester, there's simply no time to communicate in-depth with each team*." This scalability pressure leads to a vicious cycle: large class sizes → no time for personalized guidance → reliance on standardized assessment → students rationally use AI → assignment homogenization → declining teaching effectiveness.

Figure 5 comprehensively displays through radar chart the systematic impact of AI intervention on six core entrepreneurial competencies, clearly presenting the "skills hollowing-out" phenomenon. The outer blue area represents the control group, the inner red area represents the experimental group, with significant area differences directly reflecting the severe degree of competency gap.

4.3 Direct impact on core objectives of entrepreneurship education

4.3.1 Homogenization tendency of innovative thinking

The core of innovative thinking lies in generating novel and valuable ideas (Amabile, 1982). This study used an adapted Creativity Assessment Scale to conduct blind evaluation of 50 randomly selected business plans (25 experimental group, 25 control group), with 5 entrepreneurship education experts scoring from three dimensions: novelty, value, and feasibility (1-7 points).

Table 4-3a Comparison of business plan creativity evaluation results

Dimension	Experimental Group (M±SD)	Control Group (M±SD)	t-value	p-value	Cohen's d	Effect Size
Novelty	3.2 ± 0.9	5.1 ± 1.2	6.54	<0.001**	1.85	Very Large
Value	4.6 ± 0.8	5.3 ± 1.0	2.89	0.006**	0.82	Large
Feasibility	5.4 ± 0.7	4.8 ± 1.1	2.41	0.020*	0.68	Medium to Large
Comprehensive Score	4.4 ± 0.6	5.1 ± 0.9	3.45	0.001**	0.98	Large

*Note: *p<0.05, **p<0.01; Inter-rater reliability ICC=0.84*

Data shows the experimental group's novelty scores were significantly lower than the control group (d=1.85, very large effect), highly consistent with the homogenization phenomenon. Expert comments reveal the problem: "*Plans in the experimental group are mostly replications or minor adjustments of existing business models, lacking unique perspectives*" (Expert Reviewer E1); "*Many business plans seem to come from the same template, just changing industries or products, but the underlying logic is all the same*" (Expert Reviewer E3).

The experimental group scored slightly higher on feasibility, but this reflects conservatism rather than advantage—tending to choose validated mature models, avoiding high-risk but potentially high-return innovation fields. Expert Reviewer E4: "*High feasibility often means low entry barriers and intense competition. Real opportunities often exist in seemingly 'not very feasible' but uniquely insightful fields.*"

4.3.2 Systematic absence of critical thinking

Critical thinking—the ability to question, analyze, and evaluate information—is the foundation for entrepreneurs to make wise decisions in uncertain environments (Neck & Greene, 2011). The research evaluated AI's impact on critical thinking through three dimensions.

Interviews with 30 experimental group students showed only 3 (10.0%) would actively verify data or viewpoints provided by AI. When asked "if AI data conflicts with your observations, what would you do," 18 (60.0%) chose "trust AI because it's more authoritative," 9 (30.0%) "would have some doubt but don't know how to judge," and only 3 (10.0%) "would find more materials for cross-validation."

Critical thinking test task showed students a business analysis containing logical flaws ("China's coffee market annual growth rate is 15%, indicating that opening a coffee shop will definitely make money"), asking them to identify problems. Control group average identified 3.8±1.2 logical issues (such as market growth not equaling individual profitability, not considering competitive saturation, not analyzing regional differences), while experimental group only identified 1.4±0.9 ($t=9.23$, $p<0.001$, $d=2.32$, very large effect).

During interviews, a University A instructor's (T-A04) observation is quite enlightening: "*Previously students might ask me 'Teacher, this industry report says the market is large, but I went to observe on-site and felt it wasn't that bustling, is there a problem with the data?'* This questioning spirit is very valuable. But now students rarely have such questions, they're more inclined to put what AI says and what reports write directly into their business plans, without thinking whether this information applies to their specific situation."

4.3.3 Practical ability and knowledge-action gap

The ultimate goal of entrepreneurship education is cultivating students' practical ability—the ability to turn ideas into actions (Neck & Greene, 2011). However, data shows AI intervention led to serious "knowledge-action gap."

Course required all teams to "take at least one actual action to advance project" over two weeks. Among control group's 75 teams, 68 (90.7%) completed, averaging 2.4±1.1 actions per team; among experimental group's 75 teams, only 34 (45.3%) completed, averaging 0.8±0.6 actions ($\chi^2=32.78$, $p<0.001$, Cramér's $V=0.47$; $t=9.87$, $p<0.001$, $d=1.87$, very large effect).

University C student S-C18: "*Our business plan was written quite well, AI helped us refine all parts. But when the teacher required actually doing it, we suddenly didn't know where to start. The plan says to do user research, but specifically how to find users, what to ask, how to analyze, we're all unclear*."

When actual situations don't match plans, among control group's 68 teams that completed actions, 58 (85.3%) could adjust plans based on feedback and continue trying; among experimental group's 34 teams that completed actions, only 12 (35.3%) could effectively adjust, 22 (64.7%) stagnated after encountering the first obstacle ($\chi^2=21.45$, $p<0.001$, Cramér's $V=0.46$).

University B instructor T-B05: "*Students using AI seem more 'fragile.'* They have mysterious confidence in business plan plans because it's 'AI-recommended.' But once reality doesn't match expectations, they're at a loss because they haven't experienced the process of conceiving, trial-and-error, and adjusting." Kolb's (1984) experiential learning theory emphasizes that true learning requires experiencing the complete cycle of "concrete experience-reflective observation-abstract conceptualization-active experimentation." AI intervention causes students to skip "concrete experience" and "reflective observation," directly obtaining "abstract concepts," this incomplete cycle cultivates "armchair strategist" abilities rather than real entrepreneurial qualities.

4.3.4 From "skills hollowing-out" to goal alienation

AI intervention's impact on core entrepreneurship education objectives can be summarized as "skills hollowing-out": students acquire surface skills of producing professional texts but lose the underlying abilities supporting these texts. Table 4-3b summarizes the degree of core competency damage, with all indicators showing large or very large effects.

Table 4-3b Summary of core competency damage levels

Core Competency	Key Indicator	Experimental vs Control	Effect Size	Damage Level
Opportunity Recognition	Proportion of topic selection based on personal observation	10.7% vs 42.7%	V = 0.56	Severe
Innovative Thinking	Novelty evaluation score	3.2 vs 5.1	d = 1.85	Severe
Critical Thinking	Number of logical problems identified	1.4 vs 3.8	d = 2.32	Extremely severe
Practical Ability	Proportion completing actual actions	45.3% vs 90.7%	V = 0.47	Severe
Dealing with Uncertainty	Proportion continuing adjustment after setbacks	35.3% vs 85.3%	V = 0.46	Severe

These data point to a grim reality: AI intervention not only changes learning methods but may lead to alienation of entrepreneurship education objectives—from cultivating "entrepreneurs who can identify opportunities and create value in uncertain environments" to regressing to training "AI users who can generate standardized business texts."

5. Discussion

5.1 Substantive impact of AI intervention on entrepreneurial competency cultivation

The large effect size in financial projection detail level (d=1.38) reveals an important pattern: students who used AI tools demonstrated surface professionalism but showed weaker underlying analytical abilities. However, this observed association does not establish that AI use caused this deficit. Alternative explanations warrant consideration, including the possibility that students with weaker analytical skills were more likely to adopt AI-reliant approaches, or that differences in pedagogical support between cohorts contributed to the observed patterns. Although experimental group students can produce business plans with more standardized structures (81.3% follow standard structure), they are noticeably more superficial in sections requiring deep analysis.

Comparison of learning processes better illustrates the problem. Control group students averaged 32.4 hours investment, including 8.2 hours on user interviews and 12.6 hours on information gathering; experimental group only 14.7 hours, of which 7.8 hours spent on AI interaction, with almost no user interview stage. The absence of user cases and primary data (only 10.7% and 16.0%) means most students skipped the most valuable learning stage—contact with real users and market observation. This is not just a difference in time allocation but a fundamental difference in learning depth.

The absence of judgment ability is equally concerning. 63.3% of students cannot fluently express project logic without AI assistance, indicating they can submit professional business plans but haven't experienced the thinking process of forming judgments. Metacognitive ability differences (52.3 vs 64.8) show students lack effective monitoring of comprehension level during learning. Metacognitive activities requiring judgment of information reliability and identification of knowledge gaps in traditional learning are simplified to accepting AI output. In the long term, this may affect their self-regulation abilities in real entrepreneurship.

These findings reveal a concerning trend: AI intervention not only failed to alleviate entrepreneurship education's overemphasis on form but actually exacerbated the disconnection between form and ability. Business plans became more refined in form, yet students moved further from real entrepreneurial abilities.

5.2 Hidden mechanisms of AI use exacerbating educational inequality

This study reveals a phenomenon easily overlooked: although all students can equally use AI, usage methods and effects have significant differences, which may exacerbate rather than alleviate educational inequality.

Students who can identify AI errors and actively verify account for only 10%, and they all have actual entrepreneurship experience or internship backgrounds. This indicates only students who have already established independent judgment frameworks can effectively use AI as an auxiliary tool. Students lacking this background are more likely to completely depend on AI, yet this dependence precisely prevents them from establishing independent judgment abilities.

Language style differences provide another perspective. The experimental group extensively uses professional terminology (frequency 3-4 times that of control group), but most students cannot accurately explain meanings. This "pseudo-professionalization" affects students from different backgrounds differently: students with opportunities to access real business environments will eventually understand concepts in practice; students lacking opportunities may become fixed in the pattern of packaging ideas with terminology while losing the ability to truly understand problems.

The hidden nature of this inequality lies in assessment mechanisms' concealment. When assessment primarily relies on business plans, students using AI receive higher scores due to more refined text forms, masking real competency gaps. More seriously, students with weaker abilities mistakenly believe they already possess entrepreneurial capabilities, only discovering huge gaps when actually practicing. From a social mobility perspective, if disadvantaged background students lose critical thinking due to over-reliance on AI, they will face double disadvantages: both lacking resource networks and lacking abilities to deal with uncertainty, weakening entrepreneurship's function as a social mobility channel.

5.3 Systematic loss of innovation ecosystem diversity

Information entropy analysis shows AI intervention led to approximately 24% loss in topic selection diversity, revealing a systematic trend: when more and more students rely on the same AI tools, entrepreneurial ideas are converging.

High-frequency word distribution clearly presents this convergence. Experimental group high-frequency words highly concentrate in generic concepts like "platform," "user," "data" (accounting for 68.3%), while control group uses more geographically and individually distinctive words like "community," "handcraft," "rural" (generic concepts account for only 41.7%). This is not just language difference but reflects differences in thinking modes and problem frameworks.

Project topic selection distribution further confirms this trend. In fields requiring local knowledge like agricultural technology, cultural creativity, and local life services, experimental group project proportions significantly dropped (from 14.7% to 4.0%, from 12.0% to 2.7%). This is not because these fields lack opportunities but because AI training data mainly comes from mainstream fields, lacking sufficient "cognition" of niche or local fields, leading to systematic neglect of these fields.

Long-term risks brought by this convergence deserve attention. Many disruptive innovations initially came from "margins"—the pioneers of personal computer revolution were not IBM but garage enthusiasts, sharing economy was initially viewed as impractical. But when a generation of students' ideas are all shaped by the same AI data, who will explore currently "non-mainstream" but potentially future-nurturing fields? 24% diversity loss means approximately one-quarter of potential innovation paths are collectively abandoned at the conception stage.

What needs more vigilance is self-reinforcement: AI recommends mainstream → students choose mainstream → successful cases concentrate in mainstream → AI further reinforces mainstream bias. In the long term, this may lead to surface prosperity but actual lack of innovative diversity, with large numbers of students competing homogeneously in already fully competitive mainstream markets.

There exists a paradox here: AI improved individual student output average quality, but group heterogeneity is declining. If innovation ecosystem value lies more in diversity rather than average level, then AI intervention may be trading short-term efficiency improvements for long-term innovation vitality. This trade-off requires serious consideration and response from entrepreneurship educators.

6. Research Limitations and Future Directions

This study has the following limitations: First, samples concentrate in three eastern universities, lacking data from central and western regions; disciplines mainly focus on business administration and economics; time span only 9 months, unable to track long-term impacts. Second, models observed based on 2023-2024 may change with technology iteration causing different homogenization manifestations. Third, mainly adopting qualitative and semi-quantitative analysis, future research can develop more refined quantitative indicators such as semantic similarity, conceptual network analysis, and creativity uniqueness scoring (Amabile, 1982). Fourth, cross-sectional study cannot track long-term impacts, requiring follow-up of students who extensively used AI in subsequent entrepreneurship practices (Pittaway & Cope, 2007). Fifth, lacking validation of response strategy effectiveness.

Future research directions: (1) Explore moderating effects of individual difference variables on homogenization; (2) Compare response strategies across different countries and cultural backgrounds; (3) Develop entrepreneurship education assessment frameworks adapted to AI era; (4) Research from neuroscience perspective how AI changes cognitive processing.

7. Conclusion

Through systematic analysis, this study found that under AI intervention background, undergraduate entrepreneurship course outcomes exhibit significant homogenization, mainly manifested in four dimensions: project topic clustering, templatized business logic, converging data citations, and standardized language style. This results from interaction of factors at three levels: technical (training data bias, information flattening), cognitive (shallow learning, cognitive authority transfer), and educational ecology (standardized assessment, insufficient faculty, scalability pressure). Homogenization poses substantive challenges to entrepreneurship education core objectives: causing “skills hollowing-out,” making students appear professional superficially but lack underlying abilities; may exacerbate educational inequality, with weaker foundation students more likely to become passive AI consumers; threatens innovation ecosystem diversity, compressing marginal and breakthrough innovation exploration space. This study’s theoretical contribution lies in constructing a multi-level analytical framework integrating technology, cognition, and ecosystem, showing AI educational impact is highly context-dependent, depending on user ability, usage methods, and educational system support conditions. At the practical level, entrepreneurship education should guide students to use AI responsibly through systematic reforms. The core lies in repositioning educational value: from “learning to write business plans” to “cultivating abilities to discover opportunities and solve problems in uncertainty”; from “mastering standardized frameworks” to “developing critical thinking and independent judgment abilities”; from “producing professional texts” to “conducting real market exploration and user insights.” In the intelligent era, entrepreneurship education’s unique value lies in cultivating humanistic abilities AI finds difficult to replace—empathy for user needs, sensitivity to business ethics, judgment in complex situations, ability to collaborate with others, and resilience facing failure. These abilities require tempering in real, uncertainty-filled situations and cannot be achieved through simple information input and text generation. AI-era entrepreneurship education stands at crossroads. Responding appropriately, AI can become a tool liberating instructor energy and supporting student exploration; leaving it unchecked may lead to declining educational quality and talent cultivation goal alienation. This study provides theoretical framework and empirical evidence for understanding this challenge, hoping more researchers and practitioners will jointly construct entrepreneurship education new paradigms adapted to intelligent era—embracing technological progress opportunities while adhering to education’s fundamental mission of cultivating people; improving teaching efficiency while ensuring learning depth; cultivating students’ abilities to use tools and more cultivating their wisdom to become tools’ masters.

Funding

There is no funding to report.

Conflicts of Interest

The author(s) declare no conflicts of interest regarding the publication of this paper.

Ethics Statement

Not applicable.

References

Aiken, R. M., & Epstein, R. G. (2023). Ethical implications of AI in education. *Communications of the ACM*, *66*(7), 46-54. https://doi.org/10.31305/rrjss.2025.v05.n01.001

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997-1013. https://doi.org/10.1037/0022-3514.43.5.997

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52-62. https://doi.org/10.61969/jai.1337500

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. https://doi.org/10.1145/3442188.344592

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347-364. https://doi.org/10.1007/BF00138871

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brown, T., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Kasneci, E., Seßler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Cowen, T. (2024). How AI will reshape education and credentialing. *Journal of Economic Perspectives*, *38*(1), 127-148. https://doi.org/10.13140/RG.2.2.14907.12327

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. https://doi.org/10.1016/S0022-*5371*(72)80001-X

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(22), 1-22. https://doi.org/10.1186/s41239-023-00392-8

Dwivedi, Y. K., Kshetri, N., Hughes, L., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, *19*(1), 17. https://doi.org/10.1007/s40979-023-00140-5

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. arXiv preprint arXiv:2303.10130. https://doi.org/10.48550/arXiv.2303.10130

Fayolle, A., & Gailly, B. (2008). From craft to science: Teaching models and learning processes in entrepreneurship education. *Journal of European Industrial Training*, *32*(7), 569-593. https://doi.org/10.1108/03090590810899838

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Fütterer, T., Fischer, C., Alekseeva, A., et al. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, *13*(1), 15310. https://doi.org/10.1038/s41598-023-42227-6

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121-127. https://doi.org/10.1136/amiajnl-2011-000089

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7), 692. https://doi.org/10.3390/educsci13070692

Hägg, G., & Kurczewska, A. (2022). Guiding the student entrepreneur – Considering the emergent adult. *Education + Training*, *64*(6), 759-777. https://doi.org/10.1108/ET-03-2020-0069

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385-16389. https://doi.org/10.1073/pnas.0403723101

Hosseini, M., Gao, C. A., Liebovitz, D. M., et al. (2023). An exploratory survey about using ChatGPT in education, healthcare, and research. *PLOS ONE*, *18*(10), e0292216. https://doi.org/10.1371/journal.pone.0292216

Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. *International Journal of Human-Computer Interaction*, *30*(3), 177-191. https://doi.org/10.1080/10447318.2013.846790

Kasneci, E., Seßler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kirzner, I. M. (1973). Competition and entrepreneurship. University of Chicago Press.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice Hall.

Koltay, T. (2015). Data literacy: In search of a name and identity. *Journal of Documentation*, *71*(2), 401-415. https://doi.org/10.1108/JD-02-2014-0026

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49, 847 - 864. https://doi.org/10.1080/03075079.2024.2323593

Makridakis, S. (2024). The forthcoming artificial intelligence revolution: Its impact on society and firms. *Futures*, 158, 103348. https://doi.org/10.1016/j.futures.2017.03.006

Mansoori, Y., Karlsson, T., & Lundqvist, M. (2023). The influence of the lean startup methodology on entrepreneur-coach relationships in accelerators. *Technovation*, 120, 102450. https://doi.org/10.1016/j.technovation.2019.03.001

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge University Press. https://doi.org/10.2307/40324289

Neck, H. M., & Greene, P. G. (2011). Entrepreneurship education: Known worlds and new frontiers. *Journal of Small Business Management*, *49*(1), 55-70. https://doi.org/10.1111/j.1540-627X.2010.00314.x

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. The Wharton School Research Paper, April 17. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4391243

Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory and discussion. *ACM Journal on Responsible Computing*, *1*(1), 1-21. https://doi.org/10.1145/3597307

Neck, H. M., & Greene, P. G. (2011). Entrepreneurship education: Known worlds and new frontiers. *Journal of Small Business Management*, *49*(1), 55-70. https://doi.org/10.1111/j.1540-627X.2010.00314.x

Pittaway, L., & Cope, J. (2007). Entrepreneurship education: A systematic review of the evidence. *International Small Business Journal*, *25*(5), 479-510. https://doi.org/10.1177/0266242607080656

Page, S. E. (2007). The difference: How the power of diversity creates better groups, firms, schools, and societies. Princeton University Press. https://doi.org/10.2307/j.ctt7sp9c

Parker, S. C. (2018). The economics of entrepreneurship (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781316756706

Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, *20*(2), 07. https://doi.org/10.53761/1.20.02.07

Pittaway, L., & Cope, J. (2007). Entrepreneurship education: A systematic review of the evidence. *International Small Business Journal*, *25*(5), 479-510. https://10.1177/0266242607080656

Ratten, V., & Jones, P. (2021). Covid-19 and entrepreneurship education: Implications for advancing research and practice. *International Journal of Management Education*, *19*(1), 100432. https://doi.org/10.1016/j.ijme.2020.100432

Rudolph, J., Tan, S., & Tan, S. (2023). *ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching*, *6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Sarasvathy, S. D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, *26*(2), 243-263. https://doi.org/10.2307/259121

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460-475. https://doi.org/10.1006/ceps.1994.1033

Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. Harvard University Press. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4499769

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepherd, D. A., Williams, T., Wolfe, M., & Patzelt, H. (2023). Learning from entrepreneurial failure: Emotions, cognitions, and actions. *Academy of Management Review*, *48*(2), 308-333.

Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*(1), 5-14. https://doi.org/10.3102/0013189X08314117

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, *6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Sweller, J., Van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261-292. https://doi.org/10.1007/s10648-019-09465-5

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, *379*(6630), 313. https://doi.org/10.1126/science.adg7879

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

White, J., Fu, Q., Hays, S., et al. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382. https://doi.org/10.48550/arXiv.2302.11382