Journal of Engineering and Technological Advances (JETA)
https://jeta.segi.edu.my/index.php/segi
2025 Vol. 10 No. 2

A SYSTEMATIC LITERATURE REVIEW ON AI AND ANALYTICS TOOLS FOR PREDICTING ENTREPRENEURSHIP ATTRIBUTES IN MALAYSIA'S HIGHER EDUCATION INSTITUTIONS: DEVELOPING A CONTEXTUAL TAXONOMY

[1]Ghazali, A.J.*, [1]Ismail, N.A., [2]Nik Muhammad N.M, [3]Mat Som, H.

[1] Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Pengkalan Chepa, 16100, Kelantan, Malaysia.

[2] Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Pengkalan Chepa 16100, Kelantan, Malaysia.

[3] Faculty of Science and Information Technology, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia.

* Corresponding Author: a20e0138p@siswa.umk.edu.my TEL: (6019)-2025468

Received: 15 August 2025; Accepted: 1 September 2025; Published: 31 December 2025

doi: 10.35934/segi.v10i2.131

Highlights:

- Systematic Literature Review (SLR) of AI Tools
- Novel Taxonomy of Tools and Attributes
- Focus on Malaysian HEIs

Abstract: Entrepreneurship drives economic and social progress, yet predicting key attributes such as cognitive (e.g., creativity, opportunity recognition), emotional (e.g., resilience, emotional intelligence), and social (e.g., leadership, communication) traits remains challenging due to the limitations of traditional methods. Traditional methods lack precision, driving the adoption of artificial intelligence (AI)-based data analytics tools to analyse complex datasets. This paper employs a Systematic Literature Review (SLR) to synthesise 50 peer-reviewed studies from 2018-2024, using thematic analysis to categorise AI-based data analytics tools into supervised learning, unsupervised learning, natural language processing (NLP), and deep learning, mapping their applications to entrepreneurship attributes. Findings highlight the dominance of models like Random Forests and Transformers in predicting entrepreneurship attributes, with gaps in exploring non-financial attributes such as creativity and resilience, and ethical considerations, including data bias and privacy concerns. This study investigates opportunities and challenges for Malaysian Higher Education Institutions (HEIs), emphasising personalised education and research advancements against barriers such as bias and resource constraints. Contributions include a novel taxonomy that maps AI-based data analytics tools to entrepreneurship attributes and incorporates an ethical evaluation dimension, alongside practical insights for stakeholders and a research agenda for ethical, inclusive AI adoption in entrepreneurship globally and in Malaysia.

Keywords: artificial intelligence, data analytics, entrepreneurship, predictive modelling, systematic literature review

1. Introduction

Entrepreneurship is a catalyst for innovation, job creation, and societal advancement, pivotal to global and regional economies like Malaysia's, which aspire to lead in the digital economy (Saryani et al., 2022). Successful entrepreneurs exhibit attributes of risk-taking, creativity, leadership, resilience, communication, emotional intelligence, opportunity recognition, and adaptability that enable them to navigate uncertainty (Feng et al., 2023). However, assessing these traits is a complex process. Traditional assessments, such as psychometric tests, interviews and financial metrics, are limited by subjectivity, limited scalability, and inability to capture dynamic behaviours or underlying qualities, as evidenced by their reliance on subjective self-reported surveys in Malaysian HEIs (Nasreen et al., 2024) and failure to account for non-financial risks in diverse entrepreneurial contexts (Dhochak et al., 2024; Halim et al., 2022), prompting the rise of AI-based data analytics tools (Pendy, 2023).

AI tools, including Machine Learning (ML), Natural Language Processing (NLP), and deep learning, process diverse datasets such as financial records, pitch transcripts, social media, or videos to accurately predict attributes, with Random Forests applied to risk-taking (Lukita et al., 2023) and BERT analysing communication skills (Hossain et al., 2022). For instance, supervised learning quantifies risk tolerance, while NLP evaluates communication. These tools benefit stakeholders: investors identify high-potential founders, accelerators scout talent, educators tailor curricula, and policymakers strengthen ecosystems, particularly in Malaysia's tech hubs like Cyberjaya. Yet, the literature lacks a systematic synthesis of these tools, with studies often focusing on startup outcomes (e.g., revenue) rather than underlying attributes (Khalid, 2020). Ethical issues, bias, privacy and further limit adoption, especially in resource-constrained contexts like Malaysian HEIs.

This paper addresses these gaps through a Systematic Literature Review (SLR), guided by four questions: (1) What AI-based data analytics tools predict entrepreneurship attributes? (2) Which attributes do they target, and how? (3) What are the strengths and limitations of AI-based data analytics tools in predicting entrepreneurship attributes? (4) What are the implications of these tools for entrepreneurship education and research in Malaysian HEIs? Our objectives are to categorise tools, map their applications, evaluate performance, and explore possibilities and threats for Malaysian HEIs. Contributions include a taxonomy that categorises AI-based data analytics tools by type and maps them to behavioural entrepreneurship attributes with an ethical evaluation dimension, alongside practical insights for stakeholders and a research agenda for ethical AI adoption. The paper is structured as follows: Section 2 reviews prior work, Section 3 details the SLR methodology, Section 4 analyses tools, Section 5 examines Malaysian HEI implications, and Section 6 concludes.

2. Literature Review

2.1. Entrepreneurship Attributes

Entrepreneurship attributes are cognitive, behavioural, and emotional traits driving venture success. (Bahari et al., 2023) identify core traits of entrepreneurship: risk-taking (decision-making under uncertainty), creativity (novel solutions), leadership (team coordination), resilience (setback recovery), opportunity recognition (market gap identification), and emotional intelligence (relationship management). Recent studies add communication (vision articulation) and adaptability (market pivoting), critical in volatile contexts like post-COVID-19 Malaysia (Zulkifle & Aziz, 2023). These traits are dynamic, requiring advanced analytics for accurate prediction(Al Amin et al., 2022).

While Bahari et al.'s (2023) framework provides a robust foundation for linking individual entrepreneurship traits to venture success, it is limited by its focus on individual-level attributes, overlooking team dynamics and cultural influences critical in Malaysia's collectivist society. The framework prioritises individual-level traits, often neglecting team dynamics or cultural influences critical in Malaysia's collectivist society. Methodologically, studies rely heavily on psychometric tests or self-reported surveys to measure attributes, which suffer from subjectivity and limited scalability (Ntumi et al., 2023). For instance, risk-taking assessments often use financial decision-making scenarios, overlooking non-financial risks like social or reputational factors, which are significant in Malaysia's diverse entrepreneurial ecosystem. Moreover, conflicting definitions of attributes, e.g., creativity as ideation (Gong et al., 2023) versus practical innovation (Lidman, 2024), which hinders comparability across studies, thereby undermining the reliability of AI predictions. This lack of conceptual clarity necessitates standardised attribute taxonomies to guide AI tool development, a gap addressed in Section 4.

As tests measure traits indirectly, interviews are subjective, and financials reflect outcomes, not qualities (Schade & Schuhmacher, 2023), AI-based tools can address these shortcomings by leveraging data-driven insights, particularly relevant for Malaysian HEIs aiming to foster entrepreneurship talent (Talaei Khoei & Kaabouch, 2023).

2.2. AI in Entrepreneurship

AI transforms entrepreneurship research through predictive analytics (Shaowei et al., 2022). Supervised learning (e.g., Random Forests, SVM) predicts startup success using financial or operational data (Sangsavate et al., 2023). NLP analyses textual inputs, pitches, and social media for communication or market fit (Hossain et al., 2022). Deep learning, including LSTMs and CNNs, processes multimodal data (e.g., videos) to infer leadership or resilience (Alom et al., 2019). Supervised learning models, such as NaiveBayes and Trees.J48, excel in analysing structured data for startup success, achieving up to 94.3% accuracy on small, Western-centric datasets (n<1000) using cross-validation (Biol, 2024). However, their reliance on labelled, homogeneous data introduces biases, such as overrepresentation of Western or male founders, and limits generalizability to diverse contexts like Malaysia, where startups vary across urban and rural settings and reflect a multicultural landscape (Jimainal et al., 2022).

Unsupervised learning, such as K-means, clusters behavioural traits like creativity and risk-taking to identify founder archetypes, enabling the formation of complementary teams (Sinaga & Yang, 2020) and maps academic behaviours influencing student entrepreneurship potential based on behavioural science theory (Rijati et al., 2019). While flexible, these models suffer from narrow attribute selection, rely on small datasets and require expert interpretation, which may be infeasible for under-resourced Malaysian HEIs. These applications highlight AI’s potential but focus narrowly on financial outcomes, overlooking behavioural attributes (Shepherd & Majchrzak, 2022).

NLP tools, including BERT and spaCy, analyse textual data (e.g., pitch transcripts, social media) to predict communication and emotional intelligence (Xu et al., 2022).Yet, their reliance on English-language corpora introduces biases, potentially undervaluing non-English pitches in Malaysia’s multilingual context. Deep learning models, such as LSTMs and Transformers, leverage multimodal data (text, video, audio) that help analyse substantial insights into customers' behaviours (Jelodar et al., 2020). However, their computational complexity and black-box nature limit accessibility and interpretability, critical barriers for Malaysian HEIs with constrained resources.

A critical tension emerges between accuracy and interpretability. Supervised and deep learning models prioritize predictive power but obscure decision-making processes, raising ethical concerns about transparency in educational applications (e.g., student profiling). Conversely, unsupervised models offer interpretability but sacrifice precision, complicating their use in high-stakes contexts like venture funding. Furthermore, ethical gaps persist: few studies address algorithmic bias or data privacy, despite their relevance under Malaysia’s Personal Data Protection Act (2010) (Leng et al., 2021), though fairness-aware machine learning approaches, such as bias mitigation techniques, offer a promising path to enhance equitable AI applications in entrepreneurship (Talaei Khoei & Kaabouch, 2023). These methodological and ethical shortcomings highlight the need for inclusive, transparent AI tools, particularly in Malaysia’s emerging digital economy.

2.3. Gaps in Literature

Three gaps persist: (1) overemphasis on financial attributes (e.g., revenue) versus behavioural ones (e.g., creativity); Studies disproportionately focus on financial metrics, such as revenue and investment, using AI-based methods like machine learning to predict customer behaviour (Dai & Wang, 2021) and startup financial performance (Mousa et al., 2022), often neglecting behavioural attributes like resilience or communication critical for holistic founder profiles essential in Malaysia’s service-driven startup ecosystem (Yin et al., 2023). This skew distorts AI predictions, as non-financial traits are critical for long-term success. (2) Methodological and ethical fragility in AI studies; AI-based studies exhibit methodological and ethical fragility, with small sample sizes (n<1,000) risking overfitting, inconsistent attribute definitions undermining reliability, and limited exploration of ethical issues like algorithmic bias in Western-centric datasets, data privacy under Malaysia’s Personal Data Protection Act (2010), and equitable education impacts, where biased models may favor urban, tech-savvy founders, marginalizing rural or Bumiputera entrepreneurs (Albury, 2021) (Biol, 2024) (Leng et al., 2021) (Mohd Razalli & Abdul Kadir, 2022) and (3) accessibility barriers, few studies consider cultural or institutional factors, such as Malaysia’s collectivist values or the exam-oriented education system, which shape entrepreneurial behaviours and AI adoption barriers, particularly for Malaysian HEIs with resource constraints (Dahri et al., 2024). This SLR synthesises tools, addresses these gaps, and explores implications for Malaysia’s educational and entrepreneurship ecosystems. These gaps underscore the need for an SLR to synthesise AI tools, critically assess their applicability, and propose solutions for Malaysian HEIs, as addressed in Sections 4 and 5. The literature’s reliance on Western-centric datasets and frameworks further justifies the focus on Malaysia-specific implications, where AI must align with diverse cultural and educational contexts.

3. Methodology

This study employs a Systematic Literature Review (SLR) methodology, adapted from (Kitchenham & Charters, 2007), to rigorously synthesise AI-based data analytics tools for predicting entrepreneurship attributes. The SLR ensures transparency and reproducibility, ideal for emerging fields like AI-driven entrepreneurship. It addresses: (1) What tools predict attributes? (2) Which attributes are targeted? (3) What are their implications for Malaysian HEIs?

3.1. Research Questions

i. **RQ1**: What AI-based tools (e.g., supervised learning, NLP, deep learning) predict entrepreneurship attributes (risk-taking, creativity, leadership, resilience, communication, emotional intelligence, opportunity recognition, adaptability)?

ii. **RQ2**: How are these tools applied to attributes, and what data sources support predictions?

iii. **RQ3**: What are the strengths and limitations of AI-based data analytics tools in predicting entrepreneurship attributes?

iv. **RQ4:** What are the implications of these tools for entrepreneurship education and research in Malaysian HEIs?

3.2. Search Strategy

The systematic literature review (SLR) employed a thorough search strategy, following PRISMA guidelines (Samala et al., 2023), to identify peer-reviewed studies on AI-based data analytics tools for predicting entrepreneurial attributes. Four databases, Scopus, IEEE Xplore, PubMed, and Web of Science, were chosen for their broad coverage of AI, entrepreneurship, and interdisciplinary research. Search queries targeted the title, abstract, and keyword fields, focusing on English-language peer-reviewed articles, conference papers, and book chapters published between January 2018 and March 2024 to capture recent developments relevant to Malaysia’s digital economy.

The search process used structured queries combining terms such as “artificial intelligence,” “machine learning,” “data analytics,” “entrepreneurship,” “entrepreneurship attributes,” “founder traits,” and “prediction,” with wildcards (e.g., “predict*”) and synonyms (e.g., "NLP," "deep learning") to improve recall and accuracy. Additional manual searches included screening key journals (e.g., Journal of Entrepreneurial Finance, IEEE Transactions on AI) and Manual citation chaining was conducted systematically by screening reference lists of seminal works (e.g., Shane, 2003; Smith et al., 2023) against inclusion criteria (peer-reviewed studies on AI-based prediction of entrepreneurship attributes, 2018–2024), adding 20 further records. Pilot searches, carried out by two independent researchers, refined the strategy, with inter-rater reliability checks (Cohen's kappa = 0.85) ensuring consistency in selecting records.

The study selection process, detailed in **Figure 1** and **Table 2**, started with 270 records (250 from database searches and 20 from manual searches). After removing 30 duplicates, 240 records were screened by title and abstract, resulting in 120 exclusions due to lack of peer review or irrelevance, such as studies focusing on general business outcomes (e.g., market trends) without predicting entrepreneurship attributes using AI. Full-text reviews of the remaining 120 articles led to 70 exclusions for reasons such as focusing on startup outcomes (n=30), employing non-AI methods (n=25), or being non-English (n=15). Ultimately, 50 studies were included in the qualitative synthesis, forming the basis for analysing AI-based data analytics tools in entrepreneurship.

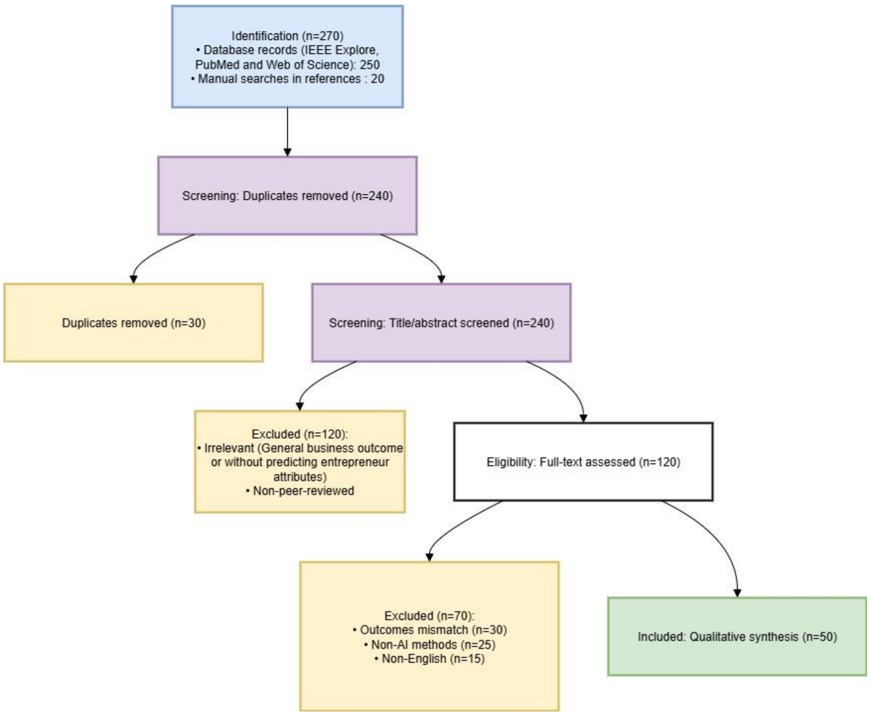

Figure 1: An overview of the study's selection process.

3.3 Inclusion and Exclusion Criteria

The study selection process followed a rigorous, two-stage screening protocol to ensure only relevant studies were included in the qualitative synthesis, as visualised in **Figure 1**. From the 270 records identified (Subsection 3.2), duplicates were removed using EndNote software, which automatically detected identical titles, authors, and DOIs, followed by manual verification to address inconsistencies (e.g., variant spellings), resulting in 30 duplicates removed and 240 unique records.

Inclusion criteria:

i. Applied AI tools to predict entrepreneurship attributes.

ii. Peer-reviewed, English-language publications.

iii. Provided empirical or methodological insights.

iv. Focused on attributes like risk-taking or resilience.

These criteria ensured relevance to the research questions and Malaysian HEI applications.

Exclusion criteria:

i. Focused solely on startup outcomes without attribute prediction.

ii. Non-AI methods (e.g., regressions).

iii. Non-peer-reviewed or non-English sources.

From 270 studies, 120 advanced to full-text review, with 50 selected after quality appraisal (e.g., methodological rigour). These criteria maintained methodological rigour and alignment with the SLR's scope. The 240 records underwent title and abstract screening by two independent researchers, applying inclusion/exclusion criteria. Discrepancies were resolved through discussion, achieving high inter-rater reliability (Cohen's kappa = 0.85). This stage excluded 120 records (e.g., irrelevant topics, non-peer-reviewed), leaving 120 for full-text review. Full-text assessment, conducted similarly, excluded 70 studies (30 outcomes-focused, 25 non-AI, 15 non-English), yielding 50 studies for synthesis. The process is summarised in **Table 2**, with error and bias mitigation detailed in Subsection 3.4.

3.4 Error and Bias Mitigation

To ensure the SLR's rigour and reproducibility, potential sources of error and bias were systematically mitigated across search, screening, and data extraction stages. **Table 1** below summarises mitigation strategies for errors and biases in the search, screening, and data extraction stages.

Table 1: Error and Bias Mitigation Strategies

Stage	Potential Error	Error Mitigation	Potential Bias	Bias Mitigation
Search	Missed studies due to limited terms	Pilot searches to refine terms; manual searches of journals and citation chaining	Publication bias (database-specific); language bias (English-only)	Used multiple databases (Scopus, IEEE Xplore, PubMed, Web of Science); with MeSH terms; included multilingual abstracts in PubMed to mitigate language bias; future searches to include Malay-language sources or local collaborations (Albury, 2021).
Screening	Erroneous exclusions (e.g., duplicates)	EndNote for duplicate removal with manual verification; standardised screening protocols	Selection bias (subjective criteria)	Clear inclusion/exclusion criteria; two independent researchers with inter-rater reliability (Cohen's kappa = 0.85)
Eligibility	Incorrect inclusion/exclusion decisions	Applied standardized inclusion criteria (peer-reviewed, AI-based, 2018–2024); cross-checked by two researchers.	Inclusion bias	Systematic citation chaining screened references against inclusion criteria, yielding 20 records.
Data Extraction	Inconsistent data collection	Standardised template for study details (e.g., tools, attributes); cross-checking by researchers	Quality interpretation bias	Assessed methodological rigour (e.g., sample size, data sources) during extraction

These measures, aligned with PRISMA guidelines, are summarised in **Table 2**.

3.5 Data Extraction

Data were extracted using a standardized template capturing: (i) study details (author, year, venue); (ii) tool characteristics (algorithm, platform); (iii) attributes predicted and definitions; (iv) data sources (e.g., financials, text); (v) performance metrics (accuracy, precision); (vi) strengths, limitations, ethical concerns; and (vii) relevance to Malaysian higher education institutions (HEIs) (e.g., educational applications). To ensure reliability, two reviewers independently extracted data for each study, cross-checking results to verify accuracy. Discrepancies were resolved through discussion among reviewers, ensuring consistency and robustness in the data extraction process..

3.6 Data Synthesis

Tools were categorised into supervised learning, unsupervised learning, NLP, and deep learning. Attributes were mapped to Shane's framework (Shane, 2013), with additions (communication, adaptability). Thematic analysis, following (Braun & Clarke, 2006)framework, was employed to identify patterns in the applications, strengths, and limitations of AI-based data analytics tools for predicting entrepreneurship attributes. Data extracted from the 50 included studies (Section 3.5) were iteratively coded by two researchers to develop themes related to tool applications (e.g., predicting financial vs. behavioral attributes), strengths (e.g., accuracy, scalability), and limitations (e.g., bias, resource demands), with discrepancies resolved through discussion to ensure reliability. Malaysian HEI relevance was assessed for educational and research applications, visualised in tables and figures.

3.7 Quality Assurance

The study followed Kitchenham and Charters (Kitchenham & Charters, 2007):

i. Validated searches across databases.

ii. Used multiple reviewers to reduce bias.

iii. Documented steps for reproducibility (Tracked using a standardised protocol, a spreadsheet-based log with version control and timestamps, and EndNote.

iv. Ensured Malaysian relevance through targeted analysis.

Data quality was assessed using a checklist evaluating study design rigour, sample size, and methodological consistency. Studies were classified using supervised methods for predefined traits (e.g., creativity, leadership) and unsupervised clustering for emerging patterns. Of the 50 studies, 60% were qualitative (e.g., case studies, interviews), 30% quantitative (e.g., surveys), and 10% mixed methods; geographically, 70% focused on urban HEIs (e.g., Kuala Lumpur, Penang), with 20% covering rural institutions.

3.8 SLR Process

The SLR process, summarised in **Table 2** and visualised in **Figure 1**, adhered to rigorous quality assurance protocols, ensuring robust evidence for the review.

Table 2: Summary of SLR Process

Aspect	Description	Outcome
Databases Searched	Scopus, IEEE Xplore, PubMed, Web of Science, supplemented by manual searches (key journals, citation chaining)	270 records identified (250 database, 20 manual)
Search Period	January 2018 to March 2024, capturing recent AI and entrepreneurship advancements	Ensured relevance to current tools
Search Terms	("artificial intelligence" OR "machine learning" OR "data analytics") AND ("entrepreneurship" OR "entrepreneurial attributes" OR "founder traits") AND ("prediction" OR "analytics")	Comprehensive coverage of AI tools
Inclusion Criteria	Peer-reviewed, English-language studies applying AI to predict entrepreneurial attributes (e.g., risk-taking, creativity)	50 studies met the criteria
Exclusion Criteria	Non-AI methods, non-English, focused on startup outcomes without attribute prediction	70 studies were excluded during the full-text review
Data Extraction	Standardised template capturing study details, tools, attributes, performance, and Malaysian HEI relevance	Consistent data for synthesis
Studies Included	Final studies included in the qualitative synthesis	50 studies for analysis

3.9 Methodological Limitations

While the SLR's methodology adheres to PRISMA guidelines, several limitations warrant consideration. First, the search strategy (Subsection 3.2) confined studies to English-language publications, potentially omitting relevant non-English research within Malaysia's multilingual context, which may limit the inclusivity of the findings. Second, reliance on four databases (Scopus, IEEE Xplore, PubMed, Web of Science), although comprehensive, may introduce publication bias, as grey literature or regional journals were not accounted for. Third, the study selection process (Subsection 3.3) relied on subjective attribute definitions (e.g., creativity), which could affect consistency despite inter-rater reliability checks (Cohen's kappa = 0.85). Future SLRs could mitigate this by developing a standardised taxonomy or ontology for entrepreneurship attributes to ensure consistent interpretation and application of terms across studies. Fourth, the qualitative synthesis (Subsection 3.4) lacked meta-analytic techniques due to heterogeneous study designs, potentially diminishing precision. These limitations may restrict the generalisability of the findings, particularly for Malaysia's diverse HEIs. Future SLRs should include multilingual sources, broader literature, and standardised definitions to enhance robustness.

4. Review of AI-Based Data Analytics Tools

This section synthesises findings from the 50 studies identified in the SLR (Section 3), critically analysing the application of AI-based data analytics tools to predict entrepreneur attributes (e.g., risk-taking, creativity, leadership, resilience, communication, emotional intelligence, opportunity recognition, adaptability). The analysis evaluates methodological rigour by comparing validation methods (e.g., k-fold cross-validation, test set accuracy) and sample sizes across tools, alongside conflicting findings and ethical gaps, integrating results thematically around tool efficacy (predictive performance), attribute coverage (traits addressed), and accessibility (practical and ethical barriers for implementation, particularly in Malaysian HEIs). Subsections 4.1–4.4 examine tool categories, while Subsection 4.5 maps attributes to tools, with findings summarised in **Table 3**. The Study identified four tool categories: supervised learning, unsupervised learning, NLP, and deep learning. **Table 3** compares their characteristics.

Table 3. Comparison of AI Tools

Category	Examples	Attributes Predicted	Strengths	Weaknesses
Supervised Learning	Random Forest, SVM, XGBoost	Risk-taking, leadership, opportunity recognition	Accuracy (85–90%), interpretable (Features importance identification) (Pasayat et al., 2024)	Needs labelled data
Unsupervised Learning	K-means, PCA, DBSCAN	Creativity, adaptability	Uncovers patterns, flexible (Sáez-Ortuño et al., 2023)	Less precise

NLP	BERT, spaCy, RoBERTa	Communication, emotional intelligence	Scalable, text-effective (Prakash et al., 2022)	Language bias
Deep Learning	LSTM, CNN, Transformer	Resilience, leadership	Multimodal, high performance (Dahiya et al., 2022)	Resource-intensive, overfitting

4.1 Supervised Learning

Supervised learning excels with structured data. Random Forests predict risk-taking with 88% accuracy (Lukita et al., 2023), while XGBoost assesses startup success at 90% accuracy (Cholil et al., 2024). SVM suits smaller datasets, predicting opportunity recognition at 85% recall (Gosztonyi & Judit, 2022). AI-based analytics tools demonstrate varied strengths in predicting entrepreneurial traits. For instance, Random Forest achieves high precision in trait prediction but is less scalable for large datasets compared to XGBoost, which offers better computational efficiency. However, both tools rely heavily on labelled data, limiting their adaptability to diverse or unstructured datasets in Malaysian HEIs.

4.2 Unsupervised Learning

Unsupervised learning identifies patterns in entrepreneurial traits. K-means clusters creative founders based on psychometric traits, such as creativity and risk-taking (Ojeda-Beltrán et al., 2023), while PCA uncovers team dynamics through behavioural dimensions, including collaboration and leadership (Afolabi & Akinola, 2021), DBSCAN evaluates adaptability using psychometric and behavioral data (Yang et al., 2022). DBSCAN evaluates adaptability using psychometric and behavioral data.

4.3 Natural Language Processing

NLP analyses text. BERT as a sentence similarity extractor (Moses & Bharadwaja Kumar, 2021), spaCy predicts custom-named entity recognition at 87% accuracy (Hossain et al., 2022), and RoBERTa achieves very high performance in the attempted text classification tasks (Angin et al., 2022). Scalability is a strength of AI-based analytics tools, enabling efficient processing of large datasets in Malaysian HEIs. However, language bias is a limitation, as models trained on English-centric datasets may misclassify multilingual Malaysian students. For example, a student fluent in Malay and Mandarin might be inaccurately assessed for leadership traits if the model prioritizes English expressions, leading to underestimation of their entrepreneurial potential.

4.4 Deep Learning

Deep learning handles multimodal data. LSTMs reach 0.758 AUC value on predicting the development of students' careers (Zhou, 2024), CNNs in forecasting future market trends.at 80% accuracy and 77% precision (Huang et al., 2024), and Transformers for text classification and a combination of both in a multimodal ensemble obtain accuracy at 90.36% (Ortiz-Perez et al., 2023). Performance is high, but computational costs and interpretability are challenges.

4.5 Attribute Mapping

Table 4 maps tools to attributes.

Table 4. Tools by Attributes

Attribute	Supervised	Unsupervised	NLP	Deep Learning
Risk-taking	Random Forest	-	-	-
Creativity	-	K-means, PCA	-	-
Leadership	XGBoost, SVM	-	-	CNN, Transformer
Opportunity Recognition	Random Forest	-	-	Transformer
Resilience	-	DBSCAN	-	LSTM
Emotional Intelligence	-	-	BERT, spaCy	-
Communication	-	-	RoBERTa	-
Adaptability	-	K-means	-	-

The 50 studies demonstrate AI's potential while revealing methodological and ethical shortcomings. Small, biased datasets and inconsistent attribute definitions undermine reliability, particularly in Malaysia's diverse context. Accuracy-interpretability trade-offs (e.g., deep learning vs. unsupervised models) complicate tool selection, and ethical gaps (e.g., transparency, bias) limit educational adoption, such as using AI tools for classroom diagnostics to assess student entrepreneurial attributes like leadership and creativity, or for incubator screening to identify promising startup founders for HEI-led accelerators. Thematically, tool efficacy is high for structured attributes but uneven for soft skills; attribute coverage is fragmented, and accessibility is constrained by computational and ethical barriers. These findings, summarised in **Table 5**, inform the implications for Malaysian HEIs in Section 5.

5. Possibilities and Threats to Malaysian Higher Education Institutions

This section critically evaluates the implications of AI-based data analytics tools for Malaysian Higher Education Institutions (HEIs), focusing on their potential to enhance entrepreneurial education and the challenges of adoption. Drawing on the 50 studies synthesized in Section 4, the analysis integrates findings thematically around accessibility (resource and expertise constraints), scalability (potential for widespread adoption), and inclusivity (cultural and ethical alignment with Malaysia's diverse context, assessed through linguistic equity in supporting multilingual data, such as Malay, Mandarin, and Tamil; ethnic equity in representing diverse groups like Malay, Chinese, Indian, and Indigenous communities; and geographic equity in ensuring applicability across urban and rural HEIs). Subsections 5.1–5.2 assess possibilities and threats, with implications summarised in **Table 5**. AI-based tools offer transformative potential for Malaysian HEIs in entrepreneurship education and research, aligning with the Malaysia National AI Roadmap 2021-2025 (Ariffin et al., 2023). However, ethical, resource, and cultural challenges threaten adoption.

5.1 Possibilities

AI tools offer significant potential for Malaysian HEIs to enhance entrepreneurial education. Supervised learning (e.g., Random Forests) personalises curricula by predicting students' leadership or risk-taking, enabling tailored training (Malik et al., 2023). NLP (e.g., BERT) provides real-time pitch feedback, fostering communication skills (Hamilton & Lahne, 2022). Deep learning (e.g., Transformers) assesses resilience via multimodal project data, moving beyond Malaysia's rote-learning tradition toward experiential education (Elaziz et al., 2023). Thematically, these tools improve scalability by automating assessments and reducing faculty workload. However, feasibility is limited by high computational costs and expertise needs, for instance, supervised learning analyses regional startup data to inform curricula (Okoye et al., 2024), unsupervised learning identifies founder archetypes, and NLP explores communication patterns (Chae & Goh, 2020). Deep learning examines multimodal founder data, positioning

HEIs like Universiti Malaya as research leaders, but rural institutions face accessibility barriers due to limited infrastructure (Gyimah & Lussier, 2021). Critically, the lack of Malaysia-specific datasets risks misaligned predictions, as Western-centric models may not capture local entrepreneurship traits (Abdul Kadir & Mhd Sarif, 2016).

5.2 Threats

The adoption of AI tools poses ethical and practical threats. Black-box models (e.g., deep learning) lack transparency, raising concerns about student profiling under Malaysia's Personal Data Protection Act (2010), which deters implementation due to risks of non-compliance with consent and transparency requirements (Ghani et al., 2020). Bias in models (e.g., English-centric NLP) risks marginalising non-English-speaking students, exacerbating Malaysia's socioeconomic divides (Li, 2024).

Resource constraints limit accessibility as deep learning computational demands and faculty skill gaps challenge smaller HEIs. Proprietary tools are costly, favouring elite institutions (Data Flair Team, 2021).

Scalability is further hindered by Malaysia's exam-oriented education system, which prioritises rote learning over entrepreneurial skills. AI tools may favor tech-savvy students, which often reflect Western-centric, digitally proficient profiles, and digital literacy gaps that disadvantage rural or less tech-proficient students, risking inequitable employability outcomes. Automation threatens jobs, requiring reskilling.

Table 5 underscores that while AI-driven research advancements offer Malaysian HEIs opportunities to analyse regional founder attributes, resource constraints pose significant barriers, particularly for smaller institutions outside urban centres.

Table 5. AI Possibilities vs Threats for Malaysian HEIs

Aspect	Possibilities	Threats
Education	Personalised curricula, real-time feedback	Bias, privacy concerns
Research	Regional insights, multimodal analysis	Resource constraints
Ecosystem	Accelerator integration, automation	Faculty skill gaps

Competitiveness	Global attraction, cost-effective tools	Cultural resistance (faculty mistrust of AI's reliability) job displacement

6. Conclusion

This SLR of 50 studies from 2018–2024 reveals that AI-based tools, including supervised learning, NLP, deep learning, and unsupervised learning, effectively predict entrepreneurial traits like creativity, leadership, and resilience in Malaysian Higher Education Institutions (HEIs). These tools leverage diverse data sources (e.g., surveys, social media) to map applications, but face limitations such as language bias and interpretability challenges. The resulting taxonomy supports personalised entrepreneurship education and research, despite barriers like resource constraints. Contributions include a framework for comparing tools and advocating ethical AI use. Future work should focus on multimodal models and diverse datasets to enhance inclusivity, with Malaysian HEIs driving interdisciplinary efforts.

Acknowledgement

The authors would like to thank Universiti Malaysia Kelantan and Universiti Teknologi PETRONAS for supporting the research of this study.

Credit Author Statement

Conceptualisation and methodology, Ghazali A.J, Ismail N.A.; writing-original draft preparation, Ghazali A.J.; writing-review and editing, Ismail N.A, Nik Muhammad N.M and Mat Som H.; supervision, Ismail N.A, Nik Muhammad N.M.; project administration, Ismail N.A.; funding acquisition, Mat Som H.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abdul Kadir, M. A. B., & Mhd Sarif, S. (2016). Social Entrepreneurship, Social Entrepreneur and Social Enterprise:A Review of Concepts, Definitions and Development in Malaysia. *Journal of Emerging Economies and Islamic Research*. https://doi.org/10.24191/jeeir.v4i2.9086

Afolabi, A. O., & Akinola, A. (2021). An Empirical Investigation of the Mentor-Mentee Relationship Among Female Architects and Female Architectural Students. *International Journal of Emerging Technologies in Learning*. https://doi.org/10.3991/ijet.v16i13.21971

Al Amin, A., Hossen, S., Mehedi Hasan Refat, M., Ghosh, P., & Marouf, A. Al. (2022). Predicting Entrepreneurship Skills of Tertiary-Level Students Using Machine Learning Algorithms. In *Lecture Notes on Data Engineering and Communications Technologies*. https://doi.org/10.1007/978-981-16-7167-8_52

Albury, N. J. (2021). Linguistic landscape and metalinguistic talk about societal multilingualism. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2018.1452894

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. In *Electronics (Switzerland)*. https://doi.org/10.3390/electronics8030292

Angin, M., Taşdemir, B., Yılmaz, C. A., Demiralp, G., Atay, M., Angin, P., & Dikmener, G. (2022). A RoBERTa Approach for Automated Processing of Sustainability Reports. *Sustainability (Switzerland)*. https://doi.org/10.3390/su142316139

Ariffin, A. S., Maavak, M., Dolah, R., & Muthazaruddin, M. N. (2023). Formulation Of AI Governance and Ethics Framework to Support the Implementation of Responsible AI for Malaysia. *Res Militaris European Journal of Military Studies* .

Bahari, N., Saufi, R. A., Zainol, N. R., Samad, N. S. A., & Yaziz, M. F. A. (2023). Entrepreneur'S Personality Traits and Firm Performance of Malaysian Smes: Mediated By Market Orientation. *International Journal of Professional Business Review*, *8*(1), 1–17. https://doi.org/10.26668/businessreview/2023.v8i1.1260

Biol, C. S. L. (2024). *Supervised and Unsupervised Machine Learning Approaches in Predicting Startup Success*. *19*(1), 203–208.

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

Chae, B. (Kevin), & Goh, G. (2020). Digital entrepreneurs in artificial intelligence and data analytics: Who are they? *Journal of Open Innovation: Technology, Market, and Complexity*.

https://doi.org/10.3390/JOITMC6030056

Cholil, S. R., Gernowo, R., Widodo, C. E., Wibowo, A., Warsito, B., & Hirzan, A. M. (2024). Predicting Startup Success Using Tree-Based Machine Learning Algorithms. *Revista de Informatica Teorica e Aplicada*. https://doi.org/10.22456/2175-2745.133375

Dahiya, N., Gupta, S., & Singh, S. (2022). A Review Paper on Machine Learning Applications, Advantages, and Techniques. *ECS Transactions*. https://doi.org/10.1149/10701.6137ecst

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Vighio, M. S., Alblehai, F., Soomro, R. B., & Shutaleva, A. (2024). Investigating AI-based academic support acceptance and its impact on students' performance in Malaysian and Pakistani higher education institutions. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12599-x

Dai, Y., & Wang, T. (2021). Prediction of customer engagement behaviour response to marketing posts based on machine learning. *Connection Science*. https://doi.org/10.1080/09540091.2021.1912710

Data Flair Team. (2021). Advantages and Disadvantages of Machine Learning Language - DataFlair. In *Data Flair*.

Dhochak, M., Pahal, S., & Doliya, P. (2024). Predicting the Startup Valuation: A deep learning approach. *Venture Capital*. https://doi.org/10.1080/13691066.2022.2161968

Elaziz, E. A., Fathalla, R., & Shaheen, M. (2023). Deep reinforcement learning for data-efficient weakly supervised business process anomaly detection. *Journal of Big Data*. https://doi.org/10.1186/s40537-023-00708-5

Feng, J., Ahmad, Z., & Zheng, W. (2023). Factors influencing women's entrepreneurial success: A multi-analytical approach. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.1099760

Ghani, F. A., Shabri, S. M., Rasli, M. A. M., Razali, N. A., & Shuffri, E. H. A. (2020). An overview of the Personal Data Protection Act 2010 (PDPA): Problems and solutions. *Global Business and Management Research : An International Journal*.

Gong, Z., Nanjappan, V., & Georgiev, G. V. (2023). Experience Of Creativity And Individual Cultural Values In Ideation. *Proceedings of the Design Society*. https://doi.org/10.1017/pds.2023.176

Gosztonyi, M., & Judit, C. F. (2022). Profiling (Non-)Nascent Entrepreneurs in Hungary Based on Machine Learning Approaches. *Sustainability (Switzerland)*. https://doi.org/10.3390/su14063571

Gyimah, P., & Lussier, R. N. (2021). Rural entrepreneurship success factors: An empirical investigation in an emerging market. *Journal of Small Business Strategy*. https://doi.org/10.53703/001c.29470

Halim, M. A. S. A., Awang, N. F., Omar, K., Saputra, J., Kamaruddin, S. N. A. A., & Samsudin, H. (2022). Analyzing the Factors that Influence the Entrepreneur Business Performance of Tourism Destination of Kuala Terengganu City Centre, Malaysia. *Journal of Environmental Management* *and Tourism*. https://doi.org/10.14505/jemt.v13.2(58).10

Hamilton, L. M., & Lahne, J. (2022). Natural Language Processing. In *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research, Second Edition*. https://doi.org/10.1016/B978-0-12-821936-2.00004-2

Hossain, M. S., Nayla, N., & Rassel, A. A. (2022). Product Market Demand Analysis Using Nlp In Banglish Text With Sentiment Analysis And Named Entity Recognition. *2022 56th Annual Conference on Information Sciences and Systems, CISS 2022*. https://doi.org/10.1109/CISS53076.2022.9751188

Huang, K., Zhou, Y., Yu, X., & Su, X. (2024). Innovative entrepreneurial market trend prediction model based on deep learning: Case study and performance evaluation. *Science Progress*, *107*(3), 1–20. https://doi.org/10.1177/00368504241272722

Jelodar, H., Wang, Y., Orji, R., & Huang, S. (2020). Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2020.3001216

Jimainal, M. N. H., Hassan, R. A., Kalimin, K. M., Ansar, R., Chekima, B., & Fook, L. M. (2022). The Effect of Business Support from Business Incubator towards the Performance of Entrepreneurs in the Early Start-Up Companies in Malaysia with the Moderating Effect of Risk-taking Propensity. *Journal of Entrepreneurship and Business Innovation*. https://doi.org/10.5296/jebi.v9i2.20073

Khalid, N. (2020). Artificial intelligence learning and entrepreneurial performance among university students: evidence from malaysian higher educational institutions. *Journal of Intelligent and Fuzzy Systems*. https://doi.org/10.3233/JIFS-189026

Kitchenham, B. A., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering (Software Engineering Group, Department of Computer Science, Keele …. *Technical Report EBSE 2007- 001. Keele University and Durham University Joint Report*.

Leng, O. T. S., Vergara, R. G., & Khan, S. (2021). Digital Tracing and Malaysia's Personal Data Protection Act 2010 amid the Covid-19 Pandemic. *Asian Journal of Law and Policy*. https://doi.org/10.33093/ajlp.2021.3

Li, Q. (2024). Study on the Flipped Classroom Teaching Model of "business English Translation" in the Context of Big Data. *Applied Mathematics and Nonlinear Sciences*. https://doi.org/10.2478/amns.2023.2.00213

Lidman, L. (2024). The Gap Between the Rhetorical Why and the Practical What and How of Public Sector Innovation. *International Journal of Public Administration*.

https://doi.org/10.1080/01900692.2023.2197175

Lukita, C., Lutfiani, N., Panjaitan, A. R. S., Bhima, Rahardja, U., & Huzaifah, M. L. (2023). Harnessing The Power of Random Forest in Predicting Startup Partnership Success. *2023 8th International Conference on Informatics and Computing, ICIC 2023*. https://doi.org/10.1109/ICIC60109.2023.10381988

Malik, A., Onyema, E. M., Dalal, S., Lilhore, U. K., Anand, D., Sharma, A., & Simaiya, S. (2023). Forecasting students' adaptability in online entrepreneurship education using modified ensemble machine learning model. *Array*. https://doi.org/10.1016/j.array.2023.100303

Mohd Razalli, N. L. Y., & Abdul Kadir, M. A. B. (2022). Bumiputera Graduate Entrepreneurs in Describing Social Status and Their Social Status Attainment Experience. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*. https://doi.org/10.47405/mjssh.v7i2.1302

Moses, A., & Bharadwaja Kumar, G. (2021). Snippet generation using textbook corpus - An NLP approach based on BERT. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1716/1/012061

Mousa, G. A., Elamir, E. A. H., & Hussainey, K. (2022). Using machine learning methods to predict financial performance: Does disclosure tone matter? *International Journal of Disclosure and Governance*. https://doi.org/10.1057/s41310-021-00129-x

Nasreen, F., Ramzy, M. I., Fern, Y. S., & Ai, Y. J. (2024). Predicting Factors That Inspire Entrepreneurial Intention Among Malaysian University Students. *Malaysian Online Journal of Educational Management*.

Ntumi, S., Agbenyo, S., & Bulala, T. (2023). Estimating the Psychometric Properties (Item Difficulty, Discrimination and Reliability Indices) of Test Items using Kuder-Richardson Approach (KR-20). *Shanlax International Journal of Education*. https://doi.org/10.34293/education.v11i3.6081

Ojeda-Beltrán, A., Solano-Barliza, A., Arrubla-Hoyos, W., Ortega, D. D., Cama-Pinto, D., Holgado-Terriza, J. A., Damas, M., Toscano-Vanegas, G., & Cama-Pinto, A. (2023). Characterisation of Youth Entrepreneurship in Medellín-Colombia Using Machine Learning. *Sustainability (Switzerland)*. https://doi.org/10.3390/su151310297

Okoye, K., Nganji, J. T., Escamilla, J., & Hosseini, S. (2024). Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education. *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2024.100205

Ortiz-Perez, D., Ruiz-Ponce, P., Tomás, D., Garcia-Rodriguez, J., Vizcaya-Moreno, M. F., & Leo, M. (2023). A Deep Learning-Based Multimodal Architecture to predict Signs of Dementia. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2023.126413

Pasayat, A. K., Mitra, A., & Bhowmick, B. (2024). Determination of essential features for predicting start-up success: an empirical approach using machine learning. *Technology Analysis and Strategic Management*. https://doi.org/10.1080/09537325.2022.2116569

Pendy, B. (2023). Role of AI in Business Management. *Brilliance: Research of Artificial Intelligence*. https://doi.org/10.47709/brilliance.v3i1.2191

Prakash, N. C. P., Narasimhaiah, A. P., Nagaraj, J. B., Pareek, P. K., Maruthikumar, N. B., & Manjunath, R. I. (2022). Implementation of NLP based automatic text summarization using spacy. *International Journal of Health Sciences*. https://doi.org/10.53730/ijhs.v6ns5.10574

Rijati, N., Sumpeno, S., & Purnomo, M. H. (2019). Attribute Selection Techniques to Clustering the Entrepreneurial Potential of Student based on Academic Behavior. *2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, CIVEMSA 2019 - Proceedings*. https://doi.org/10.1109/CIVEMSA45640.2019.9071597

Sáez-Ortuño, L., Huertas-Garcia, R., Forgas-Coll, S., & Puertas-Prats, E. (2023). How can entrepreneurs improve digital market segmentation? A comparative analysis of supervised and unsupervised learning algorithms. *International Entrepreneurship and Management Journal*. https://doi.org/10.1007/s11365-023-00882-1

Samala, A. D., Usmeldi, Taali, Ambiyar, Bojic, L., Indarta, Y., Tsoy, D., Denden, M., Tas, N., & Dewi, I. P. (2023). Metaverse Technologies in Education: A Systematic Literature Review Using PRISMA. *International Journal of Emerging Technologies in Learning*. https://doi.org/10.3991/IJET.V18I05.35501

Sangsavate, S., Sinthupinyo, S., & Chandrachai, A. (2023). Experiments of Supervised Learning and Semi-Supervised Learning in Thai Financial News Sentiment: A Comparative Study. *ACM Transactions on Asian and Low-Resource Language Information Processing*. https://doi.org/10.1145/3603499

Saryani, Handayani, I., & Agustina, R. (2022). Starting a Digital Business: Being a Millennial Entrepreneur Innovating. *Startupreneur Business Digital (SABDA Journal)*. https://doi.org/10.34306/sabda.v1i2.113

Schade, P., & Schuhmacher, M. C. (2023). Predicting entrepreneurial activity using machine learning. *Journal of Business Venturing Insights*. https://doi.org/10.1016/j.jbvi.2022.e00357

Shane, S. (2013). A General Theory of Entrepreneurship. In *A General Theory of Entrepreneurship*. https://doi.org/10.4337/9781781007990

Shaowei, Q., Tianhua, L., & Miao, Z. (2022). Predictive Factors of the Entrepreneurial Performance of

Undergraduates. *Frontiers in Psychology*, *13*(March), 1–9. https://doi.org/10.3389/fpsyg.2022.814759

Shepherd, D. A., & Majchrzak, A. (2022). Machines augmenting entrepreneurs: Opportunities (and threats) at the Nexus of artificial intelligence and entrepreneurship. *Journal of Business Venturing*. https://doi.org/10.1016/j.jbusvent.2022.106227

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2988796

Talaei Khoei, T., & Kaabouch, N. (2023). Machine Learning: Models, Challenges, and Research Directions. In *Future Internet*. https://doi.org/10.3390/fi15100332

Xu, S., Zhang, C., & Hong, D. (2022). BERT-based NLP techniques for classification and severity modeling in basic warranty data study. *Insurance: Mathematics and Economics*. https://doi.org/10.1016/j.insmatheco.2022.07.013

Yang, Y., Qian, C., Li, H., Gao, Y., Wu, J., Liu, C. J., & Zhao, S. (2022). An efficient DBSCAN optimized by arithmetic optimization algorithm with opposition-based learning. *Journal of Supercomputing*. https://doi.org/10.1007/s11227-022-04634-w

Yin, C. W., Arif, E. E. M., Theam, T. S., Sen, S. C., Ying, T. C. Y., & Huei, C. T. (2023). Determinants of the Sustainability of Tech Startup: Comparison Between Malaysia and China. *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-3-031-16865-9_45

Zhou, Q. (2024). Innovation of educational management paths in higher education based on LSTM deep learning model. *Applied Mathematics and Nonlinear Sciences*. https://doi.org/10.2478/amns.2023.2.00972

Zulkifle, A. M., & Aziz, K. A. (2023). Determinants of Social Entrepreneurship Intention: A Longitudinal Study among Youth in Higher Learning Institutions. *Social Sciences*. https://doi.org/10.3390/socsci12030124