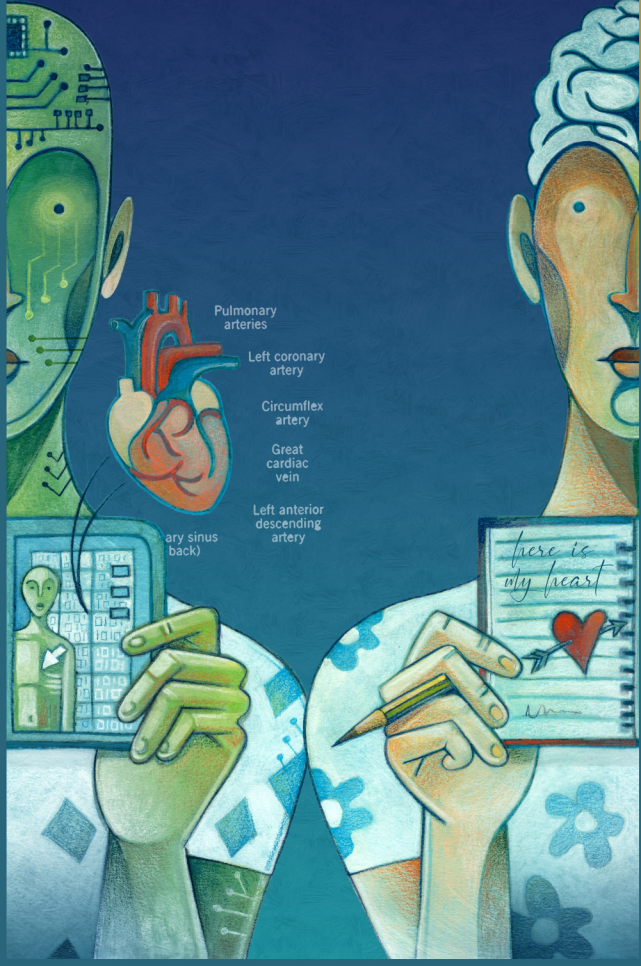
Lorella Giannandrea
Francesca Gratani
Lorenza Maria Capolla
[a cura di]



VALUTAZIONE E INTELLIGENZA ARTIFICIALE

Prospettive didattiche, etiche e metodologiche

sentieri scenari postmoderni

Pensa MULTIMEDIA

sentieri
scenari
postmoderni

Collana diretta da
Michele Corsi
Massimiliano Stramaglia

12

Lorella Giannandrea, Francesca Gratani
Lorenza Maria Capolla
[a cura di]

VALUTAZIONE E INTELLIGENZA ARTIFICIALE

PROSPETTIVE DIDATTICHE, ETICHE E METODOLOGICHE

ATTI DEL CONVEGNO PRIN "AI&F - ARTIFICIAL INTELLIGENCE & FEEDBACK FOR EFFECTIVE LEARNING".
22 E 23 GENNAIO 2026
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Pensa MULTIMEDIA

Il presente lavoro è stato pubblicato con i fondi del Progetto PRIN 2022
AI&F - Artificial intelligence & feedback for effective learning
n. 2022ZMYYTH - CUP D53D23013110006 - PNRR M4.C2.1.1

ISBN volume 979-12-5568-510-4
ISSN collana 2421-1133

2026 © by Pensa Evolution Srl
73100 Lecce • Via Arturo Maria Caprioli, 8 • Tel. 0832.230435
www.pensamultimedia.it

II.10
Conceptual Domains of AI in Education for Neuroinclusive Measurement Using Topic Modelling

Irene Gianeselli
PhD, Research Assistant (Psychometrics), Free University of Bozen-Bolzano, irene.gianeselli@unibz.it
Dario Ianes
PhD, Full Professor (Special Education), Free University of Bozen-Bolzano, dario.ianes@unibz.it
Rita Cosoli
PhD Candidate, University of Geneva, rita.cosoli@unige.ch
Marco Cadavero
PhD Candidate, University of Bologna, marco.cadavero@unibo.it
Gaetano Scianatico
PhD Candidate, University of Bari, gaetano.scianatico@uniba.it
Alessandro Oronzo Caffò
PhD, Associate Professor (Psychometrics), University of Bari, alessandro.caffo@uniba.it,
Andrea Bosco
PhD, Full Professor (Psychometrics), University of Bari, andrea.bosco@uniba.it[1]

Abstract

The increasing integration of artificial intelligence (AI) in educational contexts calls for measurement instruments that adequately capture learners' experiences from a neuroinclusive perspective. This study examines the conceptual structure of recent AI-in-education research to inform early phases of scale development. A systematic Scopus search identified 411 peer-reviewed journal articles published between 2020 and 2025, which were analysed using Latent Dirichlet Allocation topic modelling. Four latent domains were identified, reflecting general educational discourse on AI, analytical research perspectives, measurement and assessment, and medical or clinical training contexts. Lexical frequency analyses further indicated a limited representation of emotional and epistemic constructs.

1 Irene Gianeselli contributed to conceptualization, methodology, formal analysis, and writing – original draft. Dario Ianes contributed to conceptualization. Rita Cosoli, Marco Cadavero, and Gaetano Scianatico contributed to conceptualization and writing – review & editing. Alessandro Oronzo Caffò contributed to methodology and writing – review & editing. Andrea Bosco contributed to methodology, writing – review & editing, and supervision.

Keywords: Artificial intelligence; neurodiversity; topic modelling; scale development; educational psychology.

L'applicazione sempre più diffusa dell'intelligenza artificiale (IA) nei contesti educativi rende necessario disporre di strumenti di misurazione capaci di cogliere in modo preciso l'interazione degli studenti con tali strumenti in una prospettiva neuroinclusiva. Questo studio analizza la struttura concettuale della recente letteratura sull'IA in ambito educativo, con l'obiettivo di orientare le fasi iniziali dello sviluppo di una scala psicometrica. Attraverso una ricerca sistematica su Scopus sono stati individuati 411 articoli scientifici peer-reviewed pubblicati tra il 2020 e il 2025, analizzati mediante tecniche di topic modelling basate su Latent Dirichlet Allocation. Sono emersi quattro domini latenti, riferibili al discorso educativo generale sull'IA, alle prospettive analitiche della ricerca, alla misurazione e valutazione e ai contesti di formazione medica o clinica. L'analisi delle frequenze lessicali ha mostrato una presenza limitata di costrutti emotivi ed epistemici.

Parole chiave: intelligenza artificiale; neurodiversità; topic modelling; sviluppo di scala; psicologia dell'educazione.

1. Introduction

The expanding presence of Artificial Intelligence (AI) in education is reshaping how learners access information, receive feedback, and regulate cognitive processes. Advances such as generative AI and intelligent tutoring systems have shifted AI from a peripheral technological aid to an interactive cognitive agent influencing learning behaviours and decision-making (Bosco, 2024). As AI becomes increasingly embedded in instructional environments, understanding learners' cognitive, emotional, and behavioural experiences has become essential. Existing measurement instruments capture only part of this complexity. Validated scales, including the AI Attitude Scale (AIAS-4; Grassini, 2023), the General Attitudes toward Artificial Intelligence Scale (GAAIS; Schepman & Rodway, 2023), and the Threats of Artificial Intelligence Scale (TAI; Kieslich et al., 2021), primarily assess attitudes or perceived risks. However, they were not developed for educational contexts and do not capture the interplay among cognitive, emotional, and behavioural processes involved in learning with AI. Although research shows that AI can influence metacognition and socioemotional engagement (Pacheco et al., 2025), emotional responses, as well as epistemic and affective–motivational states such as curiosity (Noordewier &

Gocłowska, 2024) and trust (McAllister, 1995), remain insufficiently measured. These gaps are particularly salient in neuroinclusive education (Dark, 2024). Learners who present distinct neurological profiles, including those associated with Autism Spectrum Disorders (ASD), Attention Deficit/Hyperactivity Disorder (ADHD), and dyslexia, as well as cognitive variations such as high intellectual ability, may engage with AI differently due to variation in cognitive functioning (e.g., executive processing, attentional styles, sensory sensitivities, and metacognitive ability). A neurodiversity perspective suggests that AI tools may reduce cognitive load for some profiles while increasing uncertainty or emotional discomfort for others (Hutson, 2024). Existing instruments do not account for neurological variability, which limits their suitability for neuro-inclusive educational research. A further limitation concerns emotional processes. Although AI has been associated with curiosity, motivation, exploratory behaviour, anxiety, and error-related concerns (Montag & Elhai, 2025; Naiseh et al., 2025), these experiences are rarely operationalised in existing measures, constraining investigations of how affect interacts with learning outcomes and self-regulation. This study addresses these gaps by identifying the conceptual domains needed to assess the multidimensional impact of AI in neuroinclusive educational contexts. Through a systematic literature search and computational extraction of semantic patterns, it provides an evidence-informed foundation for developing a new measurement instrument and supports subsequent phases of scale development.

2. Methods

2.1 Literature Search and Corpus Preparation

The Scopus search strategy was operationalised to retrieve peer-reviewed publications simultaneously addressing (a) artificial intelligence, (b) educational contexts, and (c) measurement or psychometric concepts. The search was conducted on titles, abstracts, and keywords (TITLE-ABS-KEY), combining AI-related terms (e.g., *artificial intelligence*, *AI*), educational terms (e.g., *learning*, *student*, *teacher*), and measurement-related terms (e.g., *scale*, *instrument*, *validation*, *factor analysis*). Specifically, records were included if their combined title, abstract, and keyword fields simultaneously contained terms related to (a) psychometric measurement, (b) artificial intelligence, and (c) educational contexts, as operationalised in the Scopus query reported in Appendix A. Automatic filtering was applied sequentially by publication year, document type, language,

and conceptual relevance inferred from text content. No manual screening was performed, ensuring full reproducibility of the document retrieval process. The search was restricted to journal articles published between 2020 and 2025 and written in English. This time frame was selected to capture the period in which AI systems (particularly generative and adaptive AI) became widely accessible in educational settings, leading to a substantive shift in both pedagogical practices and research on learner-AI interaction. Scopus was selected as the sole data source because its journal coverage is substantially broader than that of Web of Science, particularly in educational and technology-enhanced learning research (Mongeon & Paul-Hus, 2016), making it well suited to capturing recent developments in AI-mediated education. Records focusing exclusively on technical AI engineering without educational or psychological relevance were excluded. Titles, abstracts, author keywords, and index keywords of all eligible publications were extracted and concatenated into a unified text corpus. Conceptual inclusion criteria were implemented exclusively through automated text-based filtering rather than manual screening. After filtering, the final corpus consisted of 411 publications, which were retained for text preprocessing and topic modelling.

2.2 Text Preprocessing and Topic Modelling

All text-mining procedures were implemented in R (version 4.5.1) using the *tidyverse*, *tidytext*, *slam*, and *topicmodels* packages. Text preprocessing followed standard natural language processing procedures, including lowercasing, removal of punctuation and stopwords (using the standard *tidytext* stop-word list), tokenisation into unigrams, removal of non-alphabetic tokens, and exclusion of tokens shorter than four characters. Tokens were aggregated into a document–term matrix (DTM) using raw term counts as cell values. The resulting DTM exhibited high sparsity (98.45%), a pattern typical of scientific text corpora and consistent with the assumptions of probabilistic topic modelling. Latent Dirichlet Allocation (LDA) was applied to identify latent semantic structures within the corpus. LDA is a hierarchical Bayesian model in which each document is represented as a mixture of latent topics and each topic as a probability distribution over words (Blei, Ng, & Jordan, 2003). Prior research has shown that LDA can recover coherent conceptual structures in scientific corpora and educational research domains (Griffiths & Steyvers, 2004; Chen et al., 2021). Although structural topic models allow the inclusion of document-level metadata (Roberts, Stewart & Tingley, 2019), classical LDA was considered appropriate

given the exploratory aim of identifying conceptual domains and the absence of theoretically relevant covariates. The number of topics was determined through combined inspection of perplexity trends and qualitative semantic coherence. Alternative solutions were examined. The k = 3 solution collapsed educational, analytical, and measurement-related content into a single broad topic, whereas the k = 5 solution produced semantically overlapping domains with reduced interpretability. In contrast, the four-topic model provided clearer conceptual separation while maintaining parsimony.

2.3 Construct Identification and Model Adequacy Checks

Construct identification involved interpretive examination of topic–word distributions. High-probability terms associated with each latent topic were inspected to identify semantically coherent patterns. Conceptual domains were derived solely from these semantic regularities and the terminology contained in the retrieved literature; no theoretical structures external to the dataset were imposed. Based on the identified domains, a preliminary set of item ideas was drafted to support later validation stages; items are not reported here as they remain in an early exploratory form. Model adequacy, stability, and interpretability were evaluated through multiple diagnostic procedures, drawing on recommendations from probabilistic topic modelling (Blei et al., 2003; Griffiths & Steyvers, 2004), educational applications (Chen et al., 2021), and text-mining quality assurance in systematic reviews (O'Mara-Eves et al., 2015). First, sparsity of the document-term matrix (DTM) was examined. The corpus exhibited high sparsity (0.98), a pattern typical of natural-language educational corpora and consistent with LDA's generative assumptions. Second, model fit across alternative values of *k* was assessed using perplexity. Inspection of the perplexity curve indicated diminishing improvements beyond the selected four-topic solution, supporting its adequacy relative to more complex models. Third, semantic coherence was calculated to evaluate the co-occurrence reliability of high-probability terms using log co-occurrence scores. For the retained four-topic solution, topic-level coherence values ranged from –70.10 to –33.64, reflecting substantial lexical sparsity and thematic dispersion typical of large educational corpora. Log co-occurrence-based coherence metrics produce negative values because they rely on the logarithm of word co-occurrence probabilities. In sparse natural-language corpora, such probabilities are necessarily smaller than 1, resulting in negative log values. Fourth, topic stability was examined through nonparametric bootstrap resampling (50 replications). Stability was evaluated by inspecting the

consistency of high-probability terms across bootstrap solutions, which showed substantial overlap in the most salient terms within each topic. This procedure was intended as a qualitative robustness check rather than a formal inferential test. Finally, interpretability of document–topic distributions was assessed by examining posterior values, which indicated clear dominant-topic probabilities (γ = .39–1.00) and relatively balanced topic prevalence across the four topics.

3. Results

The LDA model identified four latent topics with distinct semantic profiles, as indicated by topic–term probability distributions. One topic primarily captured general educational discourse around AI, characterised by terms such as education, learning, teachers, students, artificial, and intelligence. A second topic reflected research and analytical perspectives, with prominent terms including *analysis*, *research*, *factor*, and *study*. A third topic centred on measurement and assessment, featuring terms such as *scale*, *factor*, *chatgpt*, and *students*. Finally, a smaller yet distinct topic was characterised by medical and clinical terminology, including *medical*, *health*, *clinical*, *study*, and *training*. Document-topic posterior distributions (γ) indicated that Topics 1, 2, and 3 accounted for most documents, whereas Topic 4 represented a smaller cluster. Each document showed a dominant topic assignment (γ range = .39–1.00), supporting the interpretability of the four-topic solution and a clear dominant-topic structure across documents. Analysis of lexical frequencies revealed a strong concentration of terms related to educational, analytical, and AI-related content. The most frequent tokens in the corpus were *education*, *analysis*, *artificial*, *intelligence*, *students*, and *learning*, each occurring over one thousand times in the pre-processed text. In contrast, emotionally relevant terms occurred at substantially lower frequencies. Tokens such as *anxiety*, *engagement*, *emotional*, *trust*, *confidence*, and *fear* together accounted for less than 0.2% of all tokens, as confirmed by aggregate frequency counts, indicating that emotional terminology constituted a marginal proportion of the lexical content in the analysed corpus. Emotional terminology was operationalised as a predefined set of affect-related tokens identified *a priori*. Relative frequencies were calculated as the proportion of total post-preprocessing tokens in the corpus, based on raw term counts rather than weighted measures.

4. Discussion

This study mapped the conceptual structure of recent literature on artificial intelligence in educational contexts with the aim of informing the early phases of developing a neuroinclusive measurement instrument. The topic-modelling results delineate a field that is currently structured around a limited set of dominant conceptual domains, with a clear emphasis on cognitive, analytical, and instructional aspects of AI use. These findings are consistent with accounts describing AI tools as cognitive partners that influence information processing, feedback interpretation, and self-regulatory demands in learning environments (Bosco, 2024; Pacheco et al., 2025). As such, the domains identified here provide an empirically grounded starting point for defining candidate constructs to be operationalised in subsequent scale-development phases. At the same time, the analysis revealed a marked underrepresentation of emotional terminology across the corpus. Although affective responses such as anxiety, trust, curiosity, and confidence are frequently discussed in conceptual and empirical work on human–AI interaction (Montag & Elhai, 2025; Naiseh et al., 2025), these constructs did not emerge as a coherent latent domain in the topic model and accounted for only a minimal proportion of lexical content. This pattern suggests that emotional processes are not systematically foregrounded within the Scopus-indexed literature captured by the present search strategy, despite their documented relevance for engagement, persistence, and help-seeking. From a measurement perspective, this gap highlights the need for future instruments to explicitly incorporate affective dimensions rather than relying solely on cognitively oriented constructs. A related observation concerns the absence of topics explicitly reflecting the involvement of different neurotypes in learning. Despite growing attention to neurodiversity in educational research (Dark, 2024), the corpus did not bear topics centred on neurodivergent learners or neurocognitive variability. This absence is more plausibly attributable to the current empirical focus of *AI in education* research rather than to a lack of theoretical relevance. Given established differences in executive functioning, attentional regulation, and uncertainty processing across neural profiles (Hutson, 2024), the exclusion of such perspectives may constrain understanding of how learners differentially experience and interpret AI-supported learning environments. Addressing this limitation represents an important direction for future research and measurement development. The topic structure also included a domain characterised by medical, clinical, and professional training terminology. This topic is best interpreted considering the disciplinary composition of the retrieved literature and the broad biomedical coverage of Scopus, rather than as a core educational

construct. Similar domain-dependent effects of corpus composition on text-mining outputs have been noted in large-scale methodological studies (O'Mara-Eves et al., 2015). While this domain is peripheral to the primary focus on educational and psychological measurement, its presence underscores the importance of contextualising topic-modelling outputs within the characteristics of the source database and exercising caution when generalising conceptual structures beyond the retrieved corpus. Findings portray a research landscape in which cognitive and analytical dimensions of AI use are well represented, whereas emotional, neurodiversity-relevant, and practice-oriented educational constructs remain comparatively underdeveloped. By systematically identifying these imbalances, the present study contributes an evidence-informed basis for subsequent stages of scale development. In particular, the results support a measurement agenda that moves beyond predominantly cognitive framings to incorporate affective processes and neuroinclusive considerations, thereby better capturing the multidimensional impact of AI in educational contexts. In the subsequent phase of scale development, the empirically derived domains will be translated into theoretically bounded constructs, which will guide item generation through a combined inductive–deductive procedure. Item formulation will be refined through expert content validation prior to formal psychometric testing.

5. Conclusion

The present study offers an evidence-based overview of the conceptual domains shaping current research on AI in education, providing a starting point for developing a novel neuroinclusive measurement instrument. The results highlight areas of conceptual concentration along with gaps (particularly in emotional, confidence-related, neurodiversity-relevant, and practice-oriented educational dimensions) that will require integration during item generation. The four identified domains will serve as an initial scaffold for constructing the exploratory item pool. The next steps include refining items, conducting expert content validation, and examining dimensionality, reliability, and measurement invariance to ensure that the emerging tool can adequately capture the multifaceted impact of AI on learning across diverse learner profiles.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research,* 3, 993-1022.

Bosco, A. (2024, March 15). *IA una questione di metodo?* In *Approcci collaborativi e Intelligenza Artificiale nelle organizzazioni: Il contributo delle Neuroscienze*. IrcCAN – International Research Center for Cognitive Applied Neuroscience, Università Cattolica del Sacro Cuore, Milan.

Chen, X., Zou, D., Xie, H., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning & Technology, 25*(3), 151-185.

Dark, J. (2024). Eight principles of neuro-inclusion; an autistic perspective on innovating inclusive research methods. *Frontiers in psychology*, 15, 1326536. https://doi.org/10.3389/fpsyg.2024.1326536

Grassini S. (2023). Development and validation of the AI attitude scale (AIAS-4): a brief measure of general attitude toward artificial intelligence. *Frontiers in psychology*, 14, 1191628. https://doi.org/10.3389/fpsyg.2023.1191628

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101 Suppl 1*(Suppl 1), 5228-5235. https://doi.org/10.1073/pnas.0307752101

Hutson, P. (2024). Embracing the irreplaceable: The role of neurodiversity in cultivating human-AI symbiosis in education. *International Journal of Emerging and Disruptive Innovation in Education: VISIONARIUM, 2*(1), 5. https://doi.org/10.62608/2831-3550.1020

Kieslich, K., Lünich, M., & Marcinkowski, F. (2021). The Threats of Artificial Intelligence Scale (TAI): Development, measurement and test over three application domains. *International Journal of Social Robotics, 13*(7), 1563-1577. https://doi.org/10.1007/s12369-020-00734-w

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24-59. https://doi.org/10.2307/256727

Mongeon, P., Paul-Hus, A. The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 106, 213-228 (2016). https://doi.org/-10.1007/s11192-015-1765-5

Montag, C., & Elhai, J. D. (2025). The darker side of positive AI attitudes: Investigating associations with (problematic) social media use. *Addictive behaviors reports*, 22, 100613. https://doi.org/10.1016/j.abrep.2025.100613

Naiseh, M., Babiker, A., Al-Shakhsi, S. *et al.* (2025). Attitudes Towards AI: The Interplay of Self-Efficacy, Well-Being, and Competency. *J. technol. behav. sci.* https://doi.org/10.1007/s41347-025-00486-2

Noordewier, M. K., & Gocłowska, M. A. (2024). Shared and unique features of epistemic emotions: Awe, surprise, curiosity, interest, confusion, and boredom. *Emotion, 24*(4), 1029-1048. https://doi.org/10.1037/emo0001314

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, *4*(1), 5. https://doi.org/10.1186/2046-4053-4-5

Pacheco, A. J., Boude Figueredo, O. R., Chiappe, A., & Fontán de Bedout, L. (2025). AI-powered learning analytics for metacognitive and socio-emotional development: A systematic review. *Frontiers in Education,* 10, 1672901. https://doi.org/-10.3389/feduc.2025.1672901

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2), 1-40. https://doi.org/10.186-37/jss.v091.i02

Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human–Computer Interaction*, *39*(13), 2724-2741. https://doi.org/10.1080/10447318.2022.2085400

Appendix A. Operational Scopus Search Strategy

The Scopus search strategy was operationalised to identify peer-reviewed journal articles addressing the intersection of artificial intelligence, educational contexts, and measurement or psychometric constructs. The search was conducted using the *TITLE-ABS-KEY* field to ensure coverage of titles, abstracts, and both author and index keywords. The operational search string was:

```
TITLE-ABS-KEY(
  scale OR questionnaire OR instrument OR psychometric OR validation
  OR “item development” OR “factor analysis” OR CFA OR EFA OR measure
)
AND
TITLE-ABS-KEY(
  “artificial intelligence” OR AI OR “intelligent system”
)
AND
TITLE-ABS-KEY(
  education OR learning OR student OR teacher OR school OR “higher education”
)
AND (PUBYEAR > 2019 AND PUBYEAR < 2026)
AND (LIMIT-TO(LANGUAGE, “English”))
AND (LIMIT-TO(DOCTYPE, “ar”))
```