Universidad Pablo de Olavide (España)
International Journal of Educational Research and Innovation
número 25, 2026
ISSN: 2386-4303
DOI: 10.46661/ijeri.12995
Sección: Artículos
Recibido: 28-01-2026
Aceptado: 08-03-2026
Publicado: 27-05-2026
Páginas: 1-16

INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH AND INNOVATION
REVISTA INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN EDUCATIVA

Explorar la Inteligencia Artificial en educación superior mediante Procesamiento del Lenguaje Natural Ligero: un estudio metodológico de prueba de concepto

Exploring Artificial Intelligence in Higher Education through Lightweight Natural Language Processing: A Proof-of-Concept Methodological Study

Antonio Matas Terrón
Universidad de Málaga
amatas@uma.es

José Manuel Ríos Ariza
Universidad de Málaga
jmrios@uma.es

Antonio Luque de la Rosa
Universidad de Almería
aluque@ual.es

José Jesús Sánchez Amate
Universidad de Almería
jsa819@ual.es

1. RESUMEN

La integración de la Inteligencia Artificial en la educación ha generado una amplia gama de perspectivas entre los actores educativos, a menudo capturadas a través de respuestas de encuestas abiertas. Sin embargo, los enfoques para analizar dichos datos bajo condiciones no ideales siguen siendo limitados. Este estudio examina la viabilidad de combinar respuestas abiertas con técnicas de "Procesamiento de Lenguaje Natural (PLN) ligero" para explorar el discurso sobre la IA en la educación superior. Con un diseño exploratorio de prueba de concepto, se contó con dos conjuntos de datos independientes de 31 estudiantes de Pedagogía de España y 35 docentes de Perú que respondieron a cuestiones abiertas similares sobre Inteligencia Artificial. Se recurrió a representaciones semánticas basadas en embeddings, derivadas de la ponderación de términos y la descomposición en valores singulares (SVD) truncada, junto con anclas semánticas y visualización exploratoria para analizar textos heterogéneos. Los resultados muestran que los flujos de trabajo semánticos ligeros pueden generar representaciones interpretables, apoyar la

Cómo citar: Matas Terrón, A; Ríos Ariza, J.M; Luque de la Rosa, A & Sánchez Amate, J.J. (2026). Explorar la Inteligencia Artificial en educación superior mediante Procesamiento del Lenguaje Natural Ligero: un estudio metodológico de prueba de concepto. *IJERI: International Journal of Educational Research and Innovation*, (25), 1-16. 10.46661/ijeri.12995

inspección basada en proximidad y permitir la exploración guiada por anclas incluso con muestras pequeñas. En concreto, los datos analizados revelan que, mientras el discurso estudiantil gravita hacia la utilidad pragmática y el aprendizaje autorregulado, la narrativa docente está dominada por la ética y la ansiedad regulatoria. Se concluye reflexionando sobre la capacidad de este enfoque como herramienta escalable para monitorizar el clima educativo en tiempo real.

PALABRAS CLAVE

Inteligencia artificial en la educación (AIEd); Respuestas abiertas a encuestas; Procesamiento ligero del lenguaje natural (Light NLP); Análisis semántico; Estudio metodológico exploratorio.

2. ABSTRACT

The integration of Artificial Intelligence (AI) in education has prompted a wide array of perspectives among educational stakeholders, frequently captured through open-ended survey responses. However, approaches to analysing such data under non-ideal conditions remain limited. This study examines the feasibility of combining open-ended responses with "Light Natural Language Processing" techniques to explore the discourse on AI in higher education. Employing an exploratory proof-of-concept design, the research utilised two independent datasets comprising 31 Pedagogy students from Spain and 35 teachers from Peru, all of whom responded to open-ended questions regarding Artificial Intelligence. Semantic representations based on embeddings—derived from term weighting and truncated Singular Value Decomposition (SVD)—were employed alongside semantic anchors and exploratory visualisation to analyse heterogeneous texts. The results demonstrate that light semantic workflows can generate interpretable representations, support proximity-based inspection, and enable anchor-guided exploration, even with small sample sizes. Specifically, the analysed data reveal that while student discourse is oriented towards pragmatic utility and self-regulated learning, the teacher narrative is dominated by ethics and regulatory anxiety. The paper concludes by reflecting on the potential of this approach as a scalable tool for monitoring the educational climate in real time.

KEYWORDS

Artificial Intelligence in Education (AIEd); Open;ended survey responses; Light Natural Language Processing (Light NLP); Semantic analysis; Exploratory methodological study.

1. INTRODUCTION

The emergence of Generative Artificial Intelligence (GenAI) has precipitated an unprecedented transformation across the educational ecosystem, creating what Zhang & Cao (2025) describe as a "digital disruption" with a profound impact on the identity and well-being of educational stakeholders. Although technological integration requires educators to develop new digital competencies, the abrupt implementation of GenAI has exacerbated phenomena such as "technostress" (Li et al., 2024; Trusz & Demeshkant, 2025), adversely affecting quality of life and mental health (Li et al., 2024; Liţan, 2025; Suso-Vega et al., 2024). Factors such as a lack of institutional support, uncertainties regarding professional identity, and concerns over privacy contribute to this tension (Khlaif et al., 2022). Simultaneously, Higher Education institutions face the challenge of navigating critical regulatory voids, where academic integrity and data ethics have become primary sources of institutional anxiety (García-López & Trujillo-Liñán, 2025; García-Peñalvo et al., 2024; Gavira & Jiménez-Preciado, 2025; Huang et al., 2023; Kangwa et al., 2025). Consequently, institutions are compelled to address these challenges through regulations that

promote safety and propriety in the face of the risks associated with unethical AI use (Cordón, 2023; Horban et al., 2025). Regarding these issues, in their meta-analysis, Chiu et al. (2023) assert the need for further research into ethical questions concerning AI in higher education. Within this context of GenAI use in education, the literature suggests two dominant dimensions:

1. Utility and Self-Regulated Learning: This dimension focuses on student expectations and usage, where AI is valued for its capacity to personalise learning and enhance academic efficiency. It is grounded in research exploring how AI alters teaching-learning dynamics (Chiu et al., 2023; Xu et al., 2025). These positive expectations are higher among students with advanced university training; however, while pragmatism and optimism dominate the student perspective, vague fears regarding the threats posed by AI also emerge (Bewersdorff et al., 2023). Regarding educators, some research indicates they perceive greater potential for AI as a tool for professional development than for teaching-learning processes (Álvarez-Herrero, 2024). This faculty perception is nuanced by studies indicating that, despite an openness towards technological evolution, there are significant concerns regarding the practical and ethical integration of AI in teaching (Antunes dos Santos & Reategui, 2025; Suso-Vega et al., 2024) particularly in Higher Education (Gavira & Jiménez-Preciado, 2025; Horban, et al., 2025; Ilieva et al., 2025; López-Chila et al., 2024; Villegas-José & Delgado-García, 2024; Wu & Yu, 2023), while presenting educators with the challenge of increasing knowledge on how students learn using AI (Antonenko & Abramowitz, 2023). Furthermore, a major concern is that AI use may compromise assessment reliability. Thus, assessment tools and processes must account for transparency, traceability, and the prevention of students' "cognitive delegation" to AI (Horban et al., 2025). Additionally, institutions must consider how GenAI use for assessment and feedback quality influences academic integrity (Ilieva et al., 2025).
2. Risks and Ethics: Centred on institutional and faculty perspectives, this dimension focuses specifically on regulation, data privacy, and plagiarism prevention (García-Peñalvo et al., 2024). Recent studies (Arranz-García et al., 2025; García-López & Trujillo-Liñán, 2025) indicate that this is the dominant framework in faculty discourse, often associated with resistance to change and professional anxiety stemming from the fear of AI as a rival rather than an ally (Mateus et al., 2024). It is, therefore, relevant to identify the factors influencing faculty resistance or adoption (Suso-Vega et al., 2024). Similarly, among the student misconceptions that entail risk, Bewersdorff et al. (2023) note that AI can yield incorrect results, biases, and manipulations (Kangwa et al., 2025)—issues that also affect a portion of the faculty (Antonenko & Abramowitz, 2023).

Student self-regulation is a key factor in achieving greater integrity when using GenAI as a learning tool, making institutional regulation vital in shaping student usage (Kangwa et al., 2025). As Chiu et al. (2023) note, there is an urgent need to investigate ethical issues and perceptions not solely through an engineering lens but by integrating the voices of teachers and students via novel research methods.

Researching these two dimensions highlights the difficulty of capturing such complex psychosocial dynamics. To advance these themes, researchers increasingly rely on survey instruments with open-ended questions. While closed-ended instruments (e.g., Likert scales) effectively measure established constructs—such as "Intention to Use" in the UTAUT2 model (Sergeeva et al., 2025, Cabero-Almenara et al., 2026)—they often fail to reveal emerging concerns and contextual nuances invisible to standard metrics (García-López & Trujillo-Liñán, 2025). However, analysing such textual data presents a persistent methodological dilemma: traditional qualitative approaches provide depth but are difficult to scale and prone to overwhelming the researcher, while conventional computational methods lack the interpretive sensitivity required for the social sciences (Grimmer et al., 2022).

Historically, this analysis required intensive manual coding. Recently, the availability of Large Language Models (LLMs) like GPT-4 has tempted researchers to automate the process. While promising, these models operate as "black boxes" that present significant challenges regar-

ding reproducibility, hallucinations, and costs, questioning their suitability as a sole standard for scientific rigour (Zhang et al., 2025; Chen & Shu, 2023). Furthermore, delegating interpretation to commercial models raises ethical concerns regarding participant data privacy (Bover, 2013).

There is, therefore, a critical need for Natural Language Processing (NLP) methodologies that are accessible, transparent, and auditable. This study addresses this gap by presenting a "Light Semantic Analysis" framework. Unlike generative LLMs, approaches such as Truncated Singular Value Decomposition (SVD) transform texts into mathematically transparent vector spaces (Deerwester et al., 1990), allowing for the measurement of semantic proximity without imposing the training bias of a massive external model (Nelson, 2020).

In this context, the present study seeks to contribute to the methodological discussion on how open-ended educational data can be analysed using transparent and accessible computational approaches. Rather than relying on large proprietary language models, this research explores the potential of lightweight Natural Language Processing workflows that combine term-weighting techniques, dimensionality reduction, and theoretically informed semantic anchors.

In this sense, the objective of this study is to examine the feasibility and analytical value of combining open-ended survey responses with these light NLP techniques to explore the discourse of educational stakeholders. Specifically, the study pursues three interrelated objectives:

1. Methodological validation in ecological settings: To demonstrate the robustness of embedding-based semantic representations (truncated SVD) in extracting coherent patterns under real-world research conditions, characterised by heterogeneity and data scarcity (microtexts).
2. Theoretical anchoring of discourse: To evaluate the efficacy of "semantic anchors" as a tool for projecting predefined theoretical dimensions—specifically the tension between Utility/Efficiency and Risk/Ethics—onto unstructured empirical data.
3. Differential characterisation of perception: To apply this workflow to explore and visualise the discursive dissonance between the two participating educational stakeholder groups: Pedagogy students (Spain) and university teachers (Peru).

2. METHOD

Research Design

This study adopts an exploratory methodological design with a proof-of-concept orientation. It is a validation design aimed at demonstrating the robustness and transferability of Light Semantic Analysis (LSA) techniques within educational research concerning Generative Artificial Intelligence. Priority is given to ecological validity (Zhang et al., 2025), utilising real-world data that exhibit characteristics typical of digital social research: heterogeneity in textual length, unstructured noise, and discursive variability.

Unlike fully automated approaches, this proposal employs embedding-based semantic representations to facilitate transparent exploration. We validate this approach by analysing the contrasts between two corpora: student narratives, centred on self-regulated learning (López-Chila et al., 2024; Xu et al., 2025), and faculty discourse, marked by ethical caution (Almazán-López et al., 2025). The objective is to illustrate how light computational tools can reveal latent structures—such as the dichotomy between pragmatism and ethics—which are essential for the design of educational policies.

Theoretical Instrumentation: Semantic Anchors

To provide theoretical meaning to mathematical vector spaces, the "Semantic Anchors" technique is employed, projecting validated concepts as exploratory "lenses" (Nelson, 2020). Three discursive dimensions were defined:

- Perceived Utility (Pragmatism): Derived from the UTAUT2 model, capturing discourse on efficiency and performance (Sergeeva et al., 2025).
- Risk and Ethics (Normative): Grouping concerns regarding academic integrity, privacy, and regulation.
- Learning and Pedagogy (Transformation): Referring to the qualitative impact on cognitive processes, beyond mere utility.

Participants and Characterisation of Textual Corpora

To ensure the robustness of the analytical workflow, two independent datasets (65 participants) with contrasting roles and contexts were intentionally selected (see table 1).

- Corpus A (Estudiantes - España): Compuesto por reflexiones de 31 estudiantes de Pedagogía de la Universidad de Málaga (80 % mujeres; edad media=20.32, desviación típica=1.45). Se utilizó una pregunta única abierta sobre el impacto futuro de la IA, generando un corpus de narrativas de expectativa.
- Corpus B (Docentes - Perú): Formado por 34 docentes universitarios de diversas instituciones peruanas (33 % mujeres; edad media=58.69, desviación típica=8.74). A diferencia del grupo anterior, la recolección se estructuró mediante tres preguntas abiertas (rol, preocupaciones y condiciones), generando "microtextos" breves y fragmentados.

Both corpora were treated as separate analytical entities without post-hoc harmonisation, demonstrating the workflow's capacity to process divergent textual structures.

Table 1. Sample description.

	Students (n=31)	Teachers (n=34)
Sex	80 % Famale	33 % Female
Country	Spain	Peru
Age	mean*=20.32; SD*=1.45 years	mean= 58.69; SD=8.74 years
Digital self-efficacy perceived level (from 1 to 10)	NA	mean=7.58; SD=1.82 points
Academic performance self-efficacy perceived	mean= 3.87; DT=0.76 points	NA

Note: *=from population scores; NA= not available; SD= standar deviation

Analysis Procedure

An independent process was applied to each corpus following four phases:

1. Minimalist Preprocessing: A "conservative cleaning" was performed using quanteda (Benoit et al., 2018), normalising characters and removing functional stopwords, while avoiding lemmatisation to preserve semantic nuances.
2. Semantic Space Construction (SVD): Texts were transformed into a TF-IDF matrix to penalise generic terms. Subsequently, Truncated Singular Value Decomposition (SVD) (Ooms, 2024) was applied, projecting the texts into a dense vector space where geometric proximity reflects similarity of meaning (LSA).
3. Anchor Projection: Rather than "blind" clustering, cosine similarity was calculated between each response and the vectorised definitions of the theoretical anchors, labelling each text according to its "dominant discursive orientation."
4. Visualisation (UMAP): The vector representations were projected into 2D using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) to inspect the topology of the discourse.

Ethical Considerations

The study adhered to the Declaration of Helsinki and the GDPR (2016/679). Participation was voluntary, non-remunerated, and mediated by explicit digital informed consent, which detailed the study objectives, data processing, and the right to withdraw at any time without penalty. No sensitive data (health, political ideology, or religion) or direct identifiers (names or emails) were collected.

Given that the analysis of open-ended responses carries risks of indirect re-identification through narrative context, a pseudonymisation protocol was applied. Prior to computational processing, a manual sweep was conducted to remove or substitute specific references (names of institutions, departments, or colleagues) that could compromise participant anonymity.

Finally, all Natural Language Processing (NLP) was conducted locally using open-source libraries in R (quanteda, text2vec). This ensured that raw data remained under the exclusive custody of the researchers, eliminating the risk of third-party exposure or usage for training commercial models.

Software and Reproducibility

All analyses were conducted using the open-source R statistical environment (R Core Team, 2024) to ensure replicability. Data manipulation relied on the tidyverse ecosystem (Wickham et al., 2019). The exclusive use of R libraries ensures the workflow is auditable, transparent, and adaptable to other educational contexts without reliance on proprietary software or "black-box" models (see Appendix for the link to the repository containing R code and coded datasets).

3. RESULTS

The results are presented as two independent case studies. Firstly, the student group (Spain) is analysed; this group is characterised by high fragmentation (microtexts), allowing for an evaluation of the model's capacity to extract semantic coherence under conditions of information scarcity. Secondly, the faculty group (Peru) is examined, featuring more extensive narratives to test the model's sensitivity to dense discourse.

1. Case Study A: The Student Perspective (Managing Microtexts)

1.1.Characterisation and Coherence of the Semantic Space

The analysis of the 31 students addressed the challenge of extreme brevity: 84% of responses contained fewer than 30 words. Despite this lack of lexical co-occurrences—which would typically hinder traditional Latent Dirichlet Allocation (LDA)—dimensionality reduction via Truncated Singular Value Decomposition (SVD) successfully inferred robust latent relationships. Nearest neighbour analysis demonstrated that the model captured conceptual affinities beyond literal word matching. For instance, brief responses regarding "changes in learning" were correctly grouped with more elaborate reflections on educational transformation (cosine similarity >.80), as evidenced in Table 2.

Tabla 2. Examples of semantic proximity and student anchoring.

Dominant Anchor	Original Text (Example)	Similarity (Cosine)
Learning	"Cambio en el aprendizaje" [Changes in learning]	0.70
Utility	"Es una buena herramienta para buscar soluciones rápidas pero no te ayuda a entender o interiorizar un aprendizaje" [It's a good tool for finding quick solutions, but it doesn't help you understand or internalize learning]	0.72
Risk/Ethics	"Pienso que un uso excesivo es malo, pero que nos ayuda bastante, sobre todo a los universitarios" [I think excessive use is bad, but it helps us quite a bit, especially university students]	0.86

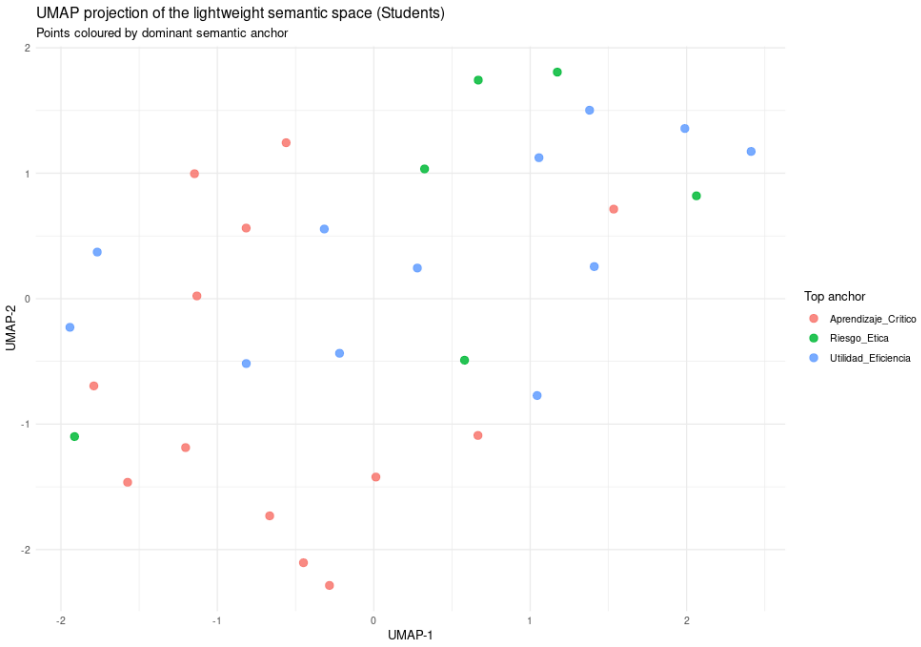
1.2.Anchor Projection and Latent Structure

Upon projecting the theoretical anchors, two clear patterns emerged:

1. Pragmatic Activation: The "Utility/Efficiency" and "Learning" anchors showed strong vectorial alignment, suggesting that student discourse gravitates toward functional impact.
2. Ethical Latency: Contrary to the initial hypothesis, the "Risk/Ethics" anchor was not absent but displayed "weak" and fragmentary activation. Normative concerns appeared subordinate to utility (e.g., "it is bad, but it helps"), without articulating a dense ethical vocabulary.

The UMAP projection (Figure 1) confirms this dispersion; no compact clusters are observed, but rather a diffuse regional trend of separation between responses oriented toward "efficiency" and those focused on "learning".

Figure 1. UMAP projection of the semantic space of students. The colors indicate the dominant theoretical anchor (Utility vs. Learning).



2. Case Study B: The Faculty Perspective

2.1. Discursive Density and Ethical Emergence

The faculty corpus (n=34), constructed through question triangulation, exhibited greater textual richness (mean length of 72 words). This "high-density" profile allowed for testing the model's robustness within complex narratives. As shown in Table 3, the semantic representation maintained its stability, coherently linking reflections on professional adaptation and instrumental use.

Table 3. Examples of semantic proximity and faculty anchoring.

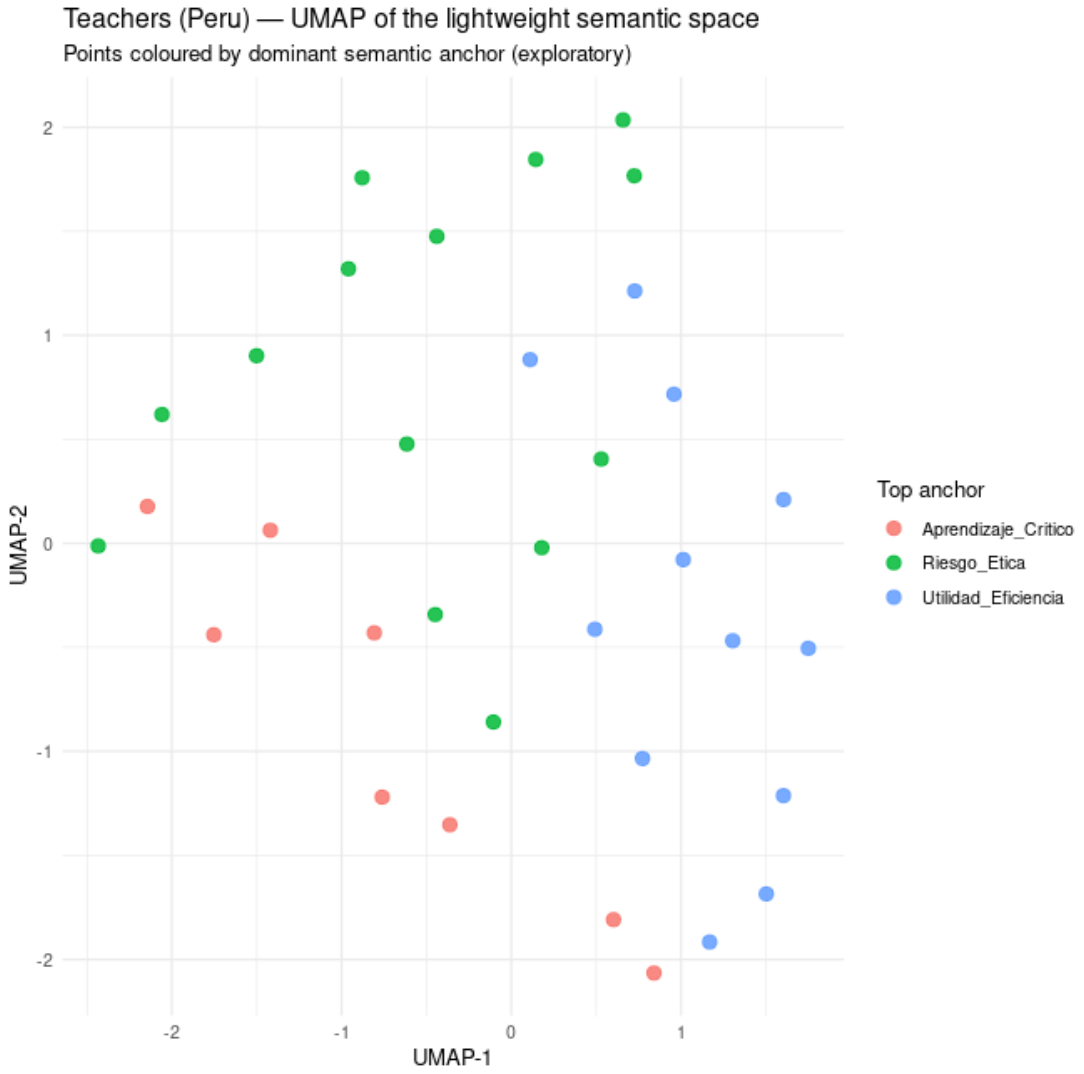
Dominant Anchor	Original Text (Example)	Similarity (Cosine)
Learning	"La AI es una herramienta indispensable en el actual contexto de desarrollo Tecnológico que vive la humanidad, ..." [AI is an indispensable tool in the current context of technological development that humanity is experiencing...]	0.70
Utility	"La IA me ayuda mucho en mi labor docente, la agiliza, la simplifica..." [AI helps me a lot in my teaching work; it streamlines it, simplifies it...]	0.70
Risk/Ethics	"La IA es una herramienta disruptiva que marca un antes y un después en la humanidad. Vino para quedarse y no queda otra cosa que adecuarse a ella y darle la mayor y mejor utilidad posible...." [AI is a disruptive tool that marks a turning point for humanity. It's here to stay, and we have no choice but to adapt to it and make the most of it...]	0.55

2.2. Anchor Projection and Topological Mapping

The most significant finding was the intense activation of the ethical dimension. Unlike the students, the Risk/Ethics anchor emerged as the dominant orientation for a considerable subset of responses, confirming that regulatory concerns are central to professional teaching identity.

The UMAP visualisation (Figure 2) reveals a more complex topology where all three orientations coexist. The spatial proximity between responses labelled as "risk" and "utility" suggests that, for faculty, these are not mutually exclusive concepts but rather simultaneous tensions that must be navigated within their practice.

Figura 2. Proyección UMAP del espacio semántico de docents.



3. Comparative Cross-Analysis

The cross-analysis validates the robustness of the method against opposing linguistic properties. Beyond technical validation, a qualitative structural divergence was detected (see table 4). While an instrumental logic with an implicit or "silenced" ethics predominates among students, an explicit "ambivalent tension" is observed among faculty, where technological adoption is heavily conditioned by normative and deontological barriers.

Table 4. Comparative Semantic Activation Matrix.

Dimension (Anchor)	Students (Microtexts)	Faculty (Narratives)	Model Interpretation
Utility	High Activation (Dominant)	High Activation	Cross-sectional consensus on functional utility.
Learning	Medium Activation	Medium-High Activation	Shared concern regarding cognitive impact.
Risk/Ethics	Low / Latent	High Activation (Dominant)	Discursive Gap: Normative concern is central for faculty; incipient and subordinate for students.

4. DISCUSSION

This study set out to validate a Light Semantic Analysis (LSA) methodological framework for exploring perceptions of Artificial Intelligence within educational contexts. Beyond the technical confirmation of the procedure's feasibility, the results illustrate the model's capacity to extract coherent discursive profiles across radically different scenarios. Whilst the geographical and role-based differences between the samples (Spanish students vs. Peruvian faculty) preclude a direct inferential comparison, the analysis confirms that the methodology is sufficiently sensitive to capture the situational specificity of each group: from instrumental pragmatism in the student sample to normative concern within the faculty context. These substantive and methodological implications are discussed below.

The primary finding lies not in the inter-group comparison, but in the efficacy of anchor-based analysis for detecting the dominant orientation inherent to each context without direct human supervision. The method demonstrated an ability to correctly "read" the theoretical literature underlying each independent sample.

In the student corpus, the strong alignment with the Utility/Efficiency anchor and the sparse lexical density regarding the Risk/Ethics anchor validate the method's capacity to identify performance-centred discourse. This characterisation resonates with the UTAUT2 model, where performance expectancy is typically the strongest predictor of adoption among university students (Sergeeva et al., 2025). Semantic analysis accurately captured how these participants conceptualise AI as an instrument for self-regulated learning (Gavira & Jiménez-Preciado, 2025; López-Chila et al., 2024; Xu et al., 2025), prioritising pragmatism. The tool proved its utility here not by detecting a total "absence" of ethics, but by revealing its discursive latency: normative concerns do exist, but they appear subordinate and hold less specific weight compared to immediate utility.

Conversely, when applying the same method to the faculty corpus, a qualitatively distinct semantic profile emerged, dominated by the Risk/Ethics anchor. This is consistent with previous research that accords central importance to this dimension (Álvarez-Herrero, 2024; García-Peñalvo et al., 2024; Mateus et al., 2024); notably, the study by Mateus et al. includes Peruvian teachers in its sample. This suggests that the model is robust enough to capture complex psychosocial phenomena such as "technostress" and regulatory anxiety (Liţan, 2025; Suso-Vega et al., 2024; Zhang & Cao, 2025). The visualisation of the coexistence between the "Utility" and "Risk" dimensions confirms that Light Semantic Analysis can represent the professional paradox described

by García-López and Trujillo-Liñán (2025): the recognition of pedagogical potential eclipsed by ethical responsibility, which necessitates a balanced training approach to mitigate detected risks (Sahín et al., 2024). Thus, the method is validated by demonstrating it can discriminate high-level concerns (integrity, regulation) and detect a profile of explicit caution among faculty (Almazán-López et al., 2025).

From a methodological perspective, this study provides empirical evidence for the debate on the use of computational tools in the social sciences, regardless of data origin. Faced with the trend of utilising Large Language Models (LLMs) for automated coding—which often operate as costly and opaque "black boxes" (Zhang et al., 2025)—our approach demonstrates that light, transparent techniques (SVD and Semantic Anchoring) are sufficient for extracting interpretable patterns under diverse data conditions. The model's ability to function with both student micro-texts and faculty narratives validates its technical transferability and robustness against the linguistic heterogeneity typical of educational research (Grimmer et al., 2022).

Unlike opaque approaches, the use of semantic anchors allows for full traceability: we can mathematically explain why a response aligns with "Ethics" based on the vector proximity of its vocabulary. This answers Nelson's (2020) call for a "computational grounded theory" that not only classifies but also facilitates theoretical abduction. Furthermore, by being conducted within a local environment (R), this workflow ensures data privacy, overcoming the ethical dilemmas of sharing sensitive narratives with third-party commercial APIs.

Despite the robustness of the workflow, several significant limitations must be acknowledged. Firstly, the disparate nature of the samples precludes any comparative generalisation; the results should be interpreted as two independent case studies validating the method, rather than a direct cross-cultural comparison. Secondly, the limited size of the corpora in this proof-of-concept design restricts population-level statistical inference. Finally, the "bag-of-words" approach inherent to TF-IDF ignores syntactic sequencing, which may result in the loss of nuances such as negation—although the use of SVD partially mitigates this by capturing latent synonymies.

Future research should:

a) Apply this design to paired samples (within the same institutional context) to enable legitimate direct comparisons.

b) Scale this method to massive corpora to longitudinally monitor the institutional climate regarding AI.

c) Integrate sentiment analysis to disaggregate emotional valence within each anchor (e.g., distinguishing between fear and ethical commitment).

5. CONCLUSIONS

The emergence of Generative Artificial Intelligence has placed educational research at a dual crossroads as the need to understand the psychological adaptation of educational stakeholders and the urgency of adopting analytical methods that are scalable without sacrificing scientific transparency. This study validates Light Semantic Analysis not merely as a viable technical alternative, but as a necessary methodological framework for navigating this complexity.

From a theoretical perspective, the findings contribute to the literature on Artificial Intelligence in Education by illustrating how discursive orientations toward AI can vary according to the role of educational actors. While student responses tend to frame AI primarily in instrumental and learning-related terms, faculty narratives incorporate stronger references to ethical considerations, responsibility, and regulatory concerns. Although these observations should not be interpreted as inferential comparisons, they highlight the importance of examining how different educational stakeholders discursively frame technological change within higher education contexts. By applying the same workflow to student microtexts and faculty narratives, we have identified a

structural dissonance within the educational ecosystem: a gap between students' pragmatic adoption, driven by efficiency, and faculty caution, mediated by ethics and regulatory anxiety.

In an other hand, and in contrast to the growing reliance on "black-box" models that privatise data interpretation, our results demonstrate that it is possible to construct robust and auditable semantic representations using accessible computational resources.

Future research should extend this methodology to longitudinal designs that monitor the evolution of technostress and ethical competence, thereby consolidating an integration of computational techniques into the social sciences that empowers all participating educational agents (Villarrubia Zúñiga et al., 2025) and remains, above all, human, transparent, and reflexive. Further, they could extend the present approach in several directions. First, applying this analytical workflow to paired samples within the same institutional or national context would allow for more direct comparisons between different educational stakeholders. Second, scaling the method to larger datasets could enable longitudinal monitoring of institutional discourse on artificial intelligence adoption in higher education. Third, integrating complementary techniques such as sentiment analysis or discourse polarity detection may provide further insight into the emotional tone associated with different semantic anchors. Finally, combining lightweight semantic analysis with qualitative interpretative approaches could strengthen the dialogue between computational text analysis and traditional qualitative research traditions in the social sciences.

Ultimately, this work's most notable contribution to the topic is providing the scientific community with a reproducible framework for "qualitative listening at scale". By combining theoretical anchors with exploratory visualisation, we transform heterogeneous open data into precise institutional diagnostics, surmounting the limitations of traditional surveys. Ultimately, this approach democratises access to advanced text mining, allowing researchers to maintain sovereignty over their data and analysis.

CONTRIBUTION

Each author mentioned has made a substantial, direct, and intellectual contribution to the scientific article in all phases of the project: conceptualization, data curation, formal analysis, methodology, drafting, and writing. Each author has approved its publication.

DATA ACCESSIBILITY AND AVAILABILITY STATEMENT

To ensure scientific transparency and reproducibility, the full analytical pipeline—including the R scripts for preprocessing, Truncated SVD construction, and UMAP visualisation—is available on the GitLab (https://gitlab.com/amataste/proof-concept-to-exploring-ai-in-higher-education) However, in strict accordance with the General Data Protection Regulation (GDPR 2016/679) and the study's pseudonymisation protocol, the raw textual datasets are not publicly shared to protect participant anonymity. The researchers maintain exclusive custody of the raw data to eliminate any risk of exposure to third-party commercial models

STATEMENT ON ETHICS

Ethical integrity was maintained throughout the study by adhering to the Declaration of Helsinki. All participants provided explicit informed written consent, confirming their voluntary agreement to take part in the research.

FUNDED:

The results presented in this study are part of the research project "Data literacy in higher education in Andalusia: Overcoming barriers to digital success" (No. 941.126), funded by the University of Malaga through the B.3 grant call linked to the II Own Plan for Research, Transfer and Scientific Dissemination.

ACKNOWLEDGEMENTS

Authors would like to thank ITED (Institute of Technology for Education) in Peru for their collaboration in data collection in Peru.

REFERENCIES

Almazán-López, O., Hasbún, H., & Osuna-Acedo, S. (2025). Inteligencia Artificial Generativa e identidad (pos) digital docente. *IJERI: International Journal of Educational Research and Innovation*, (24), 1-17. https://doi.org/10.46661/ijeri.11160

Álvarez-Herrero, J.-F. (2024). Opinion of Spanish Teachers About Artificial Intelligence and Its Use in Education. En S. Papadakis (Ed.), *IoT, AI, and ICT for Educational Applications* (pp. 163-172). Springer Nature. https://doi.org/10.1007/978-3-031-50139-5_8

Antonenko, P. & Abramowitz, B. (2023) In-service teachers' (mis)conceptions of artificial intelligence in K-12 science education, *Journal of Research on Technology in Education*, *55*(1), 64-78. https://doi.org/10.1080/15391523.2022.2119450

Antunes dos Santos, R., y Reategui, E. B. (2025). Uso de inteligencia artificial generativa y análisis de palabras clave para apoyar la planificación de proyectos de investigación en la educación superior. *RELATEC. Revista Latinoamericana de Tecnología Educativa*, *24*(2), 87-104. https://doi.org/10.17398/1695-288X.24.2.87

Arranz-García, O., Romero García, M. del C., & Alonso-Secades, V. (2025). Perceptions, strategies, and challenges of teachers in the integration of artificial intelligence in primary education: A systematic review. *Journal of Information Technology Education: Research*, *24*, Article 6. https://doi.org/10.28945/5458

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bewersdorff, A., Zhai, X., Roberts, J., & Nerdel, C. (2023). Myths, mis-and preconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence*, *4*, 100143. https://doi.org/10.1016/j.caeai.2023.100143

Bover, A. (2013). Herramientas de reflexividad y posicionalidad para promover la coherencia teórico-metodológica al inicio de una investigación cualitativa. *Enfermería Clínica*, *23*(1), 33-36. https://doi.org/10.1016/j.enfcli.2012.11.007

Cabero-Almenara, J., Palacios-Rodríguez, A., Llorente-Cejudo, C., y Barroso-Osuna, J. (2026). Aceptación de ChatGPT en Educación Superior: Actitudes y Percepciones del modelo UTAUT2. REICE. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *24*(1), 1-17. https://doi.org/10.15366/reice2026.24.1.001

Chen, C., & Shu, K. (2023). Combating Misinformation in the Age of LLMs: Opportunities and Challenges. *arXiv Computers and Society*, arXiv:2311.05656. https://doi.org/10.48550/arXiv.2311.05656

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cordón, O. (2023). Inteligencia Artificial en Educación Superior: Oportunidades y Riesgos. *RiiTE Revista Interuniversitaria de Investigación en Tecnología Educativa*, (15), 16-27. https://doi.org/10.6018/riite.591581

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

García Peñalvo, F. J., Alier, M., Pereira, J., & Casany, M. J. (2024). Safe, Transparent, and Ethical Artificial Intelligence: Keys to Quality Sustainable Education (SDG4). *IJERI: International Journal of Educational Research and Innovation*, (22), 1-21. https://doi.org/10.46661/ijeri.11036

García-López, I. M., & Trujillo-Liñán, L. (2025). Ethical and regulatory challenges of Generative AI in education: A systematic review. *Frontiers in Education*, (10), 1565938. https://doi.org/10.3389/feduc.2025.1565938

Gavira Durón, N., & Jiménez-Preciado, A. L. (2025). Exploring the role of AI in higher education: A natural language processing analysis of emerging trends and discourses. *The TQM Journal*, *37*(19). https://doi.org/10.1108/TQM-10-2024-0376

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Horban, O., Stadnyk, M., Vintoniv-Bakharieva, S., Panasiuk, L., & Yatyshchuk, O. (2025). Artificial Intelligence as an Integral Component of the Digital Culture within Contemporary Higher Education. *Journal of Educational Technology and Learning Creativity*. *3*(2), 451-467. https://doi.org/10.37251/jetlc.v3i2.2333

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, *4*(4), 799-819. https://doi.org/10.1109/TAI.2022.3194503

Ilieva, G., Yankova, T., Ruseva, M., & Kabaivanov, S. (2025). A Framework for Generative AI-Driven Assessment in Higher Education. *Information*, *16*(6), 472. https://doi.org/10.3390/info16060472

Kangwa, D., Msafiri, M. M., & Fute, A. 2025. Exploring the Factors That Promote a Balance Between Academic Integrity and the Effective Use of GenAI Tools in Higher Education: A Systematic Review. *Journal of Computer Assisted Learning*, *41*(5), e70109. https://doi.org/10.1111/jcal.70109

Khlaif, Z., Sanmugam, M., Joma, A., Odeh, A., & Barham, K. (2022). Factors Influencing Teacher's Technostress Experienced in Using Emerging Technology: A Qualitative Study. T*echnology, Knowledge and Learning*, *28*, 865-899. https://doi.org/10.1007/s10758-022-09607-9

Li, L., Li, L., Zhong, B., & Yang, Y. (2024). A scientometric analysis of technostress in education from 1991 to 2022. *Education and Information Technologies*, *29*, 23155-23183. https://doi.org/10.1007/s10639-024-12781-1

Lițan, D. E. (2025). The impact of technostress generated by Artificial Intelligence on the quality of life: The mediating role of positive and negative affect. *Behavioral Sciences*, *15*(4), 552. https://doi.org/10.3390/bs15040552

López-Chila, R., Llerena-Izquierdo, J., Sumba-Nacipucha, N., & Cueva-Estrada, J. (2024). Artificial Intelligence in Higher Education: An Analysis of Existing Bibliometrics. *Education Sciences*, *14*(1), 47. https://doi.org/10.3390/educsci14010047

Mateus, J., Lugo, N., Cappello, G., & Guerrero-Pico, M. (2024). Communication educators facing the arrival of generative artificial intelligence: Exploration in Mexico, Peru, and Spain. *Digital Education Review*, (45), 38-55. https://doi.org/10.1344/der.2024.45.106-114

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*. https://doi.org/10.21105/joss.00861

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, *49*(1), 3-42. https://doi.org/10.1177/0049124117729703

Ooms, J. (2024). *irlba: Fast truncated singular value decomposition and principal components analysis for large sparse matrices* (R package version 2.3.5.1) [Software de ordenador]. CRAN. https://CRAN.R-project.org/package=irlba

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sahin, C. (2024). Artificial intelligence technologiesand ethics in educational processes: solution suggestions and results. *Innoeduca. International Journal of Technology and Educational Innovation*, *10*(2), 201-216. https://doi.org/10.24310/ijtei.102.2024.19806

Sergeeva, O. V., Zheltukhina, M. R., Shoustikova, T., Tukhvatullina, L. R., Dobrokhotov, D. A., & Kondrashev, S. V. (2025). Understanding higher education students' adoption of generative AI technologies: An empirical investigation using UTAUT2. *Contemporary Educational Technology*, *17*(2), ep571. https://doi.org/10.30935/cedtech/16039

Suso-Vega, J. A., Meneses-La-Riva, M. E., & Fernández-Bedoya, V. H. (2024). Adapting to a New Normal: Peruvian University Faculty's Experiences with Techno-Stress Post-Covid-19. *F1000Research*, (12), 1381. https://doi.org/10.12688/f1000research.141432.3

Trusz, S., & Demeshkant, N. (2025). Teachers' technological, pedagogical, and content knowledge related to artificial intelligence as a protective factor against technostress and techno-anxiety. *Studia z Teorii Wychowania*, *16*(4), 303-327. https://doi.org/10.5604/01.3001.0055.5084

Villarrubia Zúñiga, M. S., Ortiz Jiménez, M., y González García, P. (2025). Artificial intelligence chatbots in language learning for disadvantaged populations (migrants and aboriginal): State of the art and challenge. EDMETIC, *Revista de Educación Mediática y TIC*, *14*(2), 1-21. https://doi.org/10.21071/edmetic.v14i2.17263

Villegas-José, V., y Delgado-García, M. (2024). Inteligencia artificial: revolución educativa innovadora en la Educación Superior. Pixel-Bit. *Revista de Medios y Educación*, *71*, 159-177. https://doi.org/10.12795/pixelbit.107760

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, R., & Yu, Z. (2023). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 1–24. https://doi.org/10.1111/bjet.13334

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *Computers and Education: Artificial Intelligence*, *8*, 100384. https://doi.org/10.1016/j.caeai.2025.100384

Zhang, H., & Cao, J. (2025). From digital disruption to mental health: The impact of AI-induced educational anxiety on teacher well-being in the era of smart education. *BMC Public Health*, (25), 4010. https://doi.org/10.1186/s12889-025-25372-7

Zhang, S., Xu, J., & Alvero, A. J. (2025). Generative AI meets open-ended survey responses: Research participant use of AI and homogenization. *Sociological Methods & Research*, *54*(3), 1197-1242. https://doi.org/10.1177/00491241251327130

ANNEX

- Annex I:
  - Students form: https://forms.gle/as6tsfcJu35qwLps9
  - Teachers form: https://forms.gle/qRgtHY3KatyzNF4h6
- Annexo II:
  - R code repository: https://gitlab.com/amataste/proof-concept-to-exploring-ai-in-higher-education