Scientia et PRAXIS

Innovating sustainable artificial intelligence citizenship: a qualitative study of the CAITIZEN model using ATLAS.ti

Innovando la ciudadanía sostenible en inteligencia artificial: un estudio cualitativo del modelo CAITIZEN usando ATLAS.ti

Juan Mejía-Trejo ORCID 0000-0003-0558-1943
Centro Universitario de Ciencias Económico Administrativas (**CUCEA**)
Universidad de Guadalajara (**UdeG**)
Zapopan, Jalisco, México
e-mail: jmejia@cucea.udg.mx

ABSTRACT

Context. The rapid integration of artificial intelligence is reshaping social, economic, and civic contexts worldwide. In higher education, **university students, understood as citizens in formation**, increasingly engage with **AI** for learning, creativity, and decision making, influencing academic practices, ethical reasoning, civic participation, and social responsibility across institutional cultures.

Problem. Current **AI** education prioritizes efficiency and automation, marginalizing, AI literacy, ethics, algorithmic fairness, data justice, and creative collaboration, among several issues creating tension between adoption and higher education's mission to form citizens.

Purpose. This study aims to develop and qualitatively substantiate the **CAITIZEN model** as a multidisciplinary framework for understanding **AI** assisted citizenship in formation. Aligned with the **OECD** Oslo Manual, the model is positioned as conceptual social innovation integrating five dimensions: Critical Artificial Intelligence Literacy (**CAIL**); Ethical Awareness and Responsibility (**EAR**); Awareness of Fairness and Data Justice (**AFDJ**); Human **AI** Creative Collaboration (**HAIC**); and Metacognitive Transparency in Prompting Practices (**MTPP**), contributing to the Sustainable Development Goals of the 2030 Agenda.

Methodology. A qualitative research design was implemented between **July and December 2025** in Jalisco, Mexico. The study involved **511 undergraduate and graduate students**. Data were collected through a **55 item questionnaire** structured around five analytical dimensions and distributed via **Google Forms**. The corpus was analyzed using thematic analysis supported by **ATLAS.ti 25**.

Theoretical and Practical Findings. The study reconceptualizes **AI** use as an ethical–cognitive– social system, advancing theory through an integrated literacy model and offering practical guidance for curriculum design, responsible **AI** use, and institutional oversight.

Originality. It lies in linking educational innovation with the **SDGs** by integrating AI literacy, ethics, fairness, prompt clarity, and human–**AI** collaboration as pillars of sustainable citizenship..

Conclusions. The study demonstrates that **AI** use in higher education functions as an ethical– cognitive–social system through the **CAITIZEN model.** Its qualitative, context-specific design represents potential for future expansion through comparative and longitudinal research, enhancing transferability and broader applicability.

RESUMEN

Contexto. La integración acelerada de la inteligencia artificial (**IA**) está transformando los contextos sociales, económicos y cívicos a escala global. En la educación superior, los estudiantes universitarios, entendidos como ciudadanos en formación, interactúan crecientemente con sistemas de **IA** en procesos de aprendizaje, producción creativa y toma de decisiones. Estas interacciones influyen en las prácticas académicas, el razonamiento ético, la participación cívica y la responsabilidad social dentro de culturas institucionales digitalmente mediadas.

Problema. Los enfoques educativos actuales sobre la **IA** privilegian la eficiencia, la automatización y el desempeño técnico, relegando dimensiones clave como la alfabetización en **IA**, la ética, la equidad algorítmica, la justicia de datos y la colaboración creativa humano–**IA**. Este desequilibrio genera tensiones entre la adopción de tecnologías inteligentes y la misión formativa de la educación superior orientada al desarrollo de ciudadanos críticos y socialmente responsables.

Propósito. Sustentar cualitativamente el **modelo CAITIZEN** en un marco multidisciplinario de la ciudadanía asistida por **IA** en proceso de formación y con el Manual de Oslo de la OCDE, posicionarlo como una innovación social de cinco variables: Alfabetización Crítica en **IA (CAIL)**; Conciencia Ética y Responsabilidad (**EAR**); Equidad y Justicia de los Datos (**AFDJ**); Colaboración Creativa Humano–**IA (HAIC)**; y Transparencia Metacognitiva en Prompts (**MTPP**) en **SDGs**.

Metodología. Se empleó un diseño de investigación cualitativo entre **julio y diciembre de 2025** en Jalisco, México, con la participación de **511 estudiantes de licenciatura y posgrado**. Los datos se recolectaron mediante un cuestionario estructurado y se analizaron mediante análisis temático con apoyo de **ATLAS.ti 25**.

Hallazgos teóricos y prácticos. El estudio reconceptualiza el uso de la **IA** como un sistema ético–cognitivo–social, aportando avances teóricos en alfabetización y orientaciones prácticas para el diseño curricular, el uso responsable y la supervisión institucional.

Originalidad. Vincular la innovación educativa con los **ODS** mediante la integración de la alfabetización en IA, la ética, la equidad, la claridad en los prompts y la colaboración humano–**IA** como pilares de una ciudadanía sostenible.

Conclusiones. La **IA** en la educación superior se configura como un sistema ético–cognitivo–social mediante el **modelo CAITIZEN**, cuyo enfoque cualitativo y contextual ofrece potencial para futuras investigaciones comparativas

1. INTRODUCTION

Within contemporary digital ecosystems, the accelerated integration of artificial intelligence (**AI**) is reshaping how individuals create, learn, communicate, and make decisions. This transition from conventional digital media to interaction with intelligent and adaptive systems requires competencies that extend beyond technical or instrumental digital literacy. Citizens increasingly need to understand how **AI** systems operate, recognize their limitations, identify algorithmic bias, interpret automated outputs, and act with ethical awareness and critical reflection in technologically mediated environments. However, dominant approaches to digital literacy and education continue to prioritize efficiency, functionality, and skill acquisition, offering limited attention to ethical judgment, reflexivity, and creative agency. This limitation poses significant challenges for societies seeking to responsibly integrate **AI** technologies that influence civic participation, knowledge production, creativity, and democratic life.

According to the Oslo Manual (OCDE & Eurostat, 2018), innovation is understood not only as technological advancement but also as the introduction of new conceptual frameworks and social practices that generate value and transform behavior. Based on this perspective, the absence of integrative models addressing **AI-assisted citizenship** reveals a gap in educational and social innovation. Addressing this gap requires frameworks capable of articulating creative, critical, and ethical dimensions of **human–AI interaction** in formative contexts.

This research proposes the **CAITIZEN model** as a multidimensional framework designed to identify **five essential competencies** guiding the development of **AI-assisted citizenship** among university students in a qualitative sense. These competencies include critical **AI** literacy, ethical awareness and responsibility, awareness of fairness and data justice, human–**AI** creative collaboration, and metacognitive transparency in prompting practices. These dimensions together constitute the **CAITIZEN profile,** understood not as a fully achieved state, but as a formative trajectory reflecting citizenship in formation.

Framed within innovation for sustainable development, this research adopts a qualitative methodology supported by **ATLAS.ti 25** to analyze categories, word clouds, co-occurrences, conceptual network, and sentiment analysis based on student-centered meanings in **AI-mediated educational contexts critically**.

129

2. CONTEXT

The increasing use of artificial intelligence (**AI**) has transformed worldwide, international, and national systems of governance, production, and, particularly, the education sector. Across these levels, multiple organizations emphasize that societies must not only adapt to **AI** technologies but also cultivate ethical, critical, and creative capacities to ensure inclusive and sustainable innovation. In line with the Oslo Manual, innovation is understood not only as technological advancement but also as the introduction of new conceptual frameworks and social practices that transform behavior and meaning (OECD & Eurostat, 2018). From this perspective, the need for integrative approaches to **AI-assisted citizenship** becomes evident. The following subsections outline this contextual evolution and justify the relevance of frameworks such as the **CAITIZEN model**.

2.1. Global Level

At the global level, the high-speed integration of **AI** into society and education has established the urgency to redefine digital literacy as a broader competence encompassing critical reflection, social responsibility, and ethics. For instance, **UNESCO's AI** Competency Framework for Teachers emphasizes that professionals and learners must develop **AI-specific**, human-centered values, knowledge, and ethical principles in order to engage responsibly with intelligent systems (Miao & Cukurova, 2024). However, despite advances in digital infrastructure, a significant gap persists between technical training and citizens' ability to participate critically in **AI-mediated** environments (OECD, 2025). The World Economic Forum (WEF, 2025) has also identified **AI** literacy as a key sustainability skill for the future of work and innovation. Collectively, these global reports illustrate that technological progress must be accompanied by reflective and ethical frameworks to ensure human well-being.

2.2. International Level

At the international level, collaborative efforts between the OECD and the European Commission produced the **AI** Literacy Framework for Primary and Secondary Education, which defines competencies for understanding, creating, and managing **AI** systems in socially responsible ways (OECD & European Commission, 2025). These initiatives reflect a paradigm shift toward

global cooperation that connects technical fluency with ethical awareness, transparency, and algorithmic fairness. Countries adopting such integrative frameworks demonstrate stronger resilience and innovation capacity, as ethical and critical citizens become essential to equitable digital transformation.

2.3. National Level (México)

In Mexico, technological adoption has grown significantly, yet gaps remain in higher-order digital and ethical competencies. The National Survey on Availability and Use of Information Technologies in Households (INEGI, 2023) reported that 81% of the population uses the internet daily, though understanding of algorithmic systems and data governance remains limited. This gap is particularly evident within higher education contexts, where future citizens are expected to develop ethical, critical, and reflective competencies for engaging responsibly with **AI** technologies (Mejía-Trejo, 2025).

The collaboration between **CANIETI** and **UNESCO** on the *Modelo México* for ethical and responsible artificial intelligence illustrates Mexico's commitment to promoting trustworthy **AI** ecosystems (UNESCO & CANIETI, 2025). This initiative demonstrates institutional awareness of the ethical, social, and educational challenges posed by **AI** adoption. However, while it focuses on organizational practices, the **CAITIZEN model** extends this vision to the civic sphere, proposing a framework that develops critical, ethical, and creative competencies for citizens to engage responsibly with **AI** technologies.

Given this global, international, and national panorama, the **CAITIZEN model** emerges as a necessary response to guide the ethical, cognitive, and creative development of citizens. It provides a structured foundation for cultivating critical, ethical, and creative competencies for citizens, particularly students in formative educational stages, enabling responsible participation in **AI-driven societies**.

3. LITERATURE REVIEW

This section presents the state of the art concerning the main variables that shaped the **CAITIZEN model**, based on academic literature published primarily between 2020 and 2025. The review is organized around five analytical dimensions—critical artificial intelligence literacy

(**CAIL**)**,** ethical awareness and responsibility (**EAR**)**,** awareness of fairness and data justice (**AFDJ**)**,** human–**AI** creative collaboration (**HAIC**)**,** and metacognitive transparency in prompting practices (**MTPP**)**—**each grounded in a specific body of scholarly work and explicitly aligned with selected United Nations Sustainable Development Goals (UN,n.d.) . Together, these dimensions frame **AI-assisted citizenship** as a competence-based construct that integrates critical understanding, ethical judgment, equity awareness, creative collaboration, and reflective self-regulation. In this study, **AI-assisted citizenship** is understood as a condition in formation, as the subject of analysis consists of university students who are still developing these competencies within higher education contexts.

The selected studies were chosen for their direct conceptual and empirical contributions to defining citizen-level competencies for responsible engagement with artificial intelligence. This literature collectively supports the alignment of the **CAITIZEN model** with sustainable development priorities related to education quality, institutional integrity, equity, innovation, and responsible technological use.

Critical Artificial Intelligence Literacy (CAIL) refers to the ability to understand how **AI** systems function, including their technical foundations, inherent biases, limitations, and broader societal implications. Long and Magerko (2020) conceptualized **AI** literacy as a critical competency that extends beyond technical skills to include conceptual understanding and reflection. Ng et al. (2021) further defined **AI** literacy as a set of interpretive and reflective skills enabling individuals to comprehend algorithmic mechanisms and evaluate **AI** outputs. Southworth et al. (2023) emphasized embedding **AI** literacy across educational curricula to deepen understanding of **AI**'s social and cognitive impacts. Wang and Wang (2025) provided empirical evidence showing that **AI-assisted** writing tools promote metacognitive reflection and critical evaluation of algorithmic outcomes. Collectively, these studies position **CAIL** as foundational for informed participation in **AI-mediated** societies, contributing to **SDG4** (Quality Education) and **SDG16** (Peace, Justice and Strong Institutions)**.**

Ethical Awareness and Responsibility (EAR) focuses on sensitivity to moral dilemmas arising from **AI** use and commitment to ethical decision-making, governance, and accountability. Gunasekara et al. (2025) identified moral sensitivity, accountability, and human oversight as core principles of responsible **AI**. Kong and Zhu (2025) validated an **AI** ethical awareness scale,

demonstrating that ethical reflection can be cultivated through **AI-based problem-solving**. Papagiannidis et al. (2025) emphasized ethical decision-making structures in **AI** governance, while Stetson et al. (2025) highlighted transparency and human agency in healthcare **AI** applications. Together, these contributions establish **EAR** as central to **AI** citizenship, aligned with **SDG16** (Peace, Justice and Strong Institutions) and **SDG12** (Responsible Consumption and Production)**.**

Awareness of Fairness and Data Justice (AFDJ) addresses the social, legal, and ethical impacts of data-driven and automated decision-making. Decker et al. (2025) emphasized procedural fairness and public engagement as prerequisites for legitimate algorithmic systems. Demirchyan (2025) examined regulatory challenges related to fairness in **AI**. González-Argote et al. (2025) identified ethical tensions arising from biased data and discriminatory outcomes, while Pham et al. (2025) demonstrated the importance of fairness for equitable educational outcomes. These studies collectively frame **AFDJ** as a mechanism for addressing inequality and strengthening social trust, contributing to **SDG 10**(Reduced Inequalities) and **SDG16** (Peace, Justice and Strong Institutions)**.**

Human–AI Creative Collaboration (HAIC) refers to the use of **AI** as a collaborative partner in creative processes and innovation. Georgieva and Georgiev (2025) showed how generative **AI** supports creative ideation and knowledge generation. McGuire et al. (2024) emphasized **AI**'s role in creative problem-solving, while Rafner et al. (2025) analyzed cognitive dynamics in co-creative writing. Wang et al. (2025) demonstrated that **AI** enhances creativity when guided by reflective strategies. These studies position **HAIC** as a driver of innovation and collaboration, aligned with **SDG9** (Industry, Innovation and Infrastructure)**, SDG8** (Decent Work and Economic Growth)**,** and **SDG17** (Partnerships for the Goals)**.**

Metacognitive Transparency in Prompting Practices (MTPP) emphasizes reflective awareness in human–**AI** interaction, particularly in how prompts are formulated and refined. Haidar et al. (2025) demonstrated that metacognitive prompts improve reflective awareness. Lee and Palmer (2025) highlighted the role of metacognition in intentional **AI** use. Tsakeni et al. (2025) showed that **AI** tools scaffold metacognitive learning, while Waaler et al. (2025) demonstrated that prompt engineering enhances ethical compliance. Together, these findings establish **MTPP** as essential for reflective and responsible **AI** use, supporting **SDG4** (Quality Education) and **SDG12** (Responsible Consumption and Production)**.**

133

Taken together, the five dimensions of the **CAITIZEN Model**, each explicitly grounded in its corresponding authors and **SDGs**, form an integrated framework for ethical, critical, creative, and reflective citizenship in formation, particularly within higher education contexts where future citizens engage with **AI-mediated societies**.

3.1. Design of the Qualitative Data Collection Instrument

The qualitative instrument was designed as a mixed-format questionnaire to operationalize the five variables derived from the literature within the **CAITIZEN model**. It combined structured Likert-type items with open-ended questions to contextualize the subjects of study and elicit reflective narratives. This design responds to the need to capture how citizenship in formation is constructed among university students through reflective, ethical, and critical engagement with artificial intelligence.

The questionnaire was administered via Google Forms between July and December 2025 to 511 undergraduate and graduate university students from the state of Jalisco, Mexico**,** identified as artificial intelligence users in training. The instrument comprised 50 closed-ended items distributed across the five **CAITIZEN variables—CAIL, EAR, AFDJ, HAIC, and MTPP**—measured using a Likert-type scale to sensitize participants to each conceptual dimension.

Additionally, one general open-ended question was included at the end of each variable block, totaling **55 questions**. Instrument construction followed a deductive–inductive logic, ensuring theoretical grounding while allowing emergent meanings from participants' qualitative responses. See **Table 1.**

Table 1. CAITZEN Variables and their SDG contribution

| Variable (meaning)	Authors	Qualitative guiding questions	SDGs contribution
1. Critical Artificial Intelligence Literacy (CAIL). The ability to understand how **AI** systems function, including their technical foundations, inherent biases, limitations, and broader educational and societal implications, fostering critical and reflective engagement with AI.	Long & Magerko (2020); Ng et al. (2021); Southworth et al. (2023); Wang & Wang (2025)	What personal or academic experiences have helped you better understand how artificial intelligence influences your learning or work?	**SDG 4 (Quality Education); SDG 16 (Peace, Justice and Strong Institutions).** Promotes critical thinking, informed decision-making, and accountable participation in **AI-mediated societies.**
2. Ethical Awareness and Responsibility (EAR). Sensitivity to moral dilemmas arising from **AI** use and	Gunasekara et al. (2025); Kong & Zhu (2025);	Why do you consider it important to act	**SDG 16 (Peace, Justice and Strong Institutions); SDG 12 (Responsible**

134

Variable (meaning)	Authors	Qualitative guiding questions	SDGs contribution
a commitment to responsible practices, ethical decision-making, governance, and accountability in educational and professional contexts.	Papagiannidis et al. (2025); Stetson et al. (2025)	ethically and responsibly when using artificial intelligence?	**Consumption and Production).** Strengthens ethical governance, responsibility, and sustainable use of **AI** technologies.
3. Awareness of Fairness and Data Justice (AFDJ). Recognition of the social, legal, and ethical impacts of data-driven and automated decisions, including algorithmic bias, discrimination, procedural fairness, and data justice.	Decker et al. (2025); Demirchyan (2025); González-Argote et al. (2025); Pham et al. (2025)	What situations make you think that **AI** systems may be unfair or favor certain groups?	**SDG 10 (Reduced Inequalities); SDG 16 (Peace, Justice and Strong Institutions).** Addresses bias, discrimination, and equity in algorithmic decision-making.
4. Human–AI Creative Collaboration (HAIC). The use of **AI** as a collaborative partner in creative processes, supporting idea generation, innovation, co-creation, and knowledge production across educational and design-oriented contexts.	Georgieva & Georgiev (2025); McGuire et al. (2024); Rafner et al. (2025); Wang et al. (2025)	How have you used artificial intelligence as an ally in developing creative ideas, problem-solving, or innovation?	**SDG 9 (Industry, Innovation and Infrastructure); SDG 8 (Decent Work and Economic Growth); SDG 17 (Partnerships for the Goals).** Fosters innovation, creative productivity, and collaborative human–**AI** ecosystems.
5. Metacognitive Transparency in Prompting Practices (MTPP). Reflective awareness of how prompts are formulated, adapted, and evaluated when interacting with **AI** systems, emphasizing metacognition, self-regulation, and intentional use of prompting strategies.	Haidar et al. (2025); Lee & Palmer (2025); Tsakeni et al. (2025); Waaler et al. (2025)	When interacting with **AI** tools, how do you reflect on and adjust your prompts to obtain better or more ethical responses?	**SDG 4 (Quality Education); SDG 12 (Responsible Consumption and Production).** Promotes reflective learning and responsible, transparent interaction with **AI** systems.

Note. Authors' own elaboration based on multiple sources.

3.2. Conceptual model design

The conceptual model is designed *ex ante* as an analytical framework to conceptually organize and interpret the dimensions that constitute **CAITIZEN** model in human–**AI** interaction contexts. The model proposes **CAIL, EAR, AFDJ, HAIC, and MTPP** as core dimensions that conceptually converge on the **CAITIZEN** construct, guiding qualitative inquiry and interpretation. Rather than testing causal relationships, the model serves as a heuristic structure to support meaning-making and thematic exploration of critical and ethical **AI** citizenship. See **Figure 1.**

Figure 1. CAITIZEN model
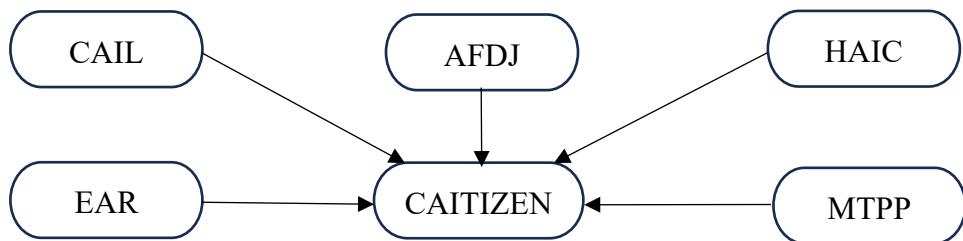
Notes: **CAIL**. *Critical Artificial Intelligence Literacy;* **EAR.** *Ethical Awareness and Responsibility*; **AFDJ**. *Awareness of Fairness and Data Justice*; **HAIC.** *Human–AI Creative Collaboration*; **MTPP**. *Metacognitive Transparency in Prompting Practices.*

4. METHODOLOGY

The study was conducted using a qualitative approach based on advanced thematic analysis, establishing a direct relationship between the subjects of analysis and the object of study, with a focus on the conceptual construction of the **CAITIZEN model** for citizenship in formation among university students in human–artificial intelligence interaction contexts. The methodological procedure was organized into clearly defined and sequential stages.

In the first stage, a comprehensive review of the scientific literature was carried out to identify and theoretically substantiate the five core variables of the model: **CAIL, EAR, AFDJ, HAIC,** and **MTPP**. This stage provided the conceptual basis for the ex ante design of the conceptual model.

In the second stage, the qualitative corpus was established through a documentary census, in which all academic documents meeting predefined inclusion criteria—namely thematic relevance, scientific rigor, and currency—were analyzed in full. No probabilistic sampling was applied, and complete textual documents constituted the primary qualitative material for analysis.

In the third stage, a previously designed mixed-format questionnaire was administered to the study participants as a contextual and elicitation mechanism aligned with the five model variables, generating additional qualitative material for analysis.

In the fourth stage, a non-structured qualitative thematic analysis was conducted using **ATLAS.ti version 25**, involving inductive coding, category construction, word cloud generation

to detect dominant lexical patterns, and code co-occurrence analysis to examine conceptual relationships.

In the fifth stage a conceptual network was constructed to visualize and interpret relationships among categories and model dimensions.

Finally, sentiment analysis was applied as a complementary interpretive strategy to identify the evaluative orientation of the analyzed discourse.

5. RESULTS

The first results derived from the advanced thematic analysis are presented below. The analysis yielded a structured set of categories and associated codes, reflecting both the conceptual organization of the **CAITIZEN** Model and the level of thematic saturation achieved among university students as citizens in formation. These results provide empirical grounding for the model's analytical dimensions. See **Table 2.**

Table 2. Categories, Codes, and Code Frequency Identified through ATLAS.ti

Category	Codes	Number of codes
CAIL	Critical analysis; Accelerated learning; Cognitive facilitation; Comprehension strategies; Idea structuring; Learning improvement; Task organization; Critical thinking; Multiple perspectives; Metacognitive reflection; Constructive feedback; Critical and responsible information use	**12**
CAITIZEN	Digital citizenship and ethical training; **AI** collaboration; Critical skills	**3**
HAIC	Participatory capacities; Integration of opinions; Creative practices; Teamwork	**4**
AFDJ	Social, cognitive, and cultural impact of AI; Bias, algorithmic influence, and transparency; Intelligent learning assistance	**3**
EARB	Ethics in **AI** use; Ethics and regulation of **AI** use	**2**
MTPP	Cognitive evaluation and performance; Prompting interaction and optimization; Resource optimization; Review of recommendations; Decision-making; Responsible AI use; Critical and responsible information use	**7**
Total		**30**

Source: Authors' own elaboration using ATLAS.ti 25

The second result is presented in Figure 2, which displays a word cloud of dominant lexical patterns. The discourse is primarily shaped by references to higher education, digital learning, academic credentials, and institutional contexts. Artificial intelligence emerges as a normalized and instrumentalized tool, embedded within narratives structured by social context, demographic markers, and ethical responsibility. This pattern indicates that **AI** is perceived less as a disruptive novelty and more as an integrated component of contemporary educational practices.

137

Figure 2. Word cloud of the citizen model.
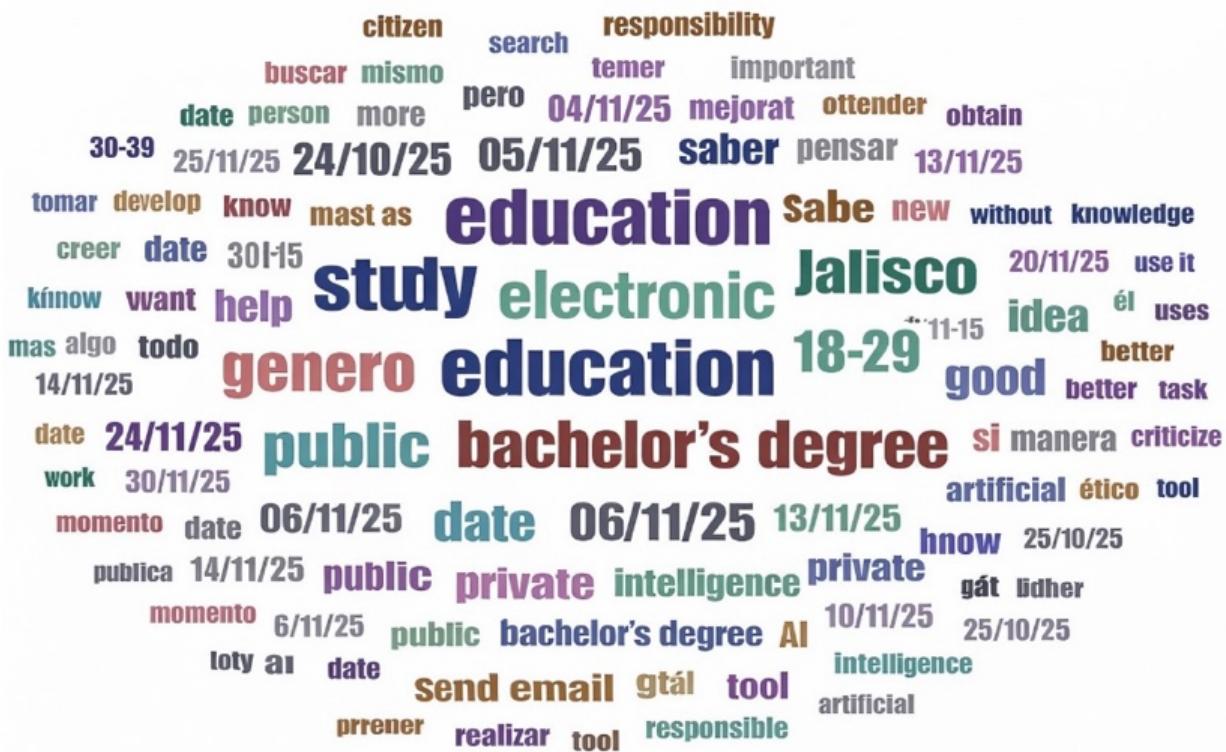
Source: Authors' own elaboration using ATLAS.ti 25.

The third result corresponds to the co-occurrence analysis, which reveals that ethics and regulation of **AI** use functions as the central axis of the discourse. This category shows the strongest associations with critical thinking, **AI** collaboration, constructive feedback, and metacognitive reflection. Such co-occurrence patterns suggest that the adoption of **AI** in educational contexts is primarily framed through ethical considerations and closely linked to the development of critical, participatory, and decision-making competencies, reinforcing a socially responsible integration of **AI technologies. See Table 3.**

Table 3. The co-occurrence analysis

#	Category A	Category B	Frequency
1	**Ethics and Regulation of AI Use**	Critical thinking	**433**
2	**Ethics and Regulation of AI Use**	**AI** collaboration	**413**
3	**Ethics and Regulation of AI Use**	Constructive feedback	**329**
4	**Ethics and Regulation of AI Use**	Metacognitive reflection	**302**
5	**Ethics and Regulation of AI Use**	Critical skills	**269**
6	AI collaboration	Constructive feedback	229
7	**Ethics and Regulation of AI Use**	Critical analysis	**222**

138

#	Category A	Category B	Frequency
8	**Ethics and Regulation of AI Use**	**Creative practices**	**189**
9	AI collaboration	Metacognitive reflection	187
10	Metacognitive reflection	Constructive feedback	183
11	Critical thinking	Critical analysis	175
12	Critical thinking	Critical skills	170
13	**Ethics and Regulation of AI Use**	**Participatory capacities**	**136**
14	AI collaboration	Creative practices	129
15	Critical analysis	Critical thinking	107
16	Critical thinking	Comprehension strategies	91
17	Critical thinking	Participatory capacities	75
18	Responsible use of AI	Critical thinking	75
19	Critical skills	Critical analysis	71
20	Constructive feedback	Critical skills	57
21	Critical skills	**AI** collaboration	57
22	Metacognitive reflection	Critical skills	55
23	**Ethics and Regulation of AI Use**	**Decision-making**	**43**
24	Intelligent learning assistance	**Ethics and Regulation of AI Use**	42
25	Comprehension strategies	Critical thinking	42
26	Decision-making	**Ethics and Regulation of AI Use**	39
27	**Ethics and Regulation of AI Use**	**Responsible use of AI**	**31**
28	Critical and responsible use of **AI**-generated information	**Ethics and Regulation of AI Use**	29
29	Responsible use of AI	**Ethics and Regulation of AI Use**	27
30	**Ethics and Regulation of AI Use**	Responsible use of AI	**27**

Source: Authors' own elaboration using ATLAS.ti 25.

As a fourth result, the **CAITIZEN** conceptual network is presented in **Table 4**. This network illustrates the directional and functional relationships among the five core dimensions of the model.

Table 4. Description of the CAITIZEN conceptual network at the level of the main model variables

Source node (→)	Arrow direction	Target node	Type of conceptual relationship	Theoretical justification
CAIL →	CAIL → EAR	EAR	Grounds	Critical **AI** literacy provides the cognitive and ethical foundations required to recognize moral implications, risks, and responsibilities in the use of **AI** systems, enabling ethical awareness as a reflective capacity (Ng et al., 2021; Kong & Zhu, 2025; Floridi et al., as synthesized in Gunasekara et al., 2025).
CAIL →	CAIL → AFDJ	AFDJ	Enables	Understanding how **AI** systems operate equips individuals to identify algorithmic bias, data injustice, and fairness challenges, thereby enabling informed engagement with issues of algorithmic fairness and data justice (Gonzalez-Argote et al., 2025; Pham et al., 2025; Demirchyan, 2025).

Source node (→)	Arrow direction	Target node	Type of conceptual relationship	Theoretical justification
CAIL →	CAIL → MTPP	MTPP	Sustains	**AI** literacy supports metacognitive prompting practices by fostering conscious formulation, evaluation, and adjustment of prompts, which enhances reflective and strategic interaction with generative **AI** systems (Haidar et al., 2025; Tsakeni et al., 2025; Waaler et al., 2025).
EAR →	EAR → AFDJ	AFDJ	Orients	Ethical awareness directs attention toward fairness, transparency, and social justice in algorithmic systems, framing data practices within broader concerns of moral responsibility and public accountability (Decker et al., 2025; Gonzalez-Argote et al., 2025).
EAR →	EAR → HAIC	HAIC	Regulates	Ethical awareness regulates human–**AI** interaction by establishing normative boundaries that guide responsible collaboration and prevent socially harmful or misaligned uses of **AI** (Papagiannidis et al., 2025; UNESCO & CANIETI, 2025; Stetson et al., 2025).
AFDJ →	AFDJ → HAIC	HAIC	Conditions	Human–**AI** collaboration is conditioned by the fairness, quality, and governance of data and algorithms, as inequitable or biased systems directly shape interaction outcomes and creative agency (Demirchyan, 2025; Pham et al., 2025; Decker et al., 2025).
MTPP →	MTPP → HAIC	HAIC	Mediates	Metacognitive prompting mediates human–**AI** collaboration by clarifying goals, constraints, and evaluation criteria, thereby improving the quality and intentionality of joint human–**AI** activity (Haidar et al., 2025; Tsakeni et al., 2025).
CAIL →	CAIL → CAITIZEN	CAITIZEN	Contributes to	**AI** literacy constitutes a structural pillar of the **CAITIZEN** profile, enabling informed, critical, and reflective engagement with **AI** across educational and social contexts (Ng et al., 2021; Southworth et al., 2023; OECD & European Commission, 2025).
EAR →	EAR → CAITIZEN	CAITIZEN	Contributes to	Ethical responsibility shapes the moral dimension of **AI**-assisted citizenship by orienting decisions and practices toward accountability, inclusion, and the common good (Papagiannidis et al., 2025; United Nations, n.d.).
AFDJ →	AFDJ → CAITIZEN	CAITIZEN	Contributes to	Algorithmic fairness and data justice are necessary conditions for inclusive and socially legitimate digital citizenship, ensuring equitable participation and trust in AI-mediated systems (Gonzalez-Argote et al., 2025; Pham et al., 2025).
HAIC →	HAIC → CAITIZEN	CAITIZEN	Contributes to	Human–**AI** interaction expresses the creative and participatory dimension of the **AI-assisted**

Source node (→)	Arrow direction	Target node	Type of conceptual relationship	Theoretical justification
				citizen, integrating human agency with algorithmic capabilities in co-creative and decision-making processes (Georgieva & Georgiev, 2025; Rafner et al., 2025; Salma et al., 2025).
MTPP →	MTPP → CAITIZEN	CAITIZEN	Contributes to	Metacognitive regulation of **AI** use consolidates an autonomous and reflective form of digital citizenship, capable of consciously governing interaction with intelligent systems (Haidar et al., 2025; Tsakeni et al., 2025; WEF, 2025).

Source: Authors' own elaboration using ATLAS.ti 25.

The **Table 4** describes a coherent conceptual network in which critical **AI** literacy **(CAIL)** operates as a foundational driver of the **CAITIZEN** model. **CAIL** grounds ethical awareness (**EAR**), enables sensitivity to algorithmic fairness and data justice (**AFDJ**), and sustains metacognitive prompting practices (**MTPP**), establishing the cognitive and reflexive conditions necessary for responsible interaction with artificial intelligence.

Ethical awareness (**EAR**) plays a regulatory and orienting role within the system. It directs concerns toward fairness and social justice in algorithmic decision-making and regulates human–**AI** collaboration (**HAIC**) by embedding technological practices within ethical frameworks. In parallel, algorithmic fairness and data justice (**AFDJ**) condition human–**AI** collaboration by ensuring that joint creative processes are based on equitable and responsibly used data.

Metacognitive prompting practices (**MTPP**) mediate the quality of human–**AI** interaction by making goals, criteria, and limits explicit, thereby shaping how collaboration with intelligent systems is enacted. Together, these relationships form a dynamic structure in which cognition, ethics, data justice, and interaction mutually reinforce one another.

At the integrative level, all variables converge in **CAITIZEN**, which represents the profile of the **AI-assisted citizen**. Critical literacy, ethical responsibility, data justice, collaborative capacity, and metacognitive regulation each contribute to the formation of a reflective, autonomous, and socially responsible digital citizenship. The model thus conceptualizes **CAITIZEN** as the outcome of an interdependent system where **AI** use is not merely technical, but ethically grounded, cognitively informed, and socially oriented.

As complementary result, **Table 5** presents the conceptual network of the **CAITIZEN model** at the level of its main categories, detailing the relationships among ethical, cognitive, competency-based, relational, and instrumental dimensions of **AI-assisted education**. The table specifies the direction and nature of these conceptual links, illustrating how ethical and regulatory frameworks structure critical thinking and responsible **AI** use, how cognitive and metacognitive processes guide decision-making, and how human–**AI** collaboration supports creativity, participation, and learning outcomes. Together, these relationships articulate the systemic logic through which the **CAITIZEN** model integrates ethics, cognition, collaboration, and educational practice into a coherent framework for responsible and reflective engagement with artificial intelligence.

Table 5. Description of the CAITIZEN conceptual network at the level of the main model categories

Source node	Conceptual role	Connected node	Link label	Conceptual direction	Theoretical justification
Ethics and Regulation of AI Use	Framework concept	Critical thinking	Sustains	Normative → critical	Ethical and regulatory frameworks provide normative standards that sustain critical thinking by enabling reflective evaluation of **AI** systems, their impacts, risks, and limitations, particularly in educational and social contexts (Decker et al., 2025; Demirchyan, 2025; Gunasekara et al., 2025).
Ethics and Regulation of AI Use	Framework concept	Responsible use of AI	Defines	Normative → applied	Responsible **AI** use represents the applied enactment of ethical and regulatory principles, translating governance frameworks into concrete practices of accountability, transparency, and risk mitigation (Papagiannidis et al., 2025; Stetson et al., 2025; UNESCO & CANIETI, 2025).
Critical thinking	Cognitive concept	Critical analysis	Is concretized in	Cognitive → evaluative	Critical thinking becomes operational through critical analysis processes that evaluate information quality, algorithmic outputs, and **AI-mediated** decisions, which are central to **AI** literacy and ethical awareness (Ng et al., 2021; Wang & Wang, 2025; Kong & Zhu, 2025).
Critical thinking	Cognitive concept	Comprehension strategies	Orients	Cognitive → strategic	Critical thinking orients comprehension strategies that allow individuals to interpret complex, data-driven, and **AI-generated** information environments in an informed and

142

Source node	Conceptual role	Connected node	Link label	Conceptual direction	Theoretical justification
					reflective manner (OECD & European Commission, 2025; Xiao et al., 2024).
Critical analysis	Evaluative concept	Constructive feedback	Grounds	Evaluative → formative	Critical analysis provides evaluative criteria that ground constructive feedback, supporting formative learning processes in **AI-supported** educational settings (Tsakeni et al., 2025; Haidar et al., 2025).
Constructive feedback	Formative concept	Metacognitive reflection	Stimulates	Formative → metacognitive	Constructive feedback stimulates metacognitive reflection by promoting awareness of cognitive strategies, learning regulation, and interaction patterns with intelligent systems (Haidar et al., 2025; Tsakeni et al., 2025).
Metacognitive reflection	Self-regulatory concept	Decision-making	Informs	Metacognitive → decisional	Metacognitive reflection informs decision-making by enabling deliberate, justified, and autonomous judgments regarding the use, reliance, and limits of **AI** systems (Waaler et al., 2025; WEF, 2025).
Critical skills	Competency-based concept	Critical thinking	Integrates	Competency → cognitive	Critical skills integrate and operationalize critical thinking as a set of applied competencies required for navigating complex **AI-mediated** educational and professional environments (Miao & Cukurova, 2024; OECD, 2025).
Critical skills	Competency-based concept	Responsible use of AI	Enables	Competency → normative	The development of critical skills enables responsible AI use by supporting informed judgment, ethical awareness, and regulatory compliance in **AI-assisted** practices (Gunasekara et al., 2025; Papagiannidis et al., 2025).
AI collaboration	Relational concept	Creative practices	Enhances	Relational → creative	Human–**AI** collaboration enhances creative practices by expanding idea generation, recombination, and exploratory processes in co-creative tasks supported by generative **AI** (Georgieva & Georgiev, 2025; Rafner et al., 2025; Wang et al., 2025).
Creative practices	Creative–applied concept	Accelerated learning	Contributes to	Creative → outcome	Creative practices contribute to accelerated learning by fostering deeper engagement, experimentation, and adaptive knowledge construction in **AI-supported educational contexts** (Córdova-Esparza, 2025; Southworth et al., 2023).

Source node	Conceptual role	Connected node	Link label	Conceptual direction	Theoretical justification
AI collaboration	Relational concept	Participatory capacities	Fosters	Relational → social	Collaborative interaction with **AI** fosters participatory capacities by supporting shared agency, collective problem-solving, and inclusive knowledge construction (Salma et al., 2025; Rafner et al., 2025).
Participatory capacities	Socio-civic concept	Decision-making	Influences	Participatory → decisional	Participatory capacities influence decision-making by strengthening deliberative, informed, and socially responsible choices in **AI-mediated** environments (Decker et al., 2025; United Nations, n.d.).
Comprehension strategies	Strategic–cognitive concept	Intelligent learning assistance	Optimizes	Strategic → instrumental	Strategic comprehension optimizes intelligent learning assistance by enabling its use as adaptive cognitive support rather than as a substitute for human reasoning (Córdova-Esparza, 2025; Tsakeni et al., 2025).
Intelligent learning assistance	Instrumental–educational concept	Accelerated learning	Supports	Instrumental → outcome	Intelligent learning assistance supports accelerated learning when integrated under ethical governance, pedagogical guidance, and metacognitive regulation (Córdova-Esparza, 2025; OECD, 2025).

Source: Authors' own elaboration using ATLAS.ti 25.

Table 6 shows the sentiment analysis of the main categories of the **CAITIZEN model,** integrating network metrics, sentiment distribution, and evaluative intensity to capture how each category is affectively and discursively framed within the corpus. By integrating category density, role in the network, predominant sentiment, and evaluative intensity, the table provides a concise overview of the emotional orientation and analytical relevance of the ethical, cognitive, collaborative, and instrumental dimensions of **AI-assisted education.** This approach reveals both sentiment prevalence and intensity, offering insight into the legitimacy, acceptance, and functional positioning of the **CAITIZEN** model components within the discourse.

Table 6. Sentiment analysis results of the CAITIZEN model.

Category	Density	Role in the network	Total cases (n)	Sentiment context of the corpus	Predominant sentiment	Evaluative intensity	Integrated analytical interpretation
Ethics and Regulation of AI Use	2,793	Central node	511	Positive 71.04% · Neutral 28.38% ·	Predominantly positive	High	Core axis of the **CAITIZEN** model, reflecting ethical legitimacy

Category	Density	Role in the network	Total cases (n)	Sentiment context of the corpus	Predominant sentiment	Evaluative intensity	Integrated analytical interpretation
				Negative 0.59%			and regulatory consensus consistent with responsible **AI** governance and fairness frameworks (Decker et al., 2025; Demirchyan, 2025).
AI Collaboration	1,029	Bridging node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Predominantly positive	Medium–High	Framed through complementarity and shared agency, aligning with human–**AI co-creation** research emphasizing augmentation over replacement (Georgieva & Georgiev, 2025; Rafner et al., 2025).
Critical Thinking	870	Central node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Predominantly positive	Medium	Cognitive core with moderate intensity, consistent with **AI** literacy frameworks focused on rational evaluation of **AI** outputs (Ng et al., 2021; Kong & Zhu, 2025).
Constructive Feedback	814	Bridging node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Positive–Neutral	Medium	Reflects formative, improvement-oriented processes embedded in pedagogical discourse rather than affective engagement (Haidar et al., 2025).
Metacognitive Reflection	804	Bridging node	511	Positive 71.04% ·	Positive–Neutral	Medium	Characterized as a deliberate self-

Category	Density	Role in the network	Total cases (n)	Sentiment context of the corpus	Predominant sentiment	Evaluative intensity	Integrated analytical interpretation
				Neutral 28.38% · Negative 0.59%			regulatory process with low polarization, aligned with reflective **AI** engagement studies (Tsakeni et al., 2025).
Critical Skills	791	Central node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Predominantly positive	Medium	Transversal competencies positively valued in education and positioned as foundational for responsible **AI** engagement (Miao & Cukurova, 2024; OECD, 2025).
Critical Analysis	681	Central node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Positive–Neutral	Medium	High structural relevance with a rational sentiment profile, reinforcing its evaluative role in assessing **AI-generated information** (Pham et al., 2025).
Creative Practices	340	Peripheral node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Predominantly positive	Low–Medium	Positively perceived but context-dependent, reflecting innovation in human–**AI** interaction with lower centrality (Georgieva & Georgiev, 2025).
Participatory Capacities	290	Peripheral node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Predominantly positive	Low–Medium	Valued yet less explicit in discourse, consistent with governance research on participation in **AI** adoption

Category	Density	Role in the network	Total cases (n)	Sentiment context of the corpus	Predominant sentiment	Evaluative intensity	Integrated analytical interpretation
							(Decker et al., 2025).
Decision-Making	226	Peripheral node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Positive–Neutral	Low	Treated as a functional and strategic competence, aligned with applied **AI** governance models (OECD, 2025).
Responsible Use of AI	224	Peripheral node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Positive–Neutral	Low	Normative and prudential orientation emphasizing accountability and regulation over affective discourse (Gunasekara et al., 2025).
Intelligent Learning Assistance	200	Peripheral node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Predominantly positive	Low	Shows acceptance of **AI-supported learning** under ethical governance, with minimal discursive conflict (Córdova-Esparza, 2025).
Comprehension Strategies	187	Peripheral node	511	Positive 71.04% · Neutral 28.38% · Negative 0.59%	Positive–Neutral	Low	Enabling cognitive mechanisms embedded in favorable but low-salience discourse (OECD & European Commission, 2025).

Source: Authors' own elaboration using ATLAS.ti 25.

6. DISCUSSION

The results offer a robust, multi-layer qualitative validation of the **CAITIZEN model** by triangulating five complementary analytical layers particularly within higher education contexts, where **AI-assisted citizenship** is still in formation.

First, **the categorical analysis** derived from advanced thematic coding reveals a highly structured conceptual architecture with strong thematic saturation. The prominence of categories related to ethics and regulation of **AI** use**,** critical ai literacy**, and** metacognitive prompting practices indicates that **AI** is primarily framed as an ethically mediated and cognitively regulated phenomenon, consistent with **AI** literacy and responsible **AI** frameworks (Ng et al., 2021; Gunasekara et al., 2025; OECD & European Commission, 2025).

Second, **the word cloud analysis** visually reinforces this interpretation by showing a discourse dominated by higher education, digital learning, and academic credentials. Artificial intelligence appears as a normalized and instrumentalized tool embedded in socially situated and ethically informed narratives, aligning with findings on **AI-powered educational agents** and institutional adoption (Córdova-Esparza, 2025).

Third, **the co-occurrence analysis** positions ethics and regulation of **AI** use as the central discursive axis, strongly associated with critical thinking, **AI** collaboration, constructive feedback, and metacognitive reflection. This pattern suggests that ethical considerations operate as the primary organizing principle through which cognitive, participatory, and evaluative practices are articulated, echoing research on procedural fairness, algorithmic accountability, and data justice (Decker et al., 2025; Demirchyan, 2025; Gonzalez-Argote et al., 2025).

Fourth, the conceptual network analysis formalizes these relationships by modeling directional dependencies among variables. Critical **AI** literacy functions as a foundational driver that grounds ethical awareness, enables sensitivity to algorithmic fairness, and sustains reflective interaction with AI. All pathways converge in **CAITIZEN**, conceptualized as an ethically grounded, cognitively informed, and socially responsible form of **AI-assisted citizenship** (Papagiannidis et al., 2025; UNESCO & CANIETI, 2025).

Fifth, **the sentiment analysis** adds an affective dimension, revealing a predominance of positive and neutral sentiment across all categories, particularly ethics, critical thinking, and collaboration. This pattern indicates legitimacy and acceptance rather than emotional polarization, supporting international evidence that responsible and transparent **AI** integration fosters trust and social alignment (WEF, 2025; United Nations, n.d.).

Together, these five results demonstrate that the **CAITIZEN model** captures **AI** use as an integrated ethical–cognitive–social system rather than a purely technical innovation.

From an innovation perspective aligned with the OCDE & Eurostat (2018) the **CAITIZEN model** can be understood as a form of social and conceptual innovation rather than a technological one. According to this framework, innovation includes new conceptual structures, organizational models, and social practices that generate value by transforming behavior and meaning. In this sense, the **CAITIZEN model** responds directly to the identified gap in **AI-assisted citizenship** by introducing an integrative framework that reconfigures how ethical regulation, critical cognition, metacognitive control, and human–**AI** collaboration are articulated within higher education contexts. Its innovative character lies in redefining **AI** use as a formative, ethically grounded, and socially situated process, contributing to sustainable educational transformation rather than technological efficiency alone.

6.1. Theoretical contribution (*Scientia*).

This study makes a substantive theoretical contribution by proposing and empirically substantiating the **CAITIZEN model** as an integrative framework for understanding artificial intelligence use in education as an ethical–cognitive–social system, rather than as a purely technological innovation. The model advances existing **AI** literacy and responsible **AI** frameworks by articulating how ethical regulation, critical cognition, metacognitive control, and human–**AI** collaboration dynamically interact to shape **AI-assisted citizenship** (Ng et al., 2021; Papagiannidis et al., 2025).

At the theoretical level, the model positions ethics and regulation of **AI** use as the foundational axis that structures **AI** engagement. This extends governance-oriented perspectives by demonstrating that ethical legitimacy and procedural fairness are not external constraints, but central organizing principles that condition cognitive practices, collaboration, and decision-making (Decker et al., 2025; Demirchyan, 2025; Gonzalez-Argote et al., 2025). In doing so, the **CAITIZEN model** bridges ethical **AI** governance with educational theory, reinforcing calls for socially grounded and accountable **AI** adoption (Gunasekara et al., 2025; UNESCO & CANIETI, 2025).

A second contribution lies in conceptualizing critical **AI** literacy and metacognitive prompting practices as key mediating mechanisms. Rather than treating **AI** literacy as a static skill set, the model theorizes it as a reflexive capacity that enables individuals to regulate interaction with **AI**

149

systems through conscious evaluation, prompt design, and decision-making (Haidar et al., 2025; Tsakeni et al., 2025; Waaler et al., 2025). This perspective extends prior literacy models by integrating metacognition and ethical awareness as co-constitutive elements.

Finally, the **CAITIZEN** model contributes to human–**AI** collaboration theory by framing creativity, participation, and agency as ethically conditioned outcomes of interaction with AI, rather than automatic byproducts of technological capability (Georgieva & Georgiev, 2025; Rafner et al., 2025; Salma et al., 2025). By converging these dimensions, the model offers a theoretically coherent explanation of **AI-assisted citizenship** aligned with global educational, workforce, and sustainability agendas (OECD, 2025; United Nations, n.d.; WEF, 2025).

6.2. Practical contributions (*Praxis*).

The **CAITIZEN model** offers several practical implications for educational institutions, policymakers, educators, and designers of **AI-based learning systems**.

First, the central role of ethics and regulation of **AI** use highlights the need for institutions to move beyond ad hoc guidelines and toward institutionalized governance frameworks that embed ethical standards, procedural fairness, and accountability into everyday educational practices involving **AI** (Decker et al., 2025; Demirchyan, 2025; Gunasekara et al., 2025). Universities and training organizations can operationalize this by integrating ethical review protocols, transparency requirements, and participatory oversight mechanisms into **AI** adoption strategies.

Second, the prominence of critical **AI** literacy and critical skills suggests that professional development and curriculum design should prioritize reflective and evaluative competencies rather than focusing solely on technical proficiency**. AI** literacy programs should explicitly train learners to assess bias, fairness, and social impact, aligning with international competency frameworks and workforce preparedness agendas (Ng et al., 2021; Miao & Cukurova, 2024; OECD, 2025; WEF, 2025). This has direct implications for teacher education and lifelong learning initiatives.

Third, findings related to metacognitive prompting practices indicate that educators should guide learners in the intentional formulation, evaluation, and revision of prompts when interacting with generative **AI**. Embedding metacognitive scaffolds into **AI** tools and instructional design can enhance learning quality while mitigating overreliance on automated outputs (Haidar et al., 2025; Tsakeni et al., 2025; Waaler et al., 2025).

150

Fourth, the positive positioning of human–**AI** collaboration and creative practices implies that **AI** should be deployed as a co-creative partner rather than as a substitute for human agency. Educational technologies should therefore be designed to support shared agency, participatory engagement, and collaborative problem-solving (Georgieva & Georgiev, 2025; Rafner et al., 2025; Salma et al., 2025).

Finally, the broadly positive and neutral sentiment surrounding **AI** use underscores the importance of sustaining trust through transparent communication and alignment with broader social and sustainability goals (United Nations, n.d.; UNESCO & CANIETI, 2025). Collectively, these implications position the **CAITIZEN model** as a practical guide for responsible, reflective, and socially legitimate **AI** integration in education.

7. CONCLUSION

This section synthesizes the main contributions of the study by reconnecting the findings with the challenges identified in the introduction regarding the accelerated integration of artificial intelligence in higher education and the limitations of predominantly technical approaches to digital literacy. The conclusions are structured around three complementary dimensions that reflect the qualitative, conceptual, and formative nature of the research.

7.1. Conceptual response to the research purpose

In alignment with the research purpose, this study demonstrates that the educational use of artificial intelligence cannot be reduced to an instrumental or efficiency-driven practice. Instead, the findings show that **AI** use among university students is configured as an ethical–cognitive–social system in which values, critical reasoning, and reflective regulation are central. Consistent with innovation perspectives informed by the Oslo Manual (OCDE & Eurostat, 2018), the **CAITIZEN Model** is articulated as a form of conceptual and social innovation that addresses the lack of integrative frameworks for **AI-assisted citizenship.** The model conceptualizes citizenship not as a fixed or fully achieved condition, but as a trajectory of citizenship in formation developed within higher education contexts.

7.2. Synthesis of the main findings

The qualitative results reveal a robust and thematically saturated categorical structure, supported by strong co-occurrence patterns among ethics, critical thinking, metacognition, and human–**AI** collaboration. The ethics and regulation of **AI** use emerge as the central organizing axis that structures other dimensions, including fairness and data justice, creative collaboration, and decision-making. Conceptual network analyses confirm the interdependence and directional relationships among the five core dimensions of the **CAITIZEN model**, while sentiment analysis indicates a predominance of positive and neutral evaluations. This affective configuration suggests legitimacy and acceptance of ethically governed **AI** use in educational environments rather than polarization or resistance.

7.3. Scope, implications, and future research

From a theoretical perspective, this study advances **AI** literacy research by integrating ethical judgment, metacognitive regulation, and collaborative agency into a unified analytical framework. From a practical standpoint, the **CAITIZEN model** offers guidance for educational policy, curriculum design, and institutional governance oriented toward responsible and sustainable **AI** integration. The study is limited by its qualitative and context-specific design.

Future research may extend the model through longitudinal, comparative, or mixed-method studies to examine how **AI-assisted citizenship** evolves across diverse educational and socio-cultural settings.

8. REFERENCIAS

Córdova-Esparza, D.-M. (2025). AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges.*Information, 16* (6), 469. https://doi.org/10.3390/info16060469

Decker, M., Wegner, L. & Leicht-Scholten, C. Procedural fairness in algorithmic decision-making: the role of public engagement. Ethics Inf Technol 27, 1 (2025). https://doi.org/10.1007/s10676-024-09811-4

Demirchyan G (2025) Algorithmic fairness: challenges to building an effective regulatory regime. Front. Artif. Intell. 8:1637134. https://doi.org/10.3389/frai.2025.1637134

Georgieva, I., & Georgiev, G. V. (2025). Exploring the use of generative text AI in design creativity inquiries. *Computers in Human Behavior: Artificial Humans, 6*, 100219. https://doi.org/10.1016/j.chbah.2025.100219

González-Argote J, Maldonado E, Maldonado K. (2025).Algorithmic Bias and Data Justice: ethical challenges in Artificial Intelligence Systems. *EthAIca* 4:159. https://ai.ageditor.ar/index.php/ai/article/view/159

Gunasekara, L., El-Haber, N., Nagpal, S., Moraliyage, H., Issadeen, Z., Manic, M., & De Silva, D. (2025). A Systematic Review of Responsible Artificial Intelligence Principles and Practice.*Applied System Innovation, 8*(4), 97. https://doi.org/10.3390/asi8040097

Haidar, H., Suryoputro, G., & Safi'i, I. (2025). Impact of the integration of metacognitive prompts by generative artificial intelligence (GenAI) in collaborative and individual learning in improving writing skills and metacognitive awareness. *International Journal of Learning, Teaching and Educational Research, 24*(6), 232–250. https://doi.org/10.26803/ijlter.24.6.11

INEGI (2023). *Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2023*. https://www.inegi.org.mx/programas/endutih/2023/

Kong, S. C., & Zhu, J. (2025). Developing and validating an artificial intelligence ethical awareness Scale for secondary and university students: Cultivating ethical awareness through problem-solving with artificial intelligence tools. *Computers and Education: Artificial Intelligence, 9*, Article 100447. https://doi.org/10.1016/j.caeai.2025.100447

Mejía-Trejo, J. (2025). *Inteligencia artificial y su repercusión en la educación superior*. AMIDI Editorial. https://doi.org/10.55965/abib.9786076984543

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. *UNESCO.* https://doi.org/10.54675/ZJTE2084

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

OECD & Eurostat. (2018). *Oslo manual 2018: Guidelines for collecting, reporting and using data on innovation* (4th ed.). OECD Publishing. https://doi.org/10.1787/9789264304604-en

OECD (2025). Bridging the AI skills gap: Is training keeping up? *OECD Publishing.* https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/04/bridging-the-aiskillsgap_b43c7c4a/66d0702e-en.pdf

OECD & European Commission (2025). *AI literacy framework for primary and secondary education.* https://learnworkecosystemlibrary.com/initiatives/ai-literacy-framework- for-primary-secondary-education-oecd-ec/

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems, 34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Pham, N., Pham Ngoc, H., & Nguyen-Duc, A. (2025). *Fairness for machine learning software in education: A systematic mapping study. Journal of Systems and Software, 219*, 112244. https://doi.org/10.1016/j.jss.2024.112244

Rafner, J., Zana, B., Bang Hansen, I., Ceh, S., Sherson, J., Benedek, M., & Lebuda, I. (2025). Agency in Human-AI Collaboration for Image Generation and Creative Writing: Preliminary Insights from Think-Aloud Protocols. Creativity Research Journal, 1–24. https://doi.org/10.1080/10400419.2025.2587803

Salma, Z., Hijón-Neira, R., & Pizarro, C. (2025). Designing Co-Creative Systems: Five Paradoxes in Human–AI Collaboration.Information, 16 (10), 909.

https://doi.org/10.3390/info16100909.

Southworth, J., Migliaccio, K., Glover, J., Glover, J. N., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). *Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence, 4*, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Stetson, P.D., Choy, J., Summerville, N. *et al.* Responsible Artificial Intelligence governance in oncology. *npj Digit. Med.* 8, 407 (2025). https://doi.org/10.1038/s41746-025-01794-w

Tsakeni, M., Nwafor, S. C., Mosia, M., & Egara, F. O. (2025). Mapping the Scaffolding of Metacognition and Learning by AI Tools in STEM Classrooms: A Bibliometric–Systematic Review Approach (2005–2025). *Journal of Intelligence, 13* (11), 148. https://doi.org/10.3390/jintelligence13110148

United Nations. (n.d.). *The 17 Sustainable Development Goals*. United Nations. https://sdgs.un.org/goals

UNESCO & Cámara Nacional de la Industria Electrónica, de Telecomunicaciones y Tecnologías de la Información (CANIETI). (2025, November 4). *UNESCO and CANIETI implement a model for ethical and responsible artificial intelligence in Mexican companies.* UNESCO. https://www.unesco.org/en/articles/unesco-and-canieti-microsoft-support-implementmodel-ethical-and-responsible-artificial-intelligence

Xiao, J., Alibakhshi, G., Zamanpour, A., Zarei, M. A., Sherafat, S., & Behzadpoor, S.-F. (2024). How AI literacy affects students' educational attainment in online learning: Testing a structural equation model in higher education context. *The International Review of Research in Open and Distributed Learning, 25*(3), 179–198. https://doi.org/10.19173/irrodl.v25i3.7720

Waaler, P. N., Hussain, M., Molchanov, I., Bongo, L. A., & Elvevåg, B. (2025). Prompt engineering an informational chatbot for education on mental health using a multiagent approach for enhanced compliance with prompt instructions: Algorithm development and validation. *JMIR AI, 4*, e69820. https://doi.org/10.2196/69820

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67,* Article 101187. https://doi.org/10.1016/j.jslw.2025.101187

Wang N, Kim H, Peng J and Wang J (2025) Exploring creativity in human–AI co-creation: a comparative study across design experience. *Front. Comput. Sci. 7*:1672735. https://www.frontiersin.org/journals/computerscience/articles/10.3389/fcomp.2025.1672735/full

WEF (2025). *Future of Jobs Report 2025.* WEF. https://www.weforum.org/stories/2025/05/why-ai-literacy-is-now-a-core-competency-in-education/