Guardians of Truth in Algerian Academia: Developing Ethical AI Frameworks for Multilingual Hypothesis-Driven Research

Hala LOUCIF

Department of English Language and Literature. Batna 2 University, Algeria.
Email : h.loucif@univ-batna2.dz

1: 0009-0008-3230-550X

Received	Accepted	Published
08/08/2026	20/08/2026	31/08/2026

DOI:https://doi.org/10.63939/JAAS.2026-Vol9.N30.62-77

Loucif, H. (2026). Guardians of truth in Algerian academia: Developing ethical AI frameworks for multilingual hypothesis-driven research. *Afro-Asian Studies, 9*(30), 61–77. https://doi.org/10.63939/JAAS.2026-Vol9.N30.62-77

Abstract

The proliferation of artificial intelligence (AI) technologies has transformed academic research practices across higher education institutions around the world. While AI-powered tools such as large language models, neural machine translation systems, and automated writing assistants offer significant opportunities for enhanced research productivity, they also pose complex ethical and linguistic challenges. These issues are particularly acute in multilingual and postcolonial academic contexts, where researchers are often required to navigate between multiple languages, traditions, and knowledge systems. This study explores the ethical implications of AI-assisted research in Algerian higher education and presents a context-sensitive framework for responsible AI use in multilingual hypothesis-based research. Using a convergent mixed-methods approach, this study examines the experience of 45 master students at the English department of Batna 2 University enrolled in a research methodology course. The results indicate that structured AI ethics education significantly improves students' ethical decision-making skills, awareness of algorithmic bias, and their ability to engage critically with AI generated outputs. The study also demonstrates that mainstream AI systems tend to favor Anglophone linguistic norms, which could marginalize local epistemologies and multilingual scholarly practices. In response, this article proposes the Algerian AI Ethics Triade: a framework based on the principles of Transparency, Cultural Relevance, and Human Primacy. The framework offers practical recommendations for researchers, educators, and policymakers who want to incorporate AI into academic research without compromising integrity, linguistic diversity, and intellectual autonomy. The findings contribute to emerging discussions on responsible AI governance and offer implications for multilingual higher education systems across the Middle East and North Africa (MENA) region.

Keywords: *artificial intelligence ethics, multilingual research, academic integrity, EFL research, higher education. .*

1. Introduction

Artificial intelligence is evolving rapidly and has become an important component of academic life. AI-powered applications are now used in universities around the world for activities such as literature review, data analysis, proofreading, translation, and even writing papers. The emergence of advanced large language models (LLMs) has accelerated this transformation by giving students and researchers the opportunity to generate sophisticated language and process texts. Consequently, AI technologies increasingly influence how people learn and share information, as well as how knowledge is produced, communicated, and evaluated within academic communities.

Although these new developments offer considerable benefits, they also raise ethical issues related to authorship, intellectual autonomy, and scientific integrity. Academic institutions have traditionally thought that people are responsible for their work and ideas. Now that AI can generate text, synthesize information, and influence the decision-making process, these assumptions are greatly challenged. This is why it has become crucial for educators, researchers, and policymakers to determine how to integrate AI responsibly without compromising fundamental academic values.

These concerns are especially important in multilingual and postcolonial settings. Most of the existing discourse on AI ethics has emerged from North America and Europe, where English dominates academic communication and technology is well established. However, these discussions often overlook the reality of researchers working in linguistically diverse environments where access to resources is unequal and contribution to global knowledge production can be limited. Therefore, new approaches are needed to address the potential benefits and risks of AI-powered technologies in these specific settings.

Algeria provides a very interesting context for investigating these issues. The country's linguistic landscape reflects a complex interaction among Modern Standard Arabic, Algerian Arabic (Darija), Tamazight, French, and increasingly English. Researchers often use many of these languages in their professional or academic routines. English is becoming the preferred language of international publication, forcing scholars to conduct, write, and share their research in a language that is neither their mother tongue nor the primary language they use in the academic environment.

In this situation, AI-powered tools offer practical solutions to linguistic barriers. Neural translation systems can facilitate access to international data, automated writing assistants can improve language accuracy, and generative AI platforms can support brainstorming and information synthesis. However, these technologies may also reproduce linguistic hierarchies mainly because they were developed using English-language corpora. They often privilege Anglophone rhetorical conventions and communicative norms. This can be problematic for multilingual researchers who have to conform to dominant academic paradigms at the expense of their local knowledge traditions, cultures, and modes of expression.

The implications of these new developments and the way they are changing research are not just about language. AI systems can shape the formulation of research questions, influence theoretical framing, and affect the interpretation of findings. Researchers who rely excessively on AI-generated outputs may weaken their critical engagement with knowledge, hinder their intellectual autonomy, and eventually undermine their scholarly originality. Furthermore, AI systems are difficult to understand, which complicates the efforts to evaluate the reliability, validity, and bias of generated content. These concerns are particularly relevant for graduate students who are still developing their research identities and methodological competencies.

This study addresses these challenges by examining how AI ethics education can support the responsible integration of AI in multilingual academic research. This study investigates the experiences of Master 02 EFL students at Batna 2 University, Algeria, and evaluates the impact of a structured educational intervention designed to promote ethical awareness, critical AI literacy, and autonomous decision-making.

This paper is guided by three research questions:

- How does structured AI ethics education influence multilingual EFL students' ethical awareness of AI-assisted research?
- To what extent do students recognize and evaluate the algorithmic biases affecting multilingual academic practices?
- What principles should guide the development of contextually relevant ethical frameworks for AI-assisted research in higher education in Algeria?

By addressing these questions, the article informs an emerging body of literature on AI ethics in higher education and responds to increasingly urgent calls for culturally responsive approaches to AI governance. The study, while acknowledging that AI systems can be both threats and solutions, interacts with alternative manifestations of a more systemic approach to responsible human-AI collaboration from the perspective of transparency, critical reflection, and respect for linguistic diversity. Ultimately, this article asserts that multilingual researchers must be positioned as **guardians of truth**: not only scholars who can leverage technological innovation, but also protectors of epistemological veracity, authenticity, and cultural richness in academic knowledge production.

2. Theoretical Framework

2.1. Artificial Intelligence and the Transformation of Academic Research

The use of AI in research represents one of the most significant transformations in higher education. Recent advances in AI-powered technologies have expanded the range of work that can be partially automated within scholarly tasks. Researchers are now using AI tools to find relevant literature, summarize academic texts, assist with the organization of qualitative data, generate statistical outputs, and even write different versions of their research papers, which changes completely the why researchers work. Surveys conducted across multiple higher education systems indicate that the majority of graduate students now regularly use AI tools in some phase of the research process—from literature review and data analysis to writing and translation (Foltynek et al., 2023; Lund et al., 2023).

Proponents of AI-assisted research emphasize its ability to make research more efficient and accessible. For multilingual scholars, AI technologies can reduce linguistic barriers that have limited participation in international academic communities in the past. Automated translation systems help them read papers in different languages, while writing assistants help researchers produce high-quality research papers that fit international publishing standards. Such capabilities have the potential to make knowledge production fairer by enabling broader participation in global scholarly discussions.

Nevertheless, the growing integration of AI raises concerns about academic expertise and intellectual work. Several scholars have argued that generative AI changes conventional understandings of authorship and originality because AI-generated content often comes from interactions that mix human and machine contributions. Hosseini et al. (2023) argue that the use of LLMs, such as ChatGPT, in academic writing raises fundamental questions about intellectual contribution, originality, and accountability. The question

now is how researchers should ethically use these technologies in a way that ensures their responsibility for the work they produce (Stokel-Walker & Van Noorden, 2023).

A central concern involves the reliability of AI-generated information. Unlike traditional search engines, generative AI systems produce sentences based on what they think is right instead of providing direct verified knowledge. Consequently, they may generate plausible but inaccurate statements, fabricate references, or misrepresent complex theoretical arguments. Bender et al. (2021), in their influential analysis of large language models, identified the phenomenon of stochastic parroting: LLMs generate fluent, plausible-sounding text by statistically modeling the distribution of language in their training data without any mechanism for grounding this text in real-world knowledge or logical reasoning.

Foltynek et al. (2023) note that the detection of AI-generated academic text remains technically unreliable, creating enforcement challenges that have led some institutions to adopt blanket prohibitions on AI use, while others have moved toward disclosure-based governance models. Neither approach adequately addresses the specific needs of multilingual researchers for whom AI tools may be genuinely necessary as accessibility accommodations. Therefore, it is important to verify every AI-generated output throughout the research process.

2.2. Algorithmic Bias and Epistemic inequality

One of the most frequently discussed AI system issues is algorithmic bias, which occurs when technological systems systematically favor certain groups, perspectives, or forms of knowledge over others. In language technologies, such biases often reflect inequalities present within training datasets, which affect how AI systems work and the results they produce. The goal is to design AI technologies that promote equality and fairness.

The use of English-language content within AI training data significantly affects multilingual researchers (Bender et al., 2021; Hovy & Spruit, 2016). Because many AI systems are optimized using data derived from English-speaking cultural settings, they often reproduce assumptions embedded within those contexts. As a result, language structure, rhetorical conventions, and knowledge associated with minority languages may not be well represented or understood. Canagarajah (2013) has documented this phenomenon and calls it the Anglophone research monoculture: the tendency of mainstream academic publishing and its associated gatekeeping mechanisms to treat Anglophone conventions as universal scholarly standards rather than as one culturally specific tradition among many.

These considerations can lead to what scholars call epistemic injustice in the sense developed by Fricker (2007): harm done to researchers as knowers, in their capacity to produce and have recognized legitimate scholarly knowledge. Epistemic injustice occurs when individuals or communities are disadvantaged because their knowledge is not valued or because institutional structures impede recognition of their intellectual contributions. Within multilingual academic environments, AI systems may accidentally reinforce epistemic inequalities by favoring dominant forms of knowledge production while marginalizing alternative perspectives (Lillis & Curry, 2010).

For Algerian researchers, these concerns are particularly important. Academic communication often involves switching between Arabic, French, and Tamazight, which systematically differ from Anglophone norms. Each of these languages carries specific historical, cultural, and epistemological significance. Tamazight, for instance, carries ideological implications that intersect with post-independence language politics (Benrabah, 2007). When AI tools systematically encourage conformity to Anglophone norms, they may diminish the visibility of locally grounded ways of knowing and communicating. Hovy and Spruit

(2016) introduce the concept of social bias in NLP, identifying ways in which natural language processing systems systematically disadvantage speakers of minority languages, non-standard dialects, and non-Anglophone varieties. Consequently, talking about AI ethics in multilingual settings should not be limited to technical considerations but should also address broader questions of linguistic justice and cultural representation.

2.3. Research Ethics in Multilingual Academic Context

Research ethics has traditionally focused on principles such as informed consent, confidentiality, and integrity. However, the rise of AI-assisted research necessitates an expansion of these frameworks to encompass new ethical challenges associated with cultural considerations. This means that research ethics frameworks that govern academic inquiry in Algerian and MENA university contexts cannot be read simply as local instantiations of universal principles (Macfarlane, 2009). The main rules for research ethics in English-speaking universities are rooted in post-World War II biomedical ethics, with their emphasis on individual autonomy, informed consent, and risk minimization. These ethics reflect specific philosophical traditions and institutional arrangements that do not map straightforwardly onto the collectivist social orientations, relational epistemologies, and differently structured institutional environments characteristic of many MENA research contexts (Hamdan, 2012).

In multilingual academic environments, these issues intersect with longstanding concerns regarding linguistic inequality. Scholars working outside Anglophone contexts frequently encounter structural barriers within international publishing systems (Curry & Lillis, 2004). The requirement to publish in English often creates additional cognitive, financial, and professional burdens. AI tools may alleviate some of these challenges; however, they can also deepen dependencies on technologies developed according to external norms and priorities.

Lillis and Curry (2010) document the practices of what they call academic literacy brokers—colleagues, editors, and commercial services who assist non-Anglophone scholars in producing internationally publishable texts. The ethical evaluation of AI use therefore requires sensitivity to contextual realities. What constitutes responsible AI use in a monolingual English-speaking institution may differ substantially from what is appropriate in multilingual postcolonial contexts. Ethical frameworks must acknowledge the complexities of language politics, cultural identity, and unequal participation in global knowledge economies.

Moreover, graduate students represent a particularly important population for AI ethics education. As emerging researchers, they simultaneously develop methodological competence, disciplinary expertise, and professional identities. Their experiences with AI technologies are likely to shape future research practices across academic fields. Flowerdew (2019) argues that non-Anglophone researchers face what he calls an asymmetric world of academic publishing—one in which native English speakers enjoy systematic advantages in peer review, editorial gatekeeping, and citation practices. Consequently, educational interventions that promote critical AI literacy may have long-term implications for the integrity of scholarly communities. An ethical framework for AI use in multilingual research must therefore simultaneously be a pedagogical tool for developing individual researchers' ethical capacities and a critical lens for interrogating the structural inequities that make AI tools both necessary and dangerous.

2.4. Emerging Approaches to AI Ethics Education

The field of AI ethics education is young and rapidly evolving. Formative contributions by Jobin et al. (2019), who conducted a systematic review of global AI ethics guidelines and identified convergent principles, including transparency, justice, non-maleficence, responsibility, and privacy, provide a useful starting point. AI literacy extends beyond technical understanding to include critical awareness of how AI systems function, how they influence social practices, and how ethical considerations should inform their deployment. However, as Hagendorff (2020) argues in a critical analysis of AI ethics guidelines, there is a significant gap between the principles articulated in ethics frameworks and the actual behavior of AI developers and users—what he calls the ethics washing problem: the production of ethical guidelines that signal virtue without generating substantive accountability.

Educational approaches increasingly encourage students to evaluate AI outputs critically rather than accepting them unreflectively. Long and Magerko (2020) proposed a competency-based model of AI literacy that includes understanding what AI is and how it works, critically evaluating AI outputs, and reflecting on the social and ethical implications of AI deployment. Selwyn (2022) extends this toward a more explicitly critical and political orientation, arguing that AI education must address the power relations and structural inequities that shape AI development and deployment, not merely the technical features of AI systems.

However, much of the existing literature focuses on general educational settings and remains grounded in Western institutional contexts. Research examining AI ethics education in multilingual and postcolonial environments remains relatively limited. The present study responds to this gap by investigating AI ethics education within the Algerian higher education context and proposing a framework that integrates the principles of transparency, cultural relevance, and human-centered scholarly agency. This study seeks to contribute to the development of ethical AI practices that support both technological innovation and epistemic diversity.

3. Methodology

3.1 Research Design

This study employed a convergent parallel mixed-methods approach (Creswell & Plano Clark, 2018) to investigate the effectiveness of a contextually adapted AI ethics intervention for multilingual EFL researchers. This design was chosen because the study aimed to provide a comprehensive understanding of how students negotiated the opportunities and challenges associated with AI-assisted research.

A pragmatist epistemological framework guided the study, prioritizing methodological choices suited to the research questions rather than adherence to a single paradigmatic tradition. Quantitative data were collected through pre- and post-intervention surveys designed to assess changes in ethical awareness and decision making. Qualitative data were obtained through semi-structured interviews and classroom observations to explore the participants' perceptions, experiences, and evolving attitudes toward AI use in research.

The intervention extended over a period of twelve weeks and was embedded within a Master's Research Methodology course. Quantitative and qualitative datasets were analyzed independently before being integrated during interpretation to identify areas of convergence, complementarity, and divergence.

3.2 Participants and Setting

The participants were 45 Master 2 EFL students enrolled in the Research Methodology course at Batna 2 University, Algeria. The institution represents a suitable site for investigating AI ethics in multilingual research because students routinely navigate Arabic, French, and English throughout their academic activities.

The ages ranged from 22 to 27 years, and 64% of the participants were female. All participants had completed at least three years of university-level English instruction and had prior experience with AI tools for academic purposes, primarily translation and writing assistance.

Purposive sampling was employed because the selected cohort represented a population actively engaged in academic research while simultaneously relying on AI technology to support English-language scholarly writing. Participation was voluntary, and informed consent was obtained from all students before data collection began.

3.3. Educational Intervention

The 12-week intervention replaced the standard research methodology course for the participants. The program was designed around three foundational principles that later informed the development of the Algerian AI Ethics Triad (see Section 5):

- **Transparency** in AI use
- **Cultural Relevance** of research tools and knowledge frameworks
- **Human Primacy** in research decision-making

These principles were operationalized through four main, interconnected pedagogical components:

Phase 01: Ethical Foundations (Week 1-3)
Ethical Awareness Workshops introduced students to fundamental concepts in AI ethics, including algorithmic bias, epistemic autonomy, and authorship in AI-augmented research. These workshops drew on Jobin et al.'s (2019) global principles review and Bender et al.'s (2021) limitation analysis adapted for the Algerian EFL research context. Classroom discussions examined real-world examples of AI misuse in academic contexts and encouraged critical reflection on responsible research practice.

Phase 2: Comparative Human–AI Inquiry (Weeks 4–6)
Comparative human–machine hypothesis testing required students to formulate research hypotheses independently, then use AI tools to generate alternative or refined versions, followed by systematic comparison and critical evaluation of the two. This component aimed to develop students' capacity to use AI as a dialogic interlocutor rather than an authority. Through structured analysis, students evaluated the strengths, limitations, and underlying assumptions embedded in both human-produced and AI-generated outputs.

Phase 3: Bias Detection Workshops (Weeks 7–9)
During these workshops, students were engaged in structured exercises using local corpora designed to identify instances of algorithmic bias. The students investigated how AI tools processed multilingual academic texts. Particular attention was paid to identifying linguistic standardization practices and the potential forms of cultural or epistemic bias.

Phase 4: Ethical Policy Development (Weeks 10–12)
Participants collaborated in groups to design institutional guidelines for the ethical use of AI in higher education. This activity encouraged students to translate their theoretical knowledge into practical policy recommendations.

3.4. Data Collection Instruments

Three instruments were employed:

The Ethical Decision-Making in AI Research Survey (EDAIRS) consisted of 28 Likert-scale items measuring four dimensions: AI literacy (understanding how AI tools work), ethical reasoning (applying ethical principles to AI use scenarios), bias awareness (recognizing algorithmic bias in AI outputs), and autonomy orientation (preference for human versus AI agency in research decisions). Items were rated on a five-point Likert scale and validated through expert review (three specialists in AI ethics and two EFL research methodology instructors) and pilot testing with a comparable student group. Internal consistency was strong (Cronbach's α = .87 pre-intervention, α = .89 post-intervention).

Semi-structured interviews were conducted with 18 participants who were purposively selected following the intervention. The interviews lasted between 40 and 60 minutes and were conducted in the participants' preferred language(s). Questions explored participants' experiences of the intervention, their evolving understanding of AI ethics, their perceptions of algorithmic bias, and their intentions regarding AI use in their own research. Interviews were audio-recorded, transcribed, and translated into English.

Classroom observations were collected across ten structured sessions that were distributed throughout the intervention. Observations were conducted using an adapted version of the AI Ethics Education Observation Protocol (AEEOP), documenting ethical reasoning episodes, critical engagement with AI outputs, collaborative discussions, and students' responses to the bias-detection activities.

3.5. Data Analysis

Quantitative data were analyzed using paired-samples *t*-tests to determine statistically significant differences between pre- and post-intervention scores. Effect sizes were calculated using Cohen's *d* (Cohen, 1988). Qualitative data were analyzed using reflexive thematic analysis (Braun & Clarke, 2022), with NVivo 14 software supporting the coding and theme development process. Integration of quantitative and qualitative findings was achieved through joint-display analysis, allowing statistical findings and qualitative themes to inform one another.

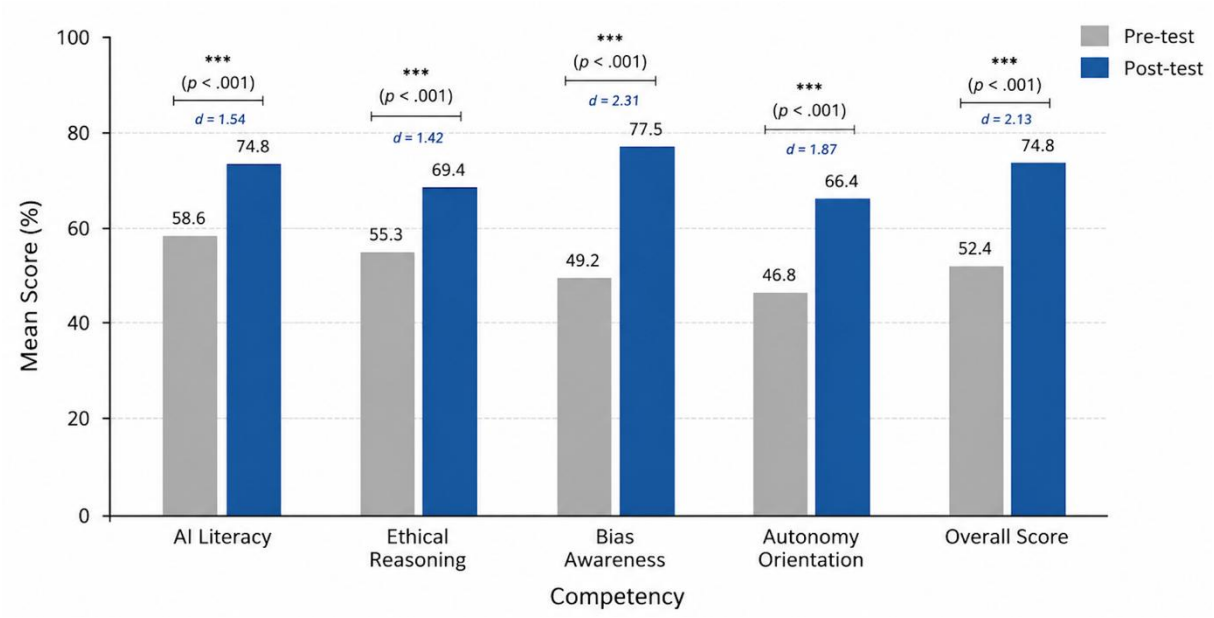

Figure 01:*Changes in Ethical Decision-Making Competence*

The results indicate statistically significant gains across all measured dimensions. The largest improvement occurred in bias awareness, suggesting that the participants became considerably more capable of recognizing how AI systems may privilege particular linguistic and epistemological norms.

These results are consistent with and extend prior findings on AI ethics education outcomes. Floridi et al. (2021) and Selwyn (2022) have argued that structured educational interventions can significantly improve students' capacity for critical ethical reasoning about AI, but evidence from multilingual, non-Anglophone contexts has been limited. The magnitude of the effect sizes observed in the present study—particularly on bias awareness—may reflect the heightened salience of algorithmic bias for students who experience its effects directly in their own multilingual research practice, as opposed to studying it as an abstract concept. This interpretation aligns with Mezirow's (1991) transformative learning theory, which holds that learning is deepest when it disrupts and reorganizes existing meaning frameworks through engagement with disorienting dilemmas—in this case, the discovery that the AI tools students had trusted were systematically biased against their own linguistic practices.

Interview data revealed a substantial shift in students' understanding of AI technologies. Prior to the intervention, many participants viewed AI-generated outputs as objective and neutral. Following the workshops, students reported increased awareness of the cultural assumptions embedded in AI systems.

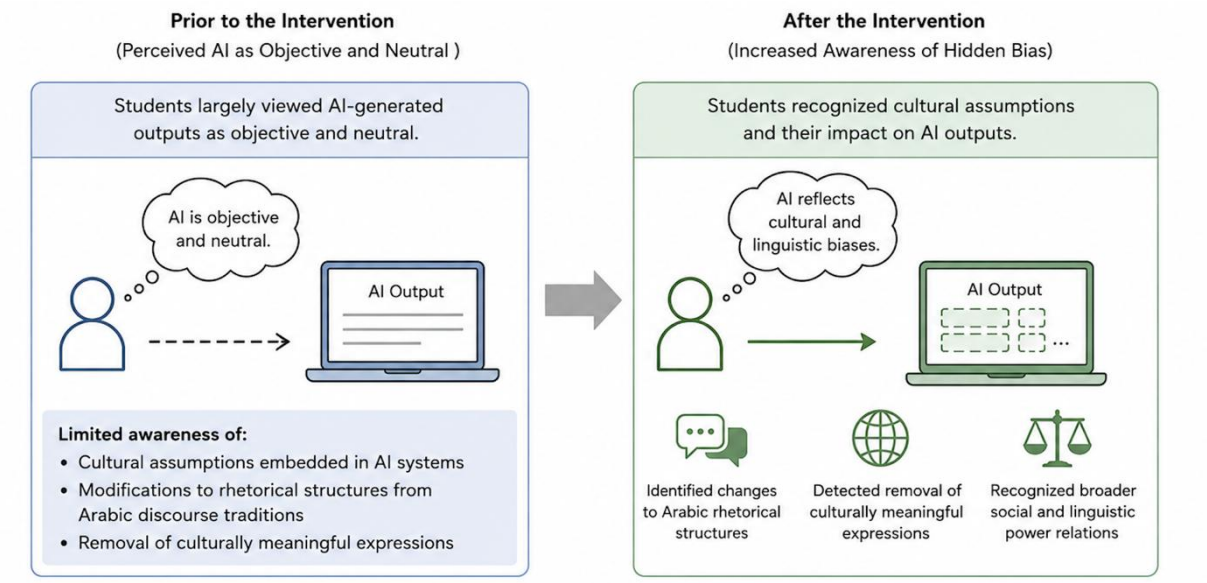

Figure 02: ***Students Understanding of Algorithmic Bias Pre and Post Intervention***

Participants frequently identified instances in which AI tools modified rhetorical structures originating from Arabic discourse traditions or removed culturally meaningful expressions from the text.

This theme suggests that ethical AI literacy involves not only understanding technological limitations but also recognizing how technologies reflect broader social and linguistic power relations.

> *"I always thought Grammarly was just correcting my mistakes. After the workshop, I realized it was also correcting things that were not mistakes—it was correcting the way I think, the way I organize ideas in Arabic, and pushing me toward a way of writing that feels foreign to me." (Participant 12, translated from Algerian Arabic and French)*

This finding resonates powerfully with Canagarajah's (2013) analysis of Anglophone research monoculture and Lillis and Curry's (2010) documentation of the invisible constraints imposed on

multilingual scholars by Anglophone publishing conventions. It extends these analyses by demonstrating that AI tools function as a new and particularly insidious vector for these constraints—insidious because they present their interventions as objective, neutral corrections rather than as culturally situated judgments.

The classroom observation data corroborated the interview findings. During the bias detection workshops, the students systematically identified AI-generated corrections that erased Arabic rhetorical structures (particularly patterns of parallel elaboration and invocation of collective authority), replaced Tamazight-inflected conceptual framings with Anglophone equivalents, and standardized French-Arabic code-mixed formulations into monolingual English. These observations align with Hovy and Spruit's (2016) theoretical analysis of social bias in Natural Language Processing (NLP) and with empirical work by Blodgett et al. (2020), who demonstrate that NLP systems consistently perform worse on text from African American Vernacular English speakers—a finding that extends, by implication, to other non-dominant language varieties.

Before the intervention, participants also reported their extensive reliance on AI systems. 87% of them used AI tools daily in their research activities, primarily for translation (DeepL), grammar correction (Grammarly), and information synthesis (ChatGPT). Many students described accepting AI outputs with minimal evaluation. After the intervention, students increasingly adopted verification practices such as cross-checking references, comparing multiple AI outputs, consulting peer-reviewed sources, and extensively revising AI-generated content.

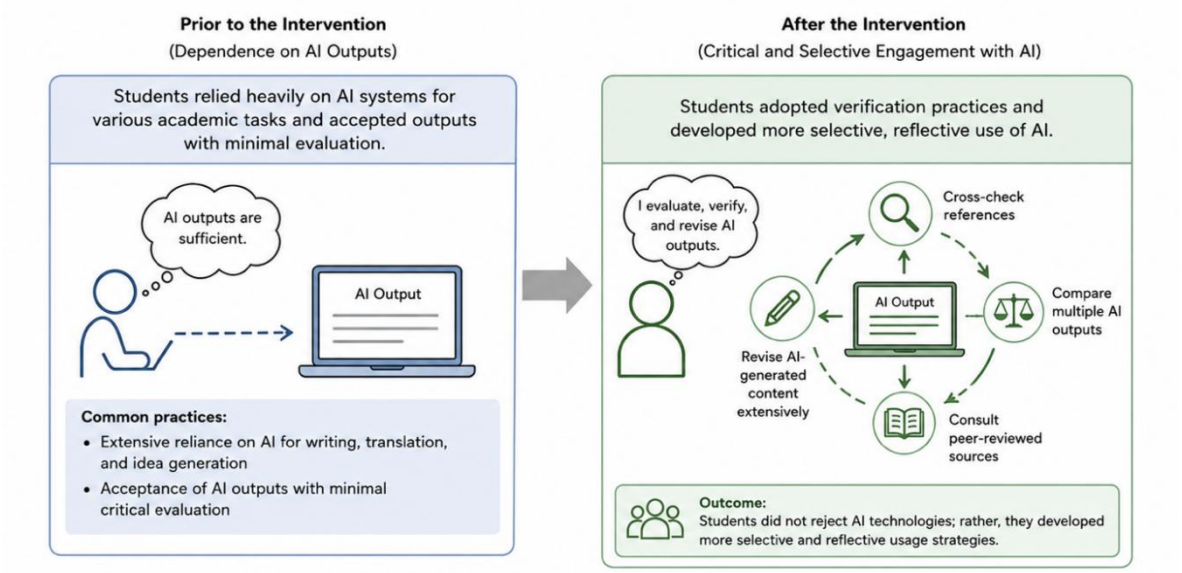

Figure 03: ***Changing Patterns of AI Use: From Dependence to. Critical Engagement***

Rather than rejecting AI technologies altogether, participants developed more selective and reflective usage strategies. Post-intervention interviews revealed a consistent pattern of what might be called calibrated appropriation: students described retaining AI tool use for tasks where the tools' affordances clearly outweighed their risks (e.g., checking specific grammatical constructions in English), while withdrawing AI involvement from higher-stakes epistemic activities (e.g., formulating research hypotheses, interpreting data, and drawing theoretical conclusions). This pattern aligns with Long and Magerko (2020) identification of evaluative AI literacy—the capacity to assess when AI tool use is appropriate, beneficial, and ethically acceptable.

The concept of calibrated appropriation also resonates with Luckin's (2018) argument for what she calls intelligent book—a model of human-AI collaboration in which AI systems function as cognitive scaffolds that support and extend human intelligence rather than replacing it. The present findings suggest that structured ethical education can move student researchers from patterns of uncritical AI dependence toward more sophisticated human-AI collaboration practices—a transition that benefits both their research quality and their development as autonomous scholars.

The question of authorship emerged as one of the most frequently discussed ethical concerns, which 71% of the participants described as extensive AI involvement in the research and writing process.

> *"When I read back a paragraph that I wrote with a lot of AI help, I sometimes cannot tell which ideas are mine and which came from the machine. It feels like my voice has been diluted. How do I put my name on something that is partly not mine?" (Participant 23, translated from Algerian Arabic)*

Many participants expressed uncertainty regarding the ownership of AI-assisted writing. Students described situations in which AI-generated text appeared academically sophisticated yet felt disconnected from their own intellectual voice. Participants repeatedly emphasized the importance of maintaining intellectual ownership of ideas while using AI as a support tool rather than a substitute for critical thinking. This concern extended beyond plagiarism to broader questions of scholarly identity. This finding connects to Norton's (2013) concept of identity investment in language learning and to Archer's (2007) sociological analysis of inner conversation as the mechanism through which individuals develop and maintain their sense of agency and identity. If AI tools interrupt or colonize the inner conversation through which researchers develop and refine their ideas, the result may be a form of cognitive and epistemic alienation that goes beyond the scope of conventional academic integrity frameworks. Addressing this concern requires not only disclosure policies but pedagogical interventions that help researchers develop and protect their capacity for autonomous intellectual engagement—precisely the goal of the Algerian AI Ethics Triad framework.

The final theme involved participants' preference for collaborative, rather than replacement-oriented, models of AI use.

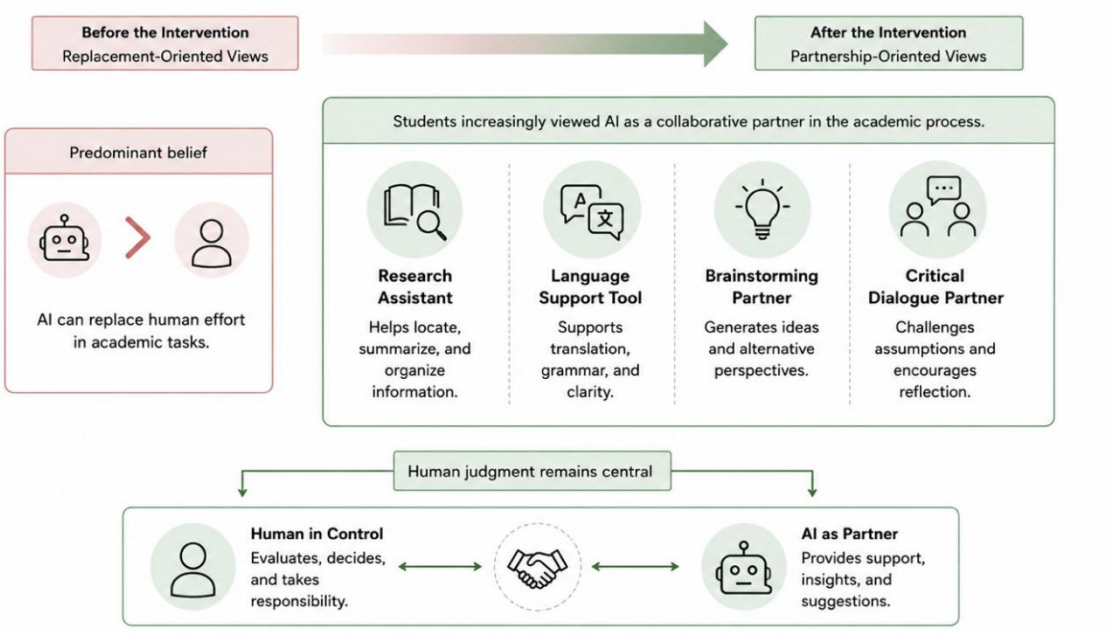

Figure 04: ***Emergence of Human-AI Partnership Models***

The interview and observation data consistently pointed towards a mode of AI use in which human judgment remains primary and AI assistance is systematically subordinated to human intellectual agency. This concept, which emerged inductively from the qualitative data, closely maps onto theoretical formulations of human-centered AI (Shneiderman, 2022) and reflects a sophisticated ethical position that neither rejects AI tools nor uncritically embraces them. Students increasingly conceptualized AI as a research assistant, a language support tool, a brainstorming partner, and a critical dialogue partner. However, they consistently emphasized the importance of maintaining personal research journals that documented their thinking process independently of AI tools; using AI outputs as a starting point for critical engagement rather than an endpoint for acceptance; regularly submitting AI-generated text to peers for evaluation of cultural and epistemic authenticity; and developing personal ethical checklists for AI use that went beyond institutional requirements. These practices represent a form of what Floridi et al. (2021) call value-sensitive design applied at the individual level—the proactive integration of ethical values into technology use through deliberate practice and reflective habit, which reflects a growing commitment to responsible human-AI collaboration rather than technological dependence.

5. The Algerian AI Ethics Triad

The findings of this study support the development of a contextually grounded framework for ethical AI use in multilingual academic environments. Existing AI ethics guidelines provide valuable general principles; however, many remain insufficiently responsive to the linguistic, cultural, and epistemological realities of researchers working in multilingual postcolonial contexts. The Algerian AI Ethics Triad was developed to address this gap by integrating global ethical principles with the specific challenges experienced by AI use in multilingual hypothesis-driven research. The framework is organized around three interdependent principles:

Transparency refers to the explicit disclosure and documentation of AI use throughout the research process. Rather than merely indicating that AI tools were used, researchers should clearly identify which AI systems were employed; the purpose of their use; the stages of the research process affected; and the extent to which outputs were modified by the researcher. The Transparency principle aligns with the convergent theme of transparency identified by Jobin et al. (2019) in their global review of AI ethics guidelines and with the disclosure recommendations advanced by Hosseini et al. (2023) in response to the ChatGPT authorship challenge. Its distinctive contribution in the Algerian context is the emphasis on transparency not only about the fact of AI use but about the evaluative process through which researchers maintain intellectual sovereignty over AI-assisted work—a dimension of transparency that is particularly important when AI tools are systematically biased against researchers' own linguistic and cultural frameworks.

Cultural Relevance emphasizes that AI systems should be evaluated not only according to technical performance but also according to their capacity to respect linguistic diversity and local knowledge traditions. This principle operationalizes what Fricker (2007) calls the value of epistemic justice—the recognition that researchers from non-Anglophone traditions have legitimate claims to produce scholarship within their own knowledge frameworks, not merely as imperfect approximations of Anglophone norms. Cultural Relevance therefore requires researchers to critically evaluate whether AI outputs preserve local meanings, respect multilingual discourse practices, represent culturally grounded knowledge accurately,

and avoid imposing external epistemological norms. By doing so, institutions can contribute to greater epistemic justice and protect intellectual diversity within academic research.

Human Primacy constitutes the central pillar of the framework. While AI technologies can assist with information processing, language support, and idea generation, responsibility for scholarly reasoning must remain fundamentally human. Human Primacy aligns with Shneiderman's (2022) concept of human-centered AI and with Luckin's (2018) vision of AI as intelligent scaffolding that supports human learning and cognition. Its implementation in pedagogical contexts involves comparative human-machine hypothesis testing exercises, which cultivate students' ability to use AI as a dialogic interlocutor rather than an authority to be followed. Research involves interpretation, judgment, creativity, and ethical responsibility. These dimensions cannot be delegated entirely to algorithmic systems. Universities should design assessment practices that reward critical thinking rather than merely polished outputs.

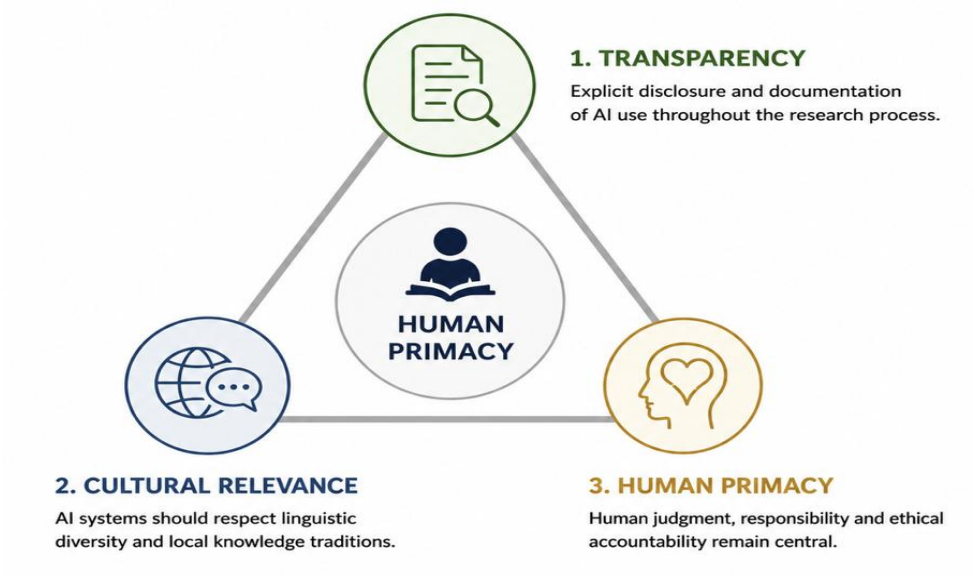

Figure 05: ***The Algerian AI Ethics Triad***

6. Recommendations and pedagogical implications

The findings underscore the need to integrate AI ethics as a core component of graduate research methodology curricula. Rather than responding reactively to the growing use of AI, universities should adopt proactive educational approaches that develop students' competencies in ethical decision-making, algorithmic bias awareness, AI literacy, academic integrity, and responsible human–AI collaboration. Importantly, AI ethics education should be contextualized to local linguistic, cultural, and academic realities instead of relying exclusively on frameworks developed in Anglophone contexts.

This study also highlights the importance of institutional policies that promote responsible AI use rather than restrictive or surveillance-based approaches. Higher education institutions should establish clear governance frameworks that include mandatory AI disclosure, guidelines for acceptable AI use, training for students and teachers, procedures for verifying AI-assisted work, and measures to protect multilingual scholarly practices. Such policies would align Algeria with emerging international discussions on AI governance and linguistic equity. The UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021), adopted by all member states, including Algeria, includes explicit provisions for protecting cultural diversity and linguistic rights in AI systems.

Finally, although the proposed Algerian AI Ethics Triad was developed within the Algerian higher education context, its principles have broader relevance for multilingual universities across the Middle East and North Africa (MENA). Given the region's shared characteristics, including linguistic diversity, postcolonial educational traditions, increasing international publication demands, and rapid AI adoption. The framework offers a contextually grounded model for ethical AI governance. As such, it provides a foundation for developing culturally responsive policies and educational practices that support innovation while preserving academic integrity and intellectual diversity.

7. Conclusion

Artificial intelligence is rapidly transforming the landscape of academic research. While these technologies offer unprecedented opportunities to improve efficiency, accessibility, and scholarly productivity, they also introduce significant ethical challenges that require careful consideration.

This study demonstrates that structured AI ethics education can substantially enhance multilingual EFL students' ethical decision-making abilities, awareness of algorithmic bias, and commitment to intellectual autonomy. The findings further reveal that AI technologies are not neutral tools but sociotechnical systems shaped by particular linguistic, cultural, and epistemological assumptions.

In response to these challenges, this study proposed the Algerian AI Ethics Triad, a framework built upon the principles of transparency, cultural relevance, and human primacy. The framework offers a practical model for responsible AI integration within multilingual academic environments and provides guidance for educators, researchers, institutions, and policymakers.

The present study has several limitations that should be acknowledged. The single-site, single-instructor design limits generalizability, and the relatively short duration of the intervention (12 weeks) may not capture the full extent of attitudinal and behavioral change that longer-term engagement with AI ethics education could produce. Future research should employ multi-site designs, include longitudinal follow-up measures, and investigate the organizational and policy conditions that support or hinder the implementation of AI ethics frameworks in diverse MENA institutional contexts.

Future studies should explore several important areas, such as teachers perspectives towards ethical AI integration within research training, the effectiveness of institutional AI governance policies, or the longitudinal effect of the investigation to examine whether the gain in ethical awareness persists beyond the immediate intervention period.

Therefore, the challenge for contemporary academia is not whether AI should be used but how it can be used responsibly while safeguarding the diversity, authenticity, and integrity of scholarly inquiry.

References

- Archer, M. S. (2007). Making our way through the world: Human reflexivity and social mobility. Cambridge University Press.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922
- Benrabah, M. (2007). Language-in-education planning in Algeria: Historical development and current issues. Language Policy, 6(2), 225–252. https://doi.org/10.1007/s10993-007-9046-7
- Blodgett, S. L., Barocas, S., Daumé, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5454–5476). ACL. https://doi.org/10.18653/v1/2020.acl-main.485
- Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. SAGE Publications.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology, 1(3), 185–216. https://doi.org/10.1177/135910457000100301
- Canagarajah, S. (2013). Translingual practice: Global Englishes and cosmopolitan relations. Routledge.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.
- Curry, M. J., & Lillis, T. (2004). Multilingual scholars and the imperative to publish in English: Negotiating interests, demands, and rewards. TESOL Quarterly, 38(4), 663–688. https://doi.org/10.2307/3588284
- Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2021). How to design AI for social good: Seven essential factors. Science and Engineering Ethics, 26(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5
- Flowerdew, J. (2019). The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. Language Teaching, 52(2), 249–260. https://doi.org/10.1017/S0261444819000041
- Foltynek, T., Bjelobaba, S., Glendinning, I., Khan, Z. R., Santos, R., Pavletic, P., & Kravjar, J. (2023). ENAI recommendations on the ethical use of artificial intelligence in education. International Journal for Educational Integrity, 19(1), 12. https://doi.org/10.1007/s40979-023-00133-4
- Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.
- Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Annals of Family Medicine, 13(6), 554–561. https://doi.org/10.1370/afm.1865
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8
- Hamdan, A. K. (2012). Reflexivity of discomfort in insider-outsider educational research. McGill Journal of Education, 47(3), 377–404. https://doi.org/10.7202/1014860ar
- Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. Accountability in Research, 30(6), 1–8. https://doi.org/10.1080/08989621.2023.2168535
- Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (pp. 591–598). ACL. https://doi.org/10.18653/v1/P16-2096

- Hyland, K. (2005). Metadiscourse: Exploring interaction in writing. Continuum.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
- Lillis, T., & Curry, M. J. (2010). Academic writing in a global context: The politics and practices of publishing in English. Routledge.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727
- Luckin, R. (2018). Machine learning and human intelligence: The future of education for the 21st century. UCL IOE Press.
- Macfarlane, B. (2009). Researching with integrity: The ethics of academic enquiry. Routledge.
- Mezirow, J. (1991). Transformative dimensions of adult learning. Jossey-Bass.
- Norton, B. (2013). Identity and language learning: Extending the conversation (2nd ed.). Multilingual Matters.
- Patton, M. Q. (2015). Qualitative research & evaluation methods (4th ed.). SAGE Publications.
- Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620–631. https://doi.org/10.1111/ejed.12532
- Shneiderman, B. (2022). Human-centered AI. Oxford University Press.
- Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. Nature, 614(7947), 214–216. https://doi.org/10.1038/d41586-023-00340-6
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). SAGE handbook of mixed methods in social and behavioral research (2nd ed.). SAGE Publications.
- UNESCO. (2021). Recommendations on the ethics of artificial intelligence. United Nations Educational, Scientific, and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., … Gabriel, I. (2021). Ethical and social risks of harm from language models. arXiv. https://arxiv.org/abs/2112.04359