

Artificial Intelligence for Women's Empowerment in Saudi Higher Education: A Human-Centred and Ethical Framework

Wafa Muhammed Alajaji

wafaalajaji@yahoo.com

University of Roehampton

Obby Phiri

University of Westminster

Research Article

Keywords: artificial intelligence, women's empowerment, Saudi Arabia, higher education, human-centred AI, AI ethics, Vision 2030, AI governance

Posted Date: August 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10587773/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Saudi Arabia's Vision 2030 has positioned both artificial intelligence (AI) and women's empowerment as twin pillars of national transformation, yet the intersection of these two agendas within higher education remains conceptually underdeveloped. Existing scholarship tends to treat AI adoption in education as a technical or pedagogical question, and treats women's empowerment as a separate social-policy concern, with little work connecting the two through a coherent governance lens. This paper addresses that gap by developing a Human-Centred Ethical AI Framework for women's empowerment in Saudi higher education. Drawing on an integrative conceptual synthesis of the Capability Approach, Human-Centred AI, Value Sensitive Design, the Technology Acceptance Model, the Unified Theory of Acceptance and Use of Technology, Self-Determination Theory, Social Cognitive Theory, and Digital Inclusion Theory, the paper proposes a multi-layered architecture linking national policy, institutional readiness, ethical AI principles, AI technologies, educational outcomes, and empowerment outcomes. The framework is positioned against established international AI governance instruments, including the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles, the EU AI Act, the NIST AI Risk Management Framework, and ISO/IEC 42001, and is adapted to the cultural, religious, and institutional context of Saudi Arabia. The framework's treatment of leadership empowerment is further grounded in original life-history evidence on Saudi women school leaders, which identifies the scarcity of structured mentoring and leadership-development infrastructure as a specific, empirically documented gap that appropriately governed AI-enabled tools could help address. The proposed framework also contributes to the achievement of the United Nations Sustainable Development Goals, particularly SDG 4 (Quality Education), SDG 5 (Gender Equality), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation and Infrastructure), SDG 10 (Reduced Inequalities), SDG 16 (Peace, Justice and Strong Institutions), and SDG 17 (Partnerships for the Goals).

1. Introduction

Artificial intelligence (AI) is rapidly transforming higher education by reshaping how teaching, learning, student support, and institutional decision-making are designed and delivered. The widespread adoption of generative AI, intelligent tutoring systems, adaptive learning platforms, learning analytics, and predictive educational technologies has created unprecedented opportunities to personalise learning, improve student outcomes, and enhance institutional efficiency. At the same time, AI introduces significant ethical challenges relating to transparency, fairness, accountability, privacy, algorithmic bias, and human autonomy. As universities increasingly integrate AI into their educational ecosystems, ensuring that these technologies promote inclusive and equitable educational opportunities rather than reinforcing existing inequalities has become an international policy priority.

This challenge is particularly significant in countries pursuing ambitious digital transformation agendas alongside broader social and economic reform. AI is increasingly recognised not only as a technological innovation but also as a strategic enabler of human capital development, economic competitiveness, and sustainable development. Consequently, responsible AI governance has emerged as a critical

requirement for ensuring that digital transformation contributes to inclusive growth while protecting human rights and promoting social equity. These priorities align closely with the United Nations Sustainable Development Goals (SDGs), particularly SDG 4 (Quality Education), SDG 5 (Gender Equality), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation and Infrastructure), SDG 10 (Reduced Inequalities), and SDG 16 (Peace, Justice and Strong Institutions), all of which emphasise the importance of inclusive, equitable, and human-centred technological innovation and, in the case of SDG 16, the accountable institutions needed to govern it.

Saudi Arabia provides a particularly compelling context within which to examine these issues. Since the launch of Vision 2030, the Kingdom has pursued an ambitious programme of economic diversification, digital transformation, and social reform, with higher education positioned as a central driver of national development. Simultaneously, Saudi Arabia has made substantial progress in expanding women's participation in higher education, employment, entrepreneurship, and leadership through initiatives such as the Human Capability Development Program and the National Strategy for Data and Artificial Intelligence coordinated by the Saudi Data and AI Authority (SDAIA). Universities therefore occupy a unique position at the intersection of these national priorities: they are expected both to prepare graduates for an AI-driven economy and to support women's increasing participation in economic and social development.

The convergence of these agendas presents both significant opportunities and complex governance challenges. AI-enabled educational technologies have considerable potential to enhance personalised learning, improve educational accessibility, strengthen digital and AI literacy, expand opportunities for lifelong learning, and increase employability among female graduates. At the same time, poorly governed AI systems risk reproducing gender bias, reinforcing historical inequalities, compromising student privacy, increasing algorithmic opacity, and introducing culturally inappropriate educational practices. Whether AI ultimately functions as an instrument of empowerment or a source of new forms of exclusion depends not simply on the technology itself but on how it is designed, governed, implemented, and evaluated within specific institutional, cultural, and regulatory contexts.

Despite rapidly growing scholarship in this area, existing research remains fragmented across several largely disconnected bodies of literature. First, studies of AI in higher education have extensively documented the capabilities of intelligent tutoring systems, learning analytics, adaptive learning platforms, automated assessment, and AI-enabled academic advising (Baidoo-Anu & Ansah, 2023; Bahroun et al., 2023). However, this literature has largely focused on technological functionality and educational performance, paying comparatively limited attention to the differentiated experiences of women or the influence of cultural and national contexts on responsible AI implementation.

Second, a substantial body of research has developed international frameworks for AI ethics and governance, including the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles, the European Union AI Act, the NIST AI Risk Management Framework, ISO/IEC 42001, and Saudi Arabia's own AI Ethics Principles developed by SDAIA. These frameworks provide valuable

guidance regarding transparency, accountability, fairness, privacy, explainability, and human oversight. Nevertheless, they are intentionally designed as broad governance instruments applicable across multiple sectors and jurisdictions. Consequently, they provide limited guidance for addressing the unique educational, gender-specific, cultural, and religious considerations associated with AI adoption within Saudi higher education.

Third, the literature on Saudi women's empowerment has documented remarkable progress in female participation in higher education, postgraduate education, STEM disciplines, employment, entrepreneurship, and leadership (Aldossari, 2025; EL-Deghaidy et al., 2025). However, relatively little attention has been given to the role of AI as a strategic enabler of this transformation or to the governance mechanisms required to ensure that AI contributes positively to women's educational and socio-economic advancement.

Finally, an emerging body of scholarship has begun examining the intersection between AI and gender in education. Recent systematic reviews have highlighted barriers to AI adoption among women, identified persistent algorithmic bias within educational AI systems, and called for more inclusive approaches to AI development (Kalim, Kanwar, Sha, et al., 2025; Kalim, Kanwar, Xu, et al., 2026). Although these studies represent important advances, they remain largely disconnected from the Gulf policy context and do not address how responsible AI governance should be operationalised within higher education institutions pursuing simultaneous digital transformation and women's empowerment agendas.

Accordingly, a significant conceptual gap remains. Existing research has yet to integrate responsible AI governance, women's empowerment, higher education, and Saudi Arabia's unique cultural, religious, and institutional context within a single coherent framework. Existing governance models remain largely generic, while studies of women's empowerment rarely engage with AI governance, and educational AI research seldom considers gender-specific empowerment outcomes. This fragmentation limits both theoretical understanding and practical guidance for universities seeking to implement AI responsibly while advancing national development priorities.

This paper addresses this gap by proposing a Human-Centred Ethical AI Framework specifically designed for women's empowerment in Saudi higher education. Unlike existing governance instruments, the proposed framework integrates national policy priorities, institutional readiness, human-centred AI principles, AI-enabled educational technologies, educational outcomes, and multidimensional empowerment outcomes within a single operational architecture. As summarised in Box 1, it is, to the authors' knowledge, the first conceptual framework explicitly designed to combine women's empowerment, higher education governance, and Saudi cultural and institutional context within a comprehensive responsible AI model.

Box 1. Positioning of the Proposed Framework Relative to Existing Governance Instruments

Framework	Human-Centred	Addresses Women	Higher-Education Specific	Governance Mechanisms	Saudi-Adapted
UNESCO Recommendation	Yes	Partial	No	Yes	No
OECD AI Principles	Yes	No	No	Yes	No
EU AI Act	Partial	No	Partial	Yes	No
NIST AI RMF	Yes	No	No	Yes	No
ISO/IEC 42001	Partial	No	No	Yes	No
SDAIA National AI Ethics Principles	Yes	No	No	Partial	Yes
Proposed Human-Centred Ethical AI Framework	Yes	Yes	Yes	Yes	Yes

The paper makes four principal contributions. First, it synthesises diverse theoretical perspectives—including the Capability Approach, Human-Centred AI, Value Sensitive Design, Technology Acceptance Model, Unified Theory of Acceptance and Use of Technology, Self-Determination Theory, Social Cognitive Theory, and Digital Inclusion Theory—into an integrated theoretical foundation for understanding AI-enabled women's empowerment. Second, it develops a novel multi-layered Human-Centred Ethical AI Framework that links national policy, institutional readiness, responsible AI governance, educational technologies, educational outcomes, and empowerment outcomes within a coherent conceptual architecture. Third, it proposes a phased implementation roadmap and practical policy recommendations for universities, SDAIA, and the Saudi Ministry of Education to facilitate responsible AI adoption. Finally, it establishes a future research agenda that advances scholarship on AI governance, women's empowerment, and higher education across Saudi Arabia and the wider Gulf Cooperation Council region while demonstrating how responsible AI can contribute to the achievement of multiple Sustainable Development Goals.

The remainder of the paper is organised as follows. Section 2 examines the evolution of women's higher education in Saudi Arabia and the policy context created by Vision 2030. Section 3 reviews contemporary AI applications in higher education. Section 4 develops a multidimensional understanding of women's empowerment and proposes a conceptual model linking AI adoption with educational and socio-economic outcomes. Section 5 examines the ethical challenges associated with AI adoption. Section 6 discusses the principles of human-centred AI, while Section 7 presents the theoretical foundations underpinning the framework. Section 8 critically reviews existing international and national AI governance frameworks. Section 9 introduces the proposed Human-Centred Ethical AI Framework, followed by a phased implementation roadmap in Section 10. Section 11 discusses policy implications, Section 12 outlines future research directions and synthesises the framework's contribution to existing

AI governance models, Section 13 discusses the paper's theoretical, social, gender, policy, educational, SDG-related, and global implications, and Section 14 concludes the paper.

2. The Saudi Context

2.1 Women's Higher Education in Saudi Arabia

Women's access to formal higher education in Saudi Arabia is a comparatively recent phenomenon, dating in institutional terms to the establishment of the first women's college in Riyadh in the early 1970s, decades after the founding of comparable institutions for men. For much of the subsequent period, women's higher education developed along a segregated track, with a narrower range of permitted fields of study, more limited access to laboratory and technical facilities, and a curriculum weighted towards education, humanities, and health sciences. Over the past two decades, however, this pattern has shifted substantially. Female enrolment in Saudi higher education has expanded to the point where women now constitute a majority or near-majority of students at several public universities, and the gender gap in tertiary gross enrolment that persisted until the mid-2010s has since closed (UNESCO, 2026), while female participation in postgraduate study, including doctoral programmes, has grown markedly. Participation in STEM fields, while still lower than in humanities and social sciences, has also increased, assisted by targeted scholarship schemes, the opening of previously restricted programmes to women, and the gradual relaxation of physical segregation requirements on many campuses, although regional evidence suggests that rising STEM enrolment has not yet translated into proportionate representation in the STEM workforce (EL-Deghaidy et al., 2025).

This educational expansion has occurred alongside, and has been reinforced by, a broader set of legal and social reforms affecting women's status, including changes to guardianship requirements, the right to drive, and expanded eligibility for a wider range of professions. These reforms are not incidental to the higher-education story; they form part of the same policy architecture, and they have altered the practical calculus facing female graduates as they consider how to translate educational credentials into employment, entrepreneurship, or further study. It is against this backdrop of rapid but still uneven and unevenly distributed change that the question of AI adoption in Saudi higher education must be situated. Female students entering Saudi universities today are simultaneously beneficiaries of one of the most rapid expansions of women's educational access in the region's recent history, and among the first cohorts to encounter AI-enabled learning technologies as a routine part of their educational experience. How these two trends interact is not yet well understood, and this paper argues that it deserves sustained scholarly attention.

2.2 Vision 2030 and the Digital-Empowerment Nexus

Vision 2030 provides the overarching policy architecture within which both AI adoption and women's empowerment in Saudi higher education must be understood. The programme is organised around three principal themes: a vibrant society, a thriving economy, and an ambitious nation, and it is supported by a set of delivery vehicles known as Vision Realization Programs. Two of these programmes are of

particular relevance to the present paper. The Human Capability Development Program is explicitly concerned with building the skills, capabilities, and labour-market readiness of the Saudi population, with particular attention to youth and to women, and it has served as the principal vehicle through which digital-skills and AI-literacy initiatives have been extended into the education sector. The Digital Government Strategy, alongside the National Strategy for Data and Artificial Intelligence coordinated by SDAIA, sets out the Kingdom's ambitions to become a global leader in data and AI, with education repeatedly identified as one of the priority sectors for AI adoption.

SDAIA occupies a distinctive institutional position in this landscape. Established in 2019 and reporting directly to the highest levels of government, it combines regulatory, strategic, and operational functions relating to data governance and AI, and it has increasingly taken on a normative role in articulating ethical principles for AI use within the Kingdom, including through its published AI ethics principles. For Saudi universities, SDAIA functions simultaneously as a source of strategic direction, a potential regulator, and a partner in specific digital-transformation initiatives. Any framework proposing responsible AI adoption in Saudi higher education must therefore take SDAIA's evolving role explicitly into account, both as a source of authority and as a potential site of institutional collaboration.

The empowerment dimension of Vision 2030 is most explicitly articulated through targets for women's labour-force participation, entrepreneurship, and representation in leadership positions, targets that have in several respects already been met or exceeded ahead of schedule, prompting subsequent upward revision. Universities are positioned within this architecture not merely as sites of credentialing but as active instruments of workforce preparation, expected to align curricula, skills development, and increasingly AI literacy with the economy that Vision 2030 envisages. This creates a specific and somewhat under-examined obligation: if universities are to be held accountable for preparing women for an AI-enabled economy, the manner in which AI is used within the university itself, and the extent to which that use is designed with women's specific circumstances, constraints, and aspirations in mind, becomes a matter of direct policy relevance rather than a peripheral technical concern.

3. Artificial Intelligence in Higher Education

The application of AI within higher education can usefully be organised into several overlapping categories. Generative AI systems, capable of producing text, code, and other content in response to natural-language prompts, have rapidly become embedded in student writing, research, and study practices, as well as in faculty course design and assessment preparation. Learning analytics platforms draw on the digital traces students generate as they interact with learning management systems to identify patterns associated with engagement, risk of attrition, and academic performance, typically presenting this information through dashboards intended to support timely intervention by advisors and instructors. Intelligent tutoring systems and adaptive learning platforms adjust the sequencing, difficulty, and presentation of instructional content in response to individual learner performance, with the aim of personalising the pace and pathway of learning. AI-based assessment tools extend from automated grading of structured responses to more sophisticated natural-language processing applications

capable of providing formative feedback on extended written work. AI-enabled academic advising and virtual assistant systems provide students with round-the-clock access to information regarding degree requirements, course selection, and administrative processes, reducing dependence on the availability of human advisors. Predictive analytics and recommendation systems extend these capabilities further, suggesting courses, study resources, or career pathways based on patterns observed across large student populations. Finally, a growing set of applications is oriented specifically towards accessibility and inclusion, including AI-enabled captioning, translation, and adaptive interfaces designed to support students with disabilities or with limited proficiency in the language of instruction.

Several features of this technological landscape carry particular significance for female students within the Saudi context. Flexibility is one of the most consistently cited benefits of AI-enabled learning technologies, and it interacts directly with the practical constraints that continue to shape many Saudi women's educational trajectories, including domestic responsibilities, transportation limitations in some regions, and, for some students, continued preference for gender-segregated or partially online modes of study. Adaptive and asynchronous learning tools, virtual advising systems, and AI-enabled distance-learning platforms can, in principle, extend access to female students in more conservative or geographically remote areas who might otherwise face greater difficulty attending in-person instruction, and can support continued study during periods such as early motherhood. AI-enabled accessibility tools carry similar significance for the smaller but important population of female students with disabilities, whose educational access has historically received less attention within Saudi higher-education policy.

At the same time, the same technologies that offer these benefits also introduce risks that fall disproportionately, or in gender-specific ways, on female students, and these risks are examined in detail in Section 5. Learning analytics systems that track engagement patterns, physical location, or time-of-day usage may interact uncomfortably with cultural expectations regarding female students' privacy and mobility. Generative AI and automated assessment tools trained predominantly on non-Arabic, non-Gulf datasets may perform less reliably for Arabic-language content or may embed assumptions about student behaviour, career aspiration, or family structure that do not reflect the lived realities of Saudi women. The remainder of this paper is concerned precisely with how these benefits can be maximised and these risks minimised through deliberate, human-centred design and governance choices, rather than treating either the benefits or the risks as inherent to the technology itself.

Table 1

summarises the principal categories of AI application in higher education discussed above, together with their particular relevance for female students.

AI Application Category	Description	Relevance for Female Students
Generative AI	Produces text, code and other content from natural-language prompts	Supports flexible, self-paced writing, study and research assistance
Learning Analytics	Tracks engagement and interaction data to flag risk and inform advising	Requires careful handling given heightened privacy sensitivities
Intelligent Tutoring / Adaptive Learning	Adjusts sequencing and difficulty to individual performance	Personalises pace around domestic and other practical constraints
AI-Based Assessment	Automated grading and feedback on structured and written work	Reduces dependence on the timing of face-to-face review
AI Advising / Virtual Assistants	Round-the-clock information on requirements, courses and processes	Reduces dependence on the availability of human advisors
Predictive Analytics & Recommenders	Suggests courses, resources and career pathways from population patterns	Risk of embedding historical gender patterns in recommendations
Accessibility Tools	Captioning, translation and adaptive interfaces	Supports students with disabilities and limited language proficiency

Table 1. **Categories of AI Applications in Higher Education and Their Relevance for Female Students**

4. Women's Empowerment Through AI

4.1 Dimensions of Empowerment

Empowerment is a multidimensional construct, and the literature on women's empowerment more broadly has long resisted collapsing it into a single indicator such as labour-force participation or educational attainment. For the purposes of this paper, empowerment is understood to encompass academic empowerment, meaning improved academic performance, persistence, and self-directed learning capacity; career empowerment, meaning improved employability, career clarity, and access to professional networks; leadership development, meaning growth in the confidence and skills associated with taking on leadership roles within academic, professional, or community settings; entrepreneurial capacity, meaning the skills and confidence associated with founding and sustaining new ventures; participation in STEM fields, understood both as enrolment and as sustained engagement; research productivity, particularly for women pursuing postgraduate study and academic careers; expanded access to flexible learning, as discussed in Section 3; digital and AI literacy, understood as a foundational competency for participation in an AI-enabled economy; the development of a secure

professional identity; academic and professional self-efficacy; confidence in independent decision-making; orientation towards lifelong learning; general employment readiness; and digital inclusion in the broadest sense, meaning meaningful rather than merely nominal access to digital tools and the skills required to use them effectively.

This multidimensional understanding of empowerment matters because it resists the temptation to treat AI adoption as automatically or uniformly empowering simply because it expands access to educational content or administrative convenience. A learning-analytics system that improves retention rates but does so by increasing a sense of surveillance and reducing students' confidence in their own autonomous judgement would, on this understanding, represent a mixed rather than a straightforwardly positive empowerment outcome. The framework developed later in this paper is designed precisely to hold these different dimensions in view simultaneously, rather than allowing gains on one dimension, such as academic performance, to stand as a proxy for empowerment as a whole.

4.2 A Conceptual Model Linking AI Adoption to Empowerment

Building on this multidimensional understanding, the paper proposes a conceptual model in which AI adoption is linked to empowerment through a sequential, though not strictly linear, chain of intermediate outcomes. Appropriately designed and governed AI adoption first supports the development of digital skills, both in the specific sense of familiarity with the AI tools themselves and in the broader sense of general digital literacy. These digital skills, in turn, are associated with improved learning engagement, as students are able to interact with course material in more personalised, responsive, and accessible ways. Improved engagement plausibly contributes to academic success, understood in terms of both performance and persistence to completion. Academic success, combined with the digital and AI literacy developed along the way, contributes to employability, as graduates enter a labour market in which familiarity with AI-enabled tools and workflows is increasingly treated as a baseline expectation across sectors. Employability, realised through actual employment, self-employment, or further study, contributes to economic participation, understood in the specific Saudi context as a key national policy objective in its own right. Economic participation, finally, contributes to the broader outcome of women's empowerment, understood in the multidimensional sense outlined above, encompassing not only economic standing but also leadership, confidence, and voice.

This chain is presented as a heuristic rather than as a strictly deterministic causal sequence, and the paper explicitly acknowledges that feedback loops operate throughout it. Confidence and self-efficacy gained through early engagement with AI-enabled learning tools plausibly reinforce further digital-skill development; leadership opportunities gained through economic participation plausibly reshape subsequent educational choices, including decisions to pursue further study. The value of the model lies not in claiming to have identified a single causal pathway, but in providing university leaders and policymakers with a structured way of tracing how a specific AI investment, such as an adaptive learning platform or an AI-enabled advising system, might be expected to contribute to empowerment outcomes,

and where along the chain that contribution might plausibly break down if the technology is poorly designed, inequitably distributed, or implemented without adequate attention to the ethical and cultural considerations discussed in the following sections.

4.3 Empirical Grounding: Evidence from Saudi Women's Leadership Pathways

The conceptual model set out above finds direct empirical corroboration in a related life history study of twelve female secondary school head teachers in Saudi Arabia (Alajaji & Phiri, forthcoming), which traced participants' leadership trajectories biographically from childhood through higher education, early career, and headship. That study's central finding is that leadership capability among Saudi women developed overwhelmingly through cumulative informal experience, including early family responsibility, mentoring relationships, and workplace challenge, rather than through formal, systematic leadership preparation. Routes into headship were similarly informal in the majority of cases, arising through nomination by a supervisor or predecessor or through a serendipitous vacancy, rather than through structured succession planning. The study's concept of developmental ecologies, denoting the relational and institutional resources, including mentoring, peer networks, and access to incremental responsibility, available to an aspiring leader at a given career stage, offers a close empirical parallel to the digital-skills and engagement stages of the conceptual model presented in Section 4.2, while extending it beyond the purely technological.

This evidence carries three implications for the present framework. First, it corroborates the paper's caution, developed in Section 4.1, against treating exposure to AI-enabled tools as automatically empowering: the life history evidence indicates that relational and institutional conditions, not individual skill acquisition alone, are decisive for whether capability translates into advancement. Second, it suggests a concrete institutional application for human-centred AI within this framework: AI-enabled mentoring platforms, peer-networking tools, and criteria-based leadership-identification systems could function as a deliberate, technologically supported extension of the developmental ecologies identified as critical in the empirical study, of particular value to women whose informal networks are comparatively weak. Third, and more cautionary, the empirical study also found that some participants' routes to headship were shaped by extensive personal and community networks, a pattern with implications for the fairness principle discussed in Section 5.4: AI-enabled succession-planning tools hold genuine promise of introducing transparent, criteria-based alternatives to network-mediated advancement, but only if the criteria and data on which they are built are not themselves derived from, and do not simply reproduce, the same informal, network-based patterns they are intended to correct. Table 2 below summarises these empirical themes and their implications for the framework developed in this paper.

Figure 1 presents the conceptual model, introduced in Section 4.2 above, linking AI adoption to the empowerment outcomes examined in this section.

Table 2
Empirical Evidence on Saudi Women's Leadership Pathways (Alajaji & Phiri, forthcoming) and
Implications for the Framework

Empirical Theme	Key Finding (12 Saudi Female Head Teachers)	Implication for Human-Centred AI Framework
Route to headship	Most participants entered headship through informal nomination by a supervisor or predecessor, or through a serendipitous vacancy, rather than structured succession planning	Supports the case for transparent, criteria-based AI-enabled leadership-identification and succession-planning tools
Formal leadership preparation	Formal leadership preparation programmes were largely absent prior to appointment	Strengthens the roadmap's emphasis (Phases 3–4) on structured, technology-supported development ahead of appointment
Developmental ecologies	Leadership confidence and smoother transitions were associated with rich mentoring, peer networks and incremental responsibility	AI-enabled mentoring and peer-networking platforms could extend developmental ecologies to women with weaker informal networks
Reliance on informal networks	Some participants' routes to appointment were shaped by extensive personal or community networks	Underscores the fairness risk (Section 5.4) that AI tools must be designed to widen, not merely digitise, access to advancement
Continuous development post-appointment	Leadership development continued after appointment through informal peer learning, in the absence of formal provision	Supports Phase 6 (monitoring and continuous evaluation) and AI-enabled continuing professional development

5. Ethical Challenges

The realisation of the empowerment potential described in Section 4 is not automatic, and AI systems are well documented to carry ethical risks capable of undermining rather than advancing the outcomes described above. This section organises these risks around six interrelated categories: bias, transparency, privacy, fairness, accountability, and cultural considerations.

5.1 Bias

Algorithmic bias arises when AI systems produce systematically different outcomes for different groups of users, typically as a consequence of imbalances or historical inequities embedded in the data on which the system was trained. Within an educational context, this can manifest as gender bias in automated assessment or recommendation systems that have learned patterns from historical data reflecting a period in which women's participation in certain fields or roles was more restricted than it is today; a recent systematic review of gender bias in educational AI applications confirms that such biases recur across content generation, assessment, and feedback systems internationally, while also

finding that research on the topic remains heavily concentrated in North American and Western settings, with the Gulf region largely absent from the evidence base this paper seeks to help address (Kalim, Kanwar, Xu, et al., 2026). Language bias presents a related and, in the Saudi context, particularly salient concern, given that many widely used generative AI and natural-language-processing systems have been trained predominantly on English-language and Western data sources, and may perform less reliably, or embed subtly different assumptions, when applied to Arabic-language content or to culturally specific forms of expression. Dataset imbalance, more generally, refers to the underlying problem that gives rise to both of these more specific forms of bias: training data that underrepresents the populations, languages, or contexts in which a system is subsequently deployed.

5.2 Transparency

A second cluster of concerns relates to the transparency and explainability of AI systems. Many of the most capable contemporary AI systems, including large language models and complex machine-learning classifiers used in learning-analytics and predictive-advising applications, operate as effective black boxes, in the sense that the specific reasoning behind any individual output or recommendation is not readily interpretable even to technically sophisticated users, let alone to students or instructors without a technical background. This opacity creates practical difficulties when an AI system flags a student as at risk of attrition, recommends against a particular course of study, or scores a piece of written work in a way the student or instructor finds surprising, since neither party may be well placed to understand, contest, or learn from the underlying reasoning. Explainable AI approaches, discussed further in Section 6, represent one response to this concern, though they do not eliminate it entirely.

5.3 Privacy

Privacy concerns arise from the extensive data collection that underpins most AI-enabled educational tools, including learning analytics platforms that track granular patterns of student interaction with course materials, virtual assistants that log the content of student queries, and adaptive systems that build detailed longitudinal profiles of individual learning behaviour. These concerns take on a particular cultural inflection within the Saudi context. Data that might appear relatively benign in other settings, such as time-of-day usage patterns, physical location at the point of access, or patterns of interaction that could reveal information about a student's domestic circumstances, may carry heightened sensitivity for Saudi women, both because of continued cultural expectations regarding privacy and mobility, and because of the potential, however unlikely in any specific instance, for such data to be accessed by family members or others in ways the student has not consented to. Universities adopting AI-enabled tracking and analytics tools therefore face a distinctive obligation to consider not only standard data-protection questions of consent, security, and retention, but also the specific privacy expectations of their female student population, and to ensure that data ownership and access arrangements are transparent and genuinely under the student's own control wherever possible.

5.4 Fairness

Fairness concerns extend beyond bias in system outputs to encompass equitable access to the AI tools themselves. Female students in more conservative or resource-constrained regions of the Kingdom may have less reliable access to the devices, connectivity, or digital infrastructure required to benefit fully from AI-enabled learning platforms, creating a risk that AI adoption widens rather than narrows existing regional and socio-economic disparities among women. Algorithmic justice, understood as the requirement that the benefits and burdens of algorithmic decision-making be distributed equitably, requires universities to monitor not only whether their AI systems perform accurately on average, but whether that accuracy, and the practical benefits that flow from it, are distributed evenly across their student population. A related and, in this context, empirically grounded fairness concern arises further along the pipeline from study into leadership. As the life history evidence discussed in Section 4.3 indicates, advancement into leadership positions within Saudi education has in practice often been mediated by informal personal and community networks rather than transparent, criteria-based processes. AI-enabled leadership-identification and succession-planning tools are frequently promoted on the grounds that they can replace such network-mediated advancement with more transparent, criteria-based alternatives; this framework treats that promise as conditional rather than automatic, since a system trained on historical advancement data risks simply learning, and thereby further entrenching, the same informal patterns it is intended to correct.

5.5 Accountability

Accountability concerns the question of who bears responsibility when an AI system produces a harmful, unfair, or simply mistaken outcome. Within a university setting, potential loci of accountability include the university itself as the institution deploying the system, the commercial or academic developer of the underlying technology, relevant government bodies including SDAIA in its regulatory capacity, and individual faculty members who may rely on AI-generated recommendations in their own decision-making. The diffusion of responsibility across this chain of actors creates a well-recognised risk that no single party takes clear ownership of outcomes, a risk that is only heightened when, as is often the case, the underlying AI system has been developed and trained outside the Kingdom by a third-party vendor with limited visibility into, or responsibility for, the specific context of its deployment in a Saudi university.

5.6 Cultural Considerations

Finally, the ethical challenges surrounding AI adoption in Saudi higher education cannot be fully understood through frameworks developed in other cultural and religious contexts without adaptation. Religious and cultural values shape expectations regarding gender interaction, privacy, modesty, and family involvement in ways that carry direct implications for how AI-enabled educational tools should be designed and deployed. A virtual advising assistant, for example, that is designed without attention to culturally appropriate modes of address, or a proctoring system that requires camera access in ways that raise concerns for students who observe modest dress or gender-segregation norms within their home environment, illustrates the kind of localisation failure that generic, internationally developed AI systems are prone to. Gender-sensitive AI design, understood as the deliberate incorporation of these cultural and religious considerations into system design from the outset, rather than as an afterthought,

is accordingly identified in this paper as a distinct and necessary component of responsible AI adoption in this context, addressed further in Sections 6 and 9.

Table 3
summarises these ethical challenges and their illustrative manifestations within the Saudi higher-education context.

Category	Core Concern	Illustrative Manifestation in this Context
Bias	Systematically different outcomes across groups	Gender and language bias in systems trained on non-Arabic, non-Gulf data
Transparency	Opacity of complex AI reasoning	Unexplained attrition-risk flags or automated grading decisions
Privacy	Extensive data collection and behavioural profiling	Heightened sensitivity of location and time-of-use data for Saudi women
Fairness	Unequal access to, and performance of, AI tools	Regional and infrastructure disparities among female students
Accountability	Diffusion of responsibility across actors	Unclear ownership between university, vendor and regulator
Cultural Considerations	Misalignment with religious and cultural norms	Poorly localised advising tools or camera-based proctoring requirements

Table 3. **Summary of Ethical Challenges in AI Adoption for Women’s Higher Education**

6. Human-Centred AI

The ethical challenges identified in Section 5 point towards the need for a design and governance paradigm capable of addressing them systematically, rather than through piecemeal, after-the-fact fixes. This paper adopts human-centred AI as that paradigm. Human-centred AI can be understood as an approach to the design, development, and deployment of AI systems that places human wellbeing, agency, and values at the centre of the technology's purpose, rather than treating human users as passive recipients of system outputs optimised primarily for technical performance metrics. Closely related concepts include responsible AI, which emphasises accountability and governance across the AI lifecycle; explainable AI, which emphasises the interpretability of system outputs; trustworthy AI, which emphasises the conditions under which users can justifiably place confidence in a system's outputs; and value-sensitive design, which emphasises the deliberate and systematic incorporation of stakeholder values into the technical design process from the earliest stages of development.

Human-AI collaboration and human oversight represent two further principles within this paradigm that carry particular relevance for the educational context under discussion. Rather than treating AI systems as autonomous decision-makers, a human-centred approach positions AI outputs as inputs to decisions that remain substantively under the control of human actors, whether that be an academic advisor reviewing an AI-generated risk flag before deciding how to engage with a student, or a student retaining

full discretion over whether to act on an AI-generated course recommendation. Inclusive and participatory design, meanwhile, requires that the populations affected by an AI system, in this case Saudi female students and the faculty who work with them, are directly involved in the design and evaluation of the systems that will be deployed, rather than being treated solely as end users of technology designed elsewhere. Digital wellbeing, finally, requires attention to the psychological and social effects of sustained AI-mediated interaction, including the risk that heavy reliance on AI-enabled advising or tutoring tools could, if poorly designed, erode rather than build the independent confidence and self-efficacy that Section 4 identified as core empowerment outcomes.

Taken together, these considerations consolidate into eight principles that form the ethical core of the framework developed in Section 9. Transparency, fairness, privacy, and accountability carry forward the concerns already established in Sections 5.2 through 5.5, now recast as positive design criteria rather than risks to be mitigated. Trust and human autonomy extend the trustworthy-AI and human-oversight concepts introduced above: confidence earned through demonstrated reliability rather than assumed by design, and AI outputs that inform rather than replace human judgement. Two further principles are introduced here for the first time: accessibility, meaning that AI-enabled tools are genuinely usable across the full range of technical proficiency, language preference, and physical ability; and inclusiveness, meaning that the benefits of AI adoption reach the full diversity of the female student population rather than concentrating among the most digitally privileged.

7. Theoretical Foundations

Rather than relying on a single theoretical lens, this paper integrates several established theoretical traditions, on the view that the empowerment question at its centre is irreducibly multidimensional and cannot be adequately addressed through any one theory alone.

The Capability Approach, developed principally by Amartya Sen, provides the paper's foundational normative orientation. Rather than evaluating development or empowerment purely in terms of resources or outcomes such as income or educational attainment, the Capability Approach directs attention to the substantive freedoms, or capabilities, that individuals actually have to lead the kind of life they have reason to value. Applied to the present context, this reorients the evaluation of AI adoption away from simple usage metrics, such as how many students interact with a given tool, towards the question of whether AI adoption genuinely expands the range of valuable choices and opportunities available to female students, including the choice to pursue particular fields of study, careers, or forms of civic and economic participation.

Human-Centred AI and Value Sensitive Design, introduced in Section 6, supply the paper's design-level theoretical resources, directing attention to how the specific technical and interface choices made in building an AI system either support or undermine the substantive capabilities identified through the Capability Approach.

Established technology-acceptance theories provide a complementary, more proximate account of how and why individual students and faculty come to adopt, or resist adopting, AI-enabled tools in practice. The Technology Acceptance Model directs attention to perceived usefulness and perceived ease of use as the principal proximate determinants of technology adoption, while the Unified Theory of Acceptance and Use of Technology extends this account to incorporate performance expectancy, effort expectancy, social influence, and facilitating conditions, the last two of which carry particular salience in the Saudi context, given the continued influence of family and social expectations on women's educational and career choices, and given the uneven distribution of the digital infrastructure that constitutes a facilitating condition for AI use.

Self-Determination Theory contributes a psychological account of the conditions under which human motivation flourishes, identifying autonomy, competence, and relatedness as basic psychological needs whose satisfaction is associated with more durable, intrinsically motivated engagement, as opposed to engagement sustained only through external pressure or reward. This theory offers a valuable diagnostic lens for evaluating whether a given AI-enabled learning tool is likely to support the kind of durable, self-directed engagement associated with genuine empowerment, or whether it risks undermining student autonomy through excessive surveillance or overly prescriptive personalisation.

Social Cognitive Theory, associated principally with Albert Bandura, contributes the closely related construct of self-efficacy, understood as an individual's belief in their own capacity to execute the actions required to achieve a given outcome. Since several of the empowerment dimensions identified in Section 4, including leadership development and entrepreneurial capacity, are substantially mediated by self-efficacy, this theory provides a further important lens through which to evaluate the likely empowerment effects of a given AI intervention.

Digital Inclusion Theory, finally, directs attention to the conditions required for digital access to translate into meaningful participation, extending beyond simple availability of devices and connectivity to encompass digital literacy, the cultural and linguistic appropriateness of digital content, and the broader social conditions that determine whether formally available digital resources are, in practice, equally accessible to all members of a given population. Empowerment Theory, understood in the broad sense developed within community psychology and social-work scholarship, supplies the paper's overarching outcome construct, treating empowerment as a multilevel process operating simultaneously at the individual, organisational, and community levels, consistent with the multidimensional treatment of empowerment adopted in Section 4.

Integrated together, these theoretical traditions yield a conceptual lattice in which the Capability Approach specifies the normative goal, Human-Centred AI and Value Sensitive Design specify the design conditions required to pursue that goal responsibly, technology-acceptance theories specify the proximate behavioural conditions required for adoption to occur at all, Self-Determination Theory and Social Cognitive Theory specify the psychological conditions required for that adoption to translate into durable engagement and confidence, and Digital Inclusion Theory and Empowerment Theory specify the

structural conditions required for individual-level engagement to aggregate into meaningful, equitably distributed empowerment outcomes. This integrated lattice underpins the conceptual framework presented in Section 9.

Table 4

summarises the theoretical foundations underpinning the proposed framework and the specific contribution each makes.

Theory	Key Construct	Contribution to the Framework
Capability Approach (Sen)	Substantive freedoms / capabilities	Normative goal: expand real opportunities, not merely usage
Human-Centred AI / Value Sensitive Design	Human wellbeing and values in design	Design-level criteria for evaluating AI tools
Technology Acceptance Model	Perceived usefulness and ease of use	Proximate determinants of individual adoption
UTAUT	Performance/effort expectancy, social influence, facilitating conditions	Accounts for social and infrastructural adoption barriers
Self-Determination Theory	Autonomy, competence, relatedness	Diagnostic lens for durable, intrinsically motivated engagement
Social Cognitive Theory (Bandura)	Self-efficacy	Explains leadership and entrepreneurial empowerment gains
Digital Inclusion Theory	Meaningful access beyond mere availability	Structural conditions for equitable participation
Empowerment Theory	Multilevel empowerment process	Overarching outcome construct across individual, organisational and community levels

Table 4. Theoretical Foundations Underpinning the Proposed Framework

8. Existing AI Governance Frameworks

A number of influential international instruments have sought to articulate general principles for the responsible governance of AI, and any framework proposed for Saudi higher education must be positioned in relation to these instruments rather than developed in isolation from them. The UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted by member states in 2021, represents the first global normative instrument on AI ethics, articulating values including human rights and dignity, environmental flourishing, and diversity and inclusiveness, alongside policy areas spanning data governance, education, and gender. The OECD AI Principles, first adopted in 2019 and subsequently endorsed by a wide range of countries beyond the OECD's own membership, articulate a set of values-based principles including inclusive growth, human-centred values and fairness, transparency and explainability, robustness and safety, and accountability. The European Union's AI Act establishes a

binding, risk-tiered regulatory regime, classifying AI applications by risk level and imposing correspondingly graduated obligations, with several categories of educational AI application, including systems used to evaluate learning outcomes or determine access to education, treated as high-risk and subject to stringent requirements. The United States National Institute of Standards and Technology has developed an AI Risk Management Framework structured around the functions of governing, mapping, measuring, and managing AI-related risk, intended as a voluntary but influential reference point for organisations seeking to operationalise responsible AI practice. ISO/IEC 42001, published in 2023, provides the first international management-system standard specifically for AI, offering organisations a certifiable framework for establishing, implementing, and continually improving an AI management system.

Within Saudi Arabia itself, SDAIA has published a set of national AI ethics principles intended to guide the responsible development and use of AI within the Kingdom, situated within the broader National Strategy for Data and Artificial Intelligence. These principles draw substantially on the international instruments described above while being framed within the specific context of Saudi national priorities, including explicit reference to alignment with Vision 2030 and with Islamic values.

These frameworks share a number of common elements, including an emphasis on transparency, fairness, accountability, privacy, and human oversight, and they collectively provide an indispensable foundation for any organisation, including a Saudi university, seeking to govern its use of AI responsibly. However, they share a common limitation for the purposes of this paper: each was developed to apply broadly, across sectors, populations, and national contexts, and none was designed with the specific circumstances of women's higher-education empowerment within a Gulf cultural and religious setting in view. This is not a criticism of these frameworks so much as an observation about the nature of general governance instruments, which necessarily trade specificity for breadth of applicability. It follows that Saudi universities cannot simply adopt one of these frameworks unmodified and expect it to adequately address the gender-specific and culturally specific considerations identified in Section 5. What is required, instead, is a framework that takes the common principles articulated across these instruments as a foundation, while adapting and extending them to the specific institutional mission of advancing women's empowerment through higher education within the Saudi cultural and religious context. Developing precisely such a framework is the task of the following section.

Figure 3 presents a policy alignment map illustrating this convergence of national and international governance instruments upon Saudi higher-education institutions.

Table 5

sets out these international and national governance frameworks in comparative summary form.

Framework	Issuing Body	Year	Core Focus
Recommendation on the Ethics of AI	UNESCO	2021	Global normative values: human rights, dignity, diversity and inclusiveness
AI Principles	OECD	2019	Inclusive growth, human-centred values, transparency, accountability
AI Act	European Union	2024	Binding, risk-tiered regulatory regime
AI Risk Management Framework	NIST (United States)	2023	Voluntary govern–map–measure–manage risk structure
ISO/IEC 42001	ISO	2023	Certifiable AI management-system standard
National AI Ethics Principles	SDAIA (Saudi Arabia)	—	National alignment with Vision 2030 and Islamic values

Table 5. **Comparison of International and National AI Governance Frameworks**

9. Proposed Framework: A Human-Centred Ethical AI Architecture for Women's Empowerment

9.1 Overview

Building on the theoretical synthesis presented in Section 7 and the governance foundation reviewed in Section 8, this section presents the paper's central conceptual contribution: a multi-layered Human-Centred Ethical AI Framework linking national policy, institutional readiness, ethical AI principles, AI technologies, educational outcomes, and women's empowerment outcomes within a single architecture. The framework is deliberately structured as a layered model rather than a simple linear causal chain, in recognition of the fact that each layer both shapes and is shaped by the layers around it, and that responsible AI adoption requires coordinated attention across all layers simultaneously rather than sequential progress through them in isolation.

9.2 The Six Layers

The first layer, the National Policy Environment, comprises Vision 2030, the National Strategy for Data and Artificial Intelligence, and Saudi higher-education strategy more broadly. This layer establishes the overarching mandate and resourcing environment within which individual universities operate, and it is the layer through which SDAIA's national AI ethics principles and regulatory expectations are principally transmitted into the higher-education sector.

The second layer, Institutional Readiness, concerns the specific capacities that a given university brings to AI adoption, including leadership commitment, technical infrastructure, faculty capability and confidence in working with AI-enabled tools, and internal governance arrangements, including the presence or absence of a clear institutional AI ethics or oversight body. Variation across Saudi universities in institutional readiness is likely to be substantial, and the framework treats this layer as a critical determinant of whether the ethical principles articulated in national policy are translated into practice at the point of actual technology deployment.

The third layer, Human-Centred AI Principles, operationalises the eight principles identified in Section 6, namely transparency, fairness, privacy, trust, human autonomy, accessibility, inclusiveness, and accountability, as the specific design and governance criteria against which any AI technology under consideration for deployment should be evaluated before adoption, and monitored against after adoption.

The fourth layer, AI Technologies, comprises the specific categories of application reviewed in Section 3, including generative AI, learning analytics, adaptive learning platforms, virtual tutors and advising assistants, intelligent feedback systems, and AI-enabled accessibility tools. This layer is deliberately positioned after, rather than before, the ethical-principles layer in the framework's architecture, in order to emphasise the paper's core argument that specific technology choices should be evaluated against, and where necessary constrained by, the human-centred principles established in the preceding layer, rather than the reverse.

The fifth layer, Educational Outcomes, comprises the proximate outcomes reviewed in Section 4.2, including academic success, digital-skills development, research-skills development, STEM participation, self-efficacy, and engagement, that mediate between AI adoption and the ultimate empowerment outcomes of interest.

The sixth and final layer, Women's Empowerment, comprises the multidimensional outcomes set out in Section 4.1: leadership, employment, entrepreneurship, innovation, lifelong learning, and economic participation, understood collectively as constituting empowerment in the substantive, capability-expanding sense articulated through the paper's engagement with the Capability Approach in Section 7.

9.3 Interpreting the Framework

The framework is intended to be read both vertically, as a layered causal architecture running from national policy through to individual empowerment outcomes, and horizontally, as a set of feedback relationships in which outcomes observed at the empowerment layer should inform ongoing adjustment at each of the preceding layers. A university observing, for example, that a given AI-enabled advising system is associated with strong academic outcomes but weak subsequent gains in students' independent decision-making confidence, would, on this framework, be prompted to revisit the human-centred principles layer, and specifically the human-autonomy principle, rather than treating the academic-outcome success as sufficient evidence of the system's overall value.

This bidirectional reading is central to the framework's claim to originality relative to the largely unidirectional models found elsewhere in the literature on AI in education, and it reflects the paper's broader argument, developed across Sections 4 through 8, that empowerment cannot be treated as an automatic downstream consequence of AI adoption, but must instead be actively designed for, monitored, and iteratively pursued through deliberate attention to the ethical and institutional conditions under which AI technologies are deployed. A conceptual figure illustrating the six layers and their principal interconnections is proposed as a core visual element of the manuscript, together with two supporting figures: a policy-alignment map showing how Vision 2030, the National Strategy for Data and Artificial Intelligence, and the international governance frameworks reviewed in Section 8 converge upon Saudi higher-education institutions, and an implementation roadmap, presented in the following section, illustrating the staged institutional adoption process through which the framework is intended to be operationalised in practice.

Figure 2 presents the proposed Human-Centred Ethical AI Framework, illustrating the six layers described above and their principal interconnections.

Table 6
sets out the six layers of the proposed framework, described in turn above, in summary form.

Layer	Core Components
1. National Policy Environment	Vision 2030; National Strategy for Data & Artificial Intelligence; higher-education strategy
2. Institutional Readiness	Leadership commitment; technical infrastructure; faculty capability; internal governance
3. Human-Centred AI Principles	Transparency, fairness, privacy, trust, human autonomy, accessibility, inclusiveness, accountability
4. AI Technologies	Generative AI; learning analytics; adaptive learning; virtual advising; intelligent feedback; accessibility tools
5. Educational Outcomes	Academic success; digital and research skills; STEM participation; self-efficacy; engagement
6. Women's Empowerment	Leadership; employment; entrepreneurship; innovation; lifelong learning; economic participation

Table 6. The Six Layers of the Proposed Human-Centred Ethical AI Framework

10. Implementation Roadmap

Translating the framework presented in Section 9 into institutional practice requires a phased approach, reflecting the reality that the capacities required for genuinely human-centred, ethically governed AI adoption are unlikely to be present in full at the outset in most Saudi universities, and are more realistically built incrementally over a period of years. The paper proposes a six-phase roadmap.

The first phase, AI readiness assessment, requires universities to conduct a systematic baseline assessment of their current technical infrastructure, faculty capability, existing AI-related initiatives, and student population characteristics, including the specific circumstances of their female student body identified in Section 2. The second phase, ethical governance, requires the establishment of institutional structures, such as a dedicated AI ethics committee or oversight body, with explicit responsibility for evaluating proposed AI deployments against the human-centred principles set out in Section 6, and with meaningful representation from female faculty, students, and, where relevant, external stakeholders with expertise in gender and cultural considerations. The third phase, faculty development, requires sustained investment in building faculty confidence and capability in working with AI-enabled tools, recognising that faculty attitudes and behaviours, per the technology-acceptance theories discussed in Section 7, exert substantial influence over student adoption and trust. The fourth phase, student AI literacy, requires the development of structured AI-literacy provision for students, extending beyond basic tool familiarity to encompass critical understanding of AI's capabilities, limitations, and ethical implications, consistent with the empowerment goals articulated in Section 4. The fifth phase, responsible AI deployment, involves the actual rollout of specific AI-enabled tools and systems, informed by the readiness assessment, governance structures, and literacy provision established in the preceding phases, and evaluated on a pilot basis against the framework's layered criteria before wider institutional scaling. The sixth and final phase, monitoring and continuous evaluation, establishes ongoing mechanisms for tracking outcomes across all six layers of the framework, including specifically the empowerment outcomes identified in Section 4, and for feeding the results of this monitoring back into governance and deployment decisions in the iterative, bidirectional manner described in Section 9.3.

This roadmap is presented as a general sequence rather than a rigid timetable, and universities are expected to vary in the pace at which they progress through it, and to revisit earlier phases as new AI technologies or governance requirements emerge. The roadmap's principal value lies in establishing governance and literacy capacity as necessary preconditions for responsible deployment, rather than treating deployment as a starting point to which governance and evaluation are subsequently, and often belatedly, appended.

Figure 4 presents the phased implementation roadmap through which the framework is intended to be operationalised in practice.

Table 7
sets out this phased implementation roadmap in summary form.

Phase	Focus	Key Activities
Phase 1: AI Readiness Assessment	Baseline capacity	Infrastructure audit; faculty capability review; student population profiling
Phase 2: Ethical Governance	Oversight structures	AI ethics committee; evaluation criteria; female faculty/student representation
Phase 3: Faculty Development	Capability building	Training; confidence-building; modelling of responsible AI use
Phase 4: Student AI Literacy	Critical literacy	Curriculum integration; awareness of capabilities and limitations
Phase 5: Responsible AI Deployment	Piloting and scaling	Pilot evaluation against framework criteria; phased institutional rollout
Phase 6: Monitoring & Continuous Evaluation	Feedback loop	Outcome tracking across all six layers; iterative adjustment

Table 7. **Phased Implementation Roadmap for Human-Centred AI Adoption**

11. Policy Implications

The framework and roadmap developed in the preceding sections carry distinct implications for each of the principal stakeholder groups involved in AI adoption within Saudi higher education. For Saudi universities, the central implication is the need to establish dedicated institutional governance capacity for AI, as described in the roadmap's second phase, and to ensure that this governance capacity includes explicit, structured attention to gender and cultural considerations rather than treating these as generic diversity concerns to be addressed through existing equality-and-inclusion mechanisms alone.

Universities should further be encouraged to adopt the framework's layered evaluation approach as a standard component of AI procurement and deployment decisions, requiring that any proposed AI tool be assessed against the human-centred principles set out in Section 6 prior to adoption. Given the empirical evidence discussed in Section 4.3 that Saudi women's advancement into leadership has historically relied on informal mentoring and network-based nomination rather than systematic succession planning, universities should further consider piloting AI-enabled mentoring and leadership-identification platforms, designed and governed in line with this framework's fairness and transparency principles, as a deliberate institutional complement to the informal developmental ecologies on which women leaders have so far had to depend.

For SDAIA, the implications concern the potential extension of the Authority's existing national AI ethics principles to include sector-specific guidance for higher education, addressing explicitly the gender and cultural considerations identified in this paper, and potentially establishing a certification or endorsement mechanism through which universities could demonstrate compliance with such guidance, drawing on the precedent set by international instruments such as ISO/IEC 42001. SDAIA is also well positioned to

convene a national community of practice among Saudi universities, facilitating the sharing of governance models, evaluation tools, and lessons learned as institutions progress through the phases of the implementation roadmap.

For the Ministry of Education, the implications concern the potential incorporation of AI-readiness and gender-sensitive AI governance criteria into existing institutional accreditation and quality-assurance processes, thereby providing a structural incentive for universities to invest in the governance and literacy capacities described in Sections 10 and beyond, and the potential development of national AI-literacy curricula standards to be implemented across the sector rather than left to individual institutional discretion.

For faculty, the implications concern the need to engage proactively with AI-literacy development, recognising the disproportionate influence that faculty attitudes and behaviours exert over student trust and adoption patterns, and the corresponding responsibility this creates for faculty to model responsible, human-centred use of AI-enabled tools within their own teaching and advising practice.

For students, and specifically for female students as the population of central concern in this paper, the implications concern the importance of developing not only technical familiarity with AI tools but also the critical literacy required to engage with these tools as an empowered user rather than a passive recipient, consistent with the emphasis on autonomy and self-efficacy developed in Section 7.

For industry and technology developers supplying AI-enabled educational tools to the Saudi market, the implications concern the need for genuine localisation, extending beyond superficial language translation to encompass the cultural, religious, and gender-specific design considerations discussed in Section 5.6, and the corresponding case for closer collaborative relationships between technology developers and Saudi universities during the design and pilot-testing phases of new AI-enabled educational tools, rather than the wholesale adoption of products designed primarily for other markets.

Table 8
summarises these policy recommendations by stakeholder.

Stakeholder	Principal Recommendation
Saudi Universities	Establish dedicated AI governance capacity; require layered ethical evaluation prior to procurement and deployment; pilot AI-enabled mentoring and leadership-identification platforms
SDAIA	Extend national AI ethics principles with higher-education-specific, gender-sensitive guidance; convene a national community of practice
Ministry of Education	Incorporate AI-readiness and gender-sensitive governance criteria into accreditation; develop national AI-literacy curricula standards
Faculty	Engage proactively with AI-literacy development; model responsible, human-centred use of AI tools
Students	Develop critical AI literacy as empowered users, not only technical familiarity
Industry & Technology Developers	Pursue genuine cultural and linguistic localisation; collaborate with universities during design and pilot testing

Table 8. **Summary of Policy Recommendations by Stakeholder**

12. Future Research Agenda

The conceptual framework developed in this paper is intended as a foundation for, rather than a substitute for, sustained empirical inquiry, and this section identifies several priority directions for future research. Longitudinal studies tracking the academic and career outcomes of female students exposed to varying degrees and forms of AI-enabled learning across their period of study and into early career would provide essential empirical grounding for the conceptual model presented in Section 4.2, allowing researchers to test the plausibility of the proposed chain linking AI adoption to empowerment outcomes, and to identify where in that chain the relationship is strongest or weakest in practice.

AI-literacy interventions specifically tailored to female students, evaluated through rigorous pre-post or comparative designs, would help establish which pedagogical approaches to building AI literacy are most effective in translating exposure to AI tools into the kind of critical, empowered engagement described in Sections 4 and 11, as opposed to superficial familiarity alone. Ethical evaluation of generative AI use in Saudi classrooms, including systematic assessment of the bias, language-performance, and cultural-appropriateness concerns identified in Section 5, would provide an essential evidence base for the kind of localisation efforts called for in Section 11.

Comparative studies extending the framework developed in this paper across other GCC higher-education systems, several of which share broadly similar policy architectures linking digital transformation and women's economic empowerment, would help establish the extent to which the framework's six layers and eight ethical principles generalise beyond the specific Saudi context, and

where meaningful national variation is likely to require further adaptation. Research attending specifically to the intersection of AI adoption with disability and other forms of underrepresentation among Saudi women, extending beyond the gender dimension emphasised in this paper, would help ensure that the inclusiveness principle articulated in Section 6 is examined with appropriate attention to intersectional variation within the female student population itself, rather than treating that population as internally homogeneous.

AI-supported entrepreneurship education represents a further priority area, given the specific salience of entrepreneurial capacity within the empowerment dimensions identified in Section 4.1 and within Vision 2030's broader economic objectives. Research developing and validating AI-governance maturity models specifically calibrated to the university context, building on the institutional-readiness layer of the framework presented in Section 9, would provide practical assessment instruments through which universities could benchmark their own progress through the implementation roadmap set out in Section 10. Finally, research examining trust and acceptance of AI among female academic staff, as distinct from female students, would extend the technology-acceptance considerations discussed in Section 7 to a population whose attitudes and behaviours, per the discussion in Section 11, exert disproportionate influence over the broader institutional adoption environment, and whose own career trajectories are likely to be shaped by the AI-enabled transformation of academic work more generally.

12.1 Limitations of the Present Contribution

As a conceptual paper, this contribution is subject to limitations that should be acknowledged directly. The framework presented in Section 9 has been developed through theoretical synthesis rather than empirical validation, and its component relationships, while grounded in established theory, have not yet been tested against primary data from Saudi universities or their female students. The paper has also, of necessity, treated Saudi women as a population with certain broadly shared characteristics arising from a common national policy and cultural environment, while recognising that this population is itself internally diverse along lines of region, socio-economic background, disability status, and field of study, diversity that the intersectional research direction identified above is intended to address in future work. Finally, the pace of change in generative AI capability in particular means that any framework of this kind risks being overtaken by technological developments not anticipated at the time of writing; the paper has sought to mitigate this risk by organising its contribution around durable ethical principles and institutional processes rather than around the specifics of any single current technology, but this risk cannot be eliminated entirely. These limitations do not, in the paper's view, undermine the case for the framework's contribution, but they do underline the priority attached to the empirical research agenda set out above.

12.2 Framework Validation Against Existing AI Governance Models

Section 8 and Box 1 established that the international and national instruments reviewed above address AI governance in general terms, without being designed for a specific population, sector, or national culture. This section makes that comparison more precise, moving from a positioning claim to a specific account of what each instrument provides, what it leaves unaddressed for the present context, and how the framework developed in Section 9 extends it. Table 9 summarises this synthesis; the discussion that follows highlights the extensions that recur across all six instruments.

Table 9

Synthesis of Gaps in Existing AI Governance Instruments and How the Proposed Framework Extends Them

Framework	What It Provides	What It Lacks for This Context	How the Proposed Framework Extends It
UNESCO Recommendation on the Ethics of AI	Global normative values (human rights, dignity, diversity, inclusiveness) spanning policy areas including education and gender	No higher-education-specific mechanisms; gender addressed as a general equality value rather than a specific empowerment outcome	Layers 5 and 6 (Section 9.2) translate its general education and gender provisions into specific, monitorable educational and empowerment outcomes
OECD AI Principles	Human-centred values, transparency, accountability, inclusive growth	No gender-specific or higher-education-specific provisions; not culturally adapted	Layer 3 operationalises these principles as concrete evaluation criteria adapted to the Saudi cultural and religious context discussed in Section 5.6
EU AI Act	Binding, risk-tiered regulation; several categories of educational AI treated as high-risk	Not designed for the GCC institutional or cultural context; no empowerment-outcome layer	The layered architecture in Section 9 supplies the institutional-readiness and empowerment-outcome layers that risk classification alone does not address
NIST AI Risk Management Framework	Voluntary govern–map–measure–manage risk structure	Generic risk functions; no sector-specific or gender-specific content	The roadmap in Section 10 operationalises these functions as six sequential phases specific to university adoption of AI for women's empowerment
ISO/IEC 42001	Certifiable AI management-system standard	No certification pathway specific to higher education or gender-equity outcomes	Section 11 proposes that SDAIA extend its national ethics principles into a comparable certification mechanism specific to gender-sensitive higher-education AI governance
SDAIA National AI Ethics Principles	National alignment with Vision 2030 and Islamic values; partial governance mechanisms	Not sector-specific to higher education; not gender-specific	Layer 1 treats SDAIA's principles as the national mandate through which the framework's remaining five layers supply the missing sector and gender specificity

Read together, three extensions recur across the comparison in Table 9. First, none of the six instruments translates its general education or gender provisions into the sequenced, layered architecture set out in Section 9; this translation, rather than the underlying principles themselves, is the

paper's central original contribution. Second, none supplies an institutional roadmap of the kind presented in Section 10, capable of guiding a Saudi university from readiness assessment through governance, literacy, deployment, and monitoring. Third, none is grounded in empirical evidence specific to Saudi women's advancement of the kind discussed in Section 4.3; the framework's fairness principle (Section 5.4) is calibrated directly against documented patterns of network-mediated appointment rather than against a generic fairness concern. These three extensions are what distinguish the proposed framework from adaptation or synthesis alone.

13. Discussion

This section considers the implications of the analysis and framework developed above across seven dimensions: theoretical, social, gender, policy, educational, implications for the Sustainable Development Goals, and global implications. Several of these dimensions extend discussion already introduced in earlier sections (Sections 9, 11 and 12.2 in particular); the purpose here is to draw out what each set of implications means for interpretation of the paper's contribution as a whole, rather than to restate those sections.

13.1 Theoretical Implications

The framework's central theoretical contribution is the integration of six previously separate literatures, the Capability Approach, Human-Centred AI, Value Sensitive Design, established technology-acceptance theories, Self-Determination Theory, Social Cognitive Theory, and Digital Inclusion Theory, into a single architecture linking national policy, institutional readiness, ethical principles, technology, and outcomes. This integration responds to a wider call in the comparative AI ethics literature for governance models that engage seriously with non-Western normative traditions rather than applying a single, implicitly Western ethical template across contexts (Elmahjub 2023). By treating Islamic and Saudi cultural values as a formal layer of the framework rather than an afterthought, the paper offers one template for the kind of pluralist, context-specific theorising that this literature calls for. The theoretical implication is that empowerment-oriented AI frameworks cannot be treated as a generic addition to existing technology-acceptance or governance models; they require dedicated integration of capability-based and culturally situated theory from the outset.

13.2 Social Implications

Read as a social intervention rather than a purely technological one, the framework implies that Saudi higher education functions as one of the principal sites through which broader change in women's roles, opportunities, and public participation is currently being negotiated (Alhawsawi and Jawhar 2023). Where AI adoption is governed in line with the principles set out in Section 6, it has genuine potential to accelerate the social transformation already under way through Vision 2030, by expanding the practical opportunities available to female graduates beyond the university itself. Where it is not, AI risks

becoming a mechanism through which existing social patterns, including the informal, network-mediated routes into leadership documented in Section 4.3, are reproduced under the appearance of technological neutrality rather than disrupted by it.

13.3 Gender Implications

The paper's gender implications extend beyond parity of access to AI tools. Genuine empowerment, in the sense developed through the Capability Approach in Section 7, requires that AI systems be designed and governed so that women can convert access into substantive capability, agency, and voice, rather than into a formally equal but substantively limited set of options. Read together with the evidence that Saudi women's educational and career trajectories continue to be shaped by social and cultural factors, particularly around field-of-study choice and progression into leadership, this implies that AI governance in this context must attend explicitly to gendered patterns of use, evaluation, and outcome, rather than treating gender as a demographic variable to be reported and then set aside.

13.4 Policy Implications

At the policy level, the paper's central implication, developed in detail in Section 11, is that AI governance and women's empowerment policy cannot be pursued as parallel, independently managed agendas. The framework implies that Vision 2030's technology and human-capital pillars need a shared point of coordination, most plausibly located at the intersection of SDAIA and the Ministry of Education, so that AI procurement, accreditation, and literacy standards are evaluated against gender-sensitive criteria as a matter of routine practice rather than through ad hoc review. Comparative evidence on AI governance across the six GCC states finds that most national strategies in the region currently rely on aspirational ethical principles rather than binding, sector-specific mechanisms (Albous et al. 2025), reinforcing the case for the kind of institutional-layer specificity the framework proposes.

13.5 Educational Implications

For pedagogy and curriculum design, the framework implies that AI literacy for female students should be taught as a form of critical, empowered engagement rather than technical competence alone, consistent with the autonomy-supportive orientation developed in Section 7. This has direct implications for how Saudi universities design AI-related coursework, advising, and mentoring: content that builds technical fluency without also building the capacity to question, evaluate, and where appropriate redirect AI-generated recommendations risks producing passive rather than empowered users, undermining the outcome the framework is intended to support.

13.6 Implications for the Sustainable Development Goals

The framework's contribution to the Sustainable Development Goals, introduced in the Abstract and Introduction, follows from its structure rather than being asserted separately from it. Its ethical-principles layer operationalises SDG 16 by translating abstract commitments to transparency and accountability into specific institutional practice; its empowerment-outcome layer operationalises SDG 5 and SDG 4 by linking AI-enabled educational access to documented gains in women's agency and participation; and its attention to accessibility and inclusiveness speaks directly to SDG 10. This pattern is consistent with recent systematic evidence that AI in higher education advances SDG 4 primarily through personalised learning and adaptive systems, and SDG 10 primarily through accessibility and inclusive support for underrepresented groups (Apata et al. 2026), the same two mechanisms through which the framework's fifth and sixth layers are designed to operate in the Saudi context. The framework further supports SDG 8 through its attention to entrepreneurship and employability, SDG 9 through its treatment of AI as digital infrastructure requiring deliberate institutional investment, and SDG 17 through the multi-stakeholder coordination it proposes between universities, SDAIA, and the Ministry of Education.

13.7 Global Implications

Although developed for the Saudi context, the framework's six-layer architecture does not depend on features unique to Saudi Arabia. This framework may be adapted for GCC countries, MENA nations and other emerging economies pursuing simultaneous digital transformation, responsible AI governance and gender equality agendas. Comparative analysis of national AI strategies across the six GCC states finds a broadly shared reliance on aspirational ethical principles rather than binding, sector-specific governance mechanisms (Albous et al. 2025), which is precisely the gap the framework's institutional and ethical-principles layers are designed to fill. Beyond the Gulf, countries pursuing national development visions that combine digital transformation with explicit gender-equality commitments, a combination increasingly common across the wider MENA region and other emerging economies, face a structurally similar coordination problem, suggesting that the framework's core contribution, the deliberate integration of AI governance with empowerment policy, generalises beyond the specific national policy environment in which it was developed.

14. Conclusion

This paper has argued that the intersection of AI adoption and women's empowerment within Saudi higher education constitutes a distinct field of inquiry, requiring conceptual resources that neither the general literature on AI in education nor the general literature on international AI governance is, by itself, well positioned to supply. The remainder of this section sets out the paper's major findings and its theoretical, practical, policy, SDG-related, and Vision 2030-related contributions, before restating its principal limitations, priorities for future research, and global relevance.

14.1 Major Findings

Three principal findings follow from the analysis in Sections 2 through 9. First, the international AI governance instruments most commonly invoked in this context are not designed for this specific population, sector, or national culture, and the focused comparison in Section 12.2 shows that none of them translates general education or gender provisions into institutional practice at the level of specificity Saudi universities require. Second, Saudi women's documented advancement into academic and educational leadership continues to rely substantially on informal, network-mediated pathways rather than systematic, transparent processes, a pattern that unmanaged AI adoption risks reproducing rather than correcting. Third, a workable response to both findings requires ethical principles, institutional readiness, and empowerment outcomes to be treated as elements of a single layered architecture rather than as separate initiatives pursued in parallel.

14.2 Theoretical Contribution

The paper's theoretical contribution is the integrative synthesis discussed in Section 13.1: six established literatures, spanning capability theory, human-centred and value-sensitive design, technology acceptance, self-determination, social cognition, and digital inclusion, are combined into a single six-layer framework rather than left to operate as parallel explanatory resources. This synthesis responds to calls in the wider AI ethics literature for governance models that engage seriously with non-Western normative traditions rather than applying a single, implicitly Western ethical template across contexts (Elmahjub 2023). The resulting framework is offered as a template that other researchers can adapt, extend, or test, rather than as a finished, self-validating model.

14.3 Practical Contribution

Beyond its theoretical synthesis, the paper offers three practical resources: a six-layer framework capable of structuring institutional evaluation of specific AI tools (Section 9); a six-phase implementation roadmap sequencing the institutional steps required to move from readiness assessment to full deployment and monitoring (Section 10); and a stakeholder-specific set of recommendations addressed to universities, SDAIA, the Ministry of Education, faculty, students, and industry (Section 11). Together, these resources are intended to give Saudi universities and their national regulators a working starting point for AI governance decisions that would otherwise have to be made without a context-specific evidence base.

14.4 Policy Contribution

The paper's policy contribution is the case, developed across Sections 11 and 13.4, for treating AI governance and women's empowerment policy as a single coordinated agenda rather than two agendas pursued by different institutional actors on separate timelines. Concretely, this implies extending SDAIA's national AI ethics principles with higher-education-specific, gender-sensitive guidance; incorporating AI-readiness and gender-sensitive governance criteria into university accreditation; and establishing a

national community of practice among Saudi universities through which governance models and lessons learned can be shared as institutions progress through the roadmap set out in Section 10.

14.5 Contribution to the Sustainable Development Goals

As set out in Section 13.6, the framework's contribution to the Sustainable Development Goals is structural rather than incidental: each of its six layers maps onto specific SDG targets, from SDG 4 and SDG 10 through its educational and inclusion layers, to SDG 5 through its empowerment-outcome layer, to SDG 16 through its ethical-principles layer, to SDG 8, SDG 9, and SDG 17 through its attention to employability, digital infrastructure, and multi-stakeholder coordination respectively. This pattern is consistent with wider evidence that AI's contribution to SDG 4 and SDG 10 in higher education operates specifically through personalisation, adaptive systems, and inclusive accessibility support (Apata et al. 2026), the mechanisms the framework's own educational and empowerment layers are designed to operationalise for the Saudi context.

14.6 Contribution to Vision 2030

The paper's most direct national contribution is to Vision 2030 itself, whose technology and human-capital pillars this paper has argued are currently pursued through largely separate institutional channels. Saudi Arabia's reform programme has already produced substantial, well-documented gains in women's economic and educational participation (Topal 2019); the framework's contribution is to propose a specific mechanism, coordinated AI governance, through which the next phase of that programme can extend those gains into the AI-enabled institutions increasingly central to Saudi higher education, rather than allowing AI adoption to proceed on a separate track from the Kingdom's empowerment commitments.

14.7 Limitations

As discussed more fully in Section 12.1, this contribution is conceptual rather than empirical: its component relationships have not yet been tested against primary data from Saudi universities or their female students, and its treatment of Saudi women as a population with shared characteristics necessarily understates the internal diversity of that population along lines of region, socio-economic background, disability, and field of study. These limitations do not undermine the case for the framework's contribution, but they do mean that the claims made here should be read as a theoretically grounded starting point for empirical work, not as an empirically validated model in its own right.

14.8 Future Research

Section 12 sets out a detailed future research agenda; three priorities are worth restating here. Longitudinal research tracking the academic and career outcomes of female students exposed to varying forms of AI-enabled learning is the single most important next step, since it would allow the conceptual model in Section 4.2 to be tested against outcome data rather than theoretical plausibility alone. Comparative research extending the framework to other GCC higher-education systems would establish which of its six layers generalise beyond the Saudi context and which require national adaptation. Research developing AI-governance maturity models calibrated specifically to the university context would give the institutional-readiness layer of the framework a practical measurement instrument it does not currently have.

14.9 Global Relevance

Although developed specifically for Saudi higher education, the framework's architecture does not depend on features unique to the Kingdom. This framework may be adapted for GCC countries, MENA nations and other emerging economies pursuing simultaneous digital transformation, responsible AI governance and gender equality agendas. Comparative evidence on AI governance across the six GCC states, which finds a broadly shared reliance on aspirational ethical principles rather than binding, sector-specific mechanisms (Albous et al. 2025), suggests that the institutional and ethical-principles layers developed here address a regional gap rather than a purely national one. Judged on those terms, the future of AI in Saudi higher education, and in the wider group of countries pursuing comparable reform agendas, should ultimately be assessed not by the sophistication of its algorithms, but by its capacity to expand women's capabilities, agency, participation, and leadership, while preserving human dignity, cultural values, and ethical governance.

Declarations

Funding: This research received no external funding.

Conflicts of interest / Competing interests: The authors declare no competing interests.

Ethics approval: Not applicable — this study did not involve human participants, human data or animals.

Data availability: No new data were generated or analysed in this conceptual review. All sources are cited in the reference list.

Acknowledgements: During the preparation of this manuscript, the authors used ChatGPT (OpenAI) for language refinement and formatting support. The author reviewed and edited all generated content and take full responsibility for the integrity, accuracy, and originality of the manuscript.

Author Contribution

WMA conceived the study, conducted the literature review, developed the conceptual framework, and drafted the manuscript. OP contributed to the study design, theoretical development, critical revision of the manuscript, and supervision. Both authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

References

1. Alajaji WM, Phiri O (forthcoming). From childhood to headship: Women's leadership development pathways in Saudi secondary schools. Manuscript in preparation
2. Albous MR, Al-Jayyousi OR, Stephens M (2025) AI governance in the GCC states: A comparative analysis of national AI strategies. *J Artif Intell Res* 82:2389–2422
3. Aldossari AS (2025) Women's empowerment: New paradigm shift in Saudi women's decision-making and choice of university majors. *Humanit Social Sci Commun* 12:623
4. Alhawsawi S, Jawhar SS (2023) Education, employment, and empowerment among Saudi women. *Gend Educ* 35(4):401–419
5. Apata OE, Ajoye ST, Apata BO et al (2026) Artificial intelligence in higher education: A systematic review of contributions to SDG 4 (quality education) and SDG 10 (reduced inequality). *Int J Educational Manage* 40(1–2):336–353
6. Baidoo-Anu D, Ansah LO (2023) Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *J AI* 7(1):52–62
7. Bahroun Z, Anane C, Ahmed V, Zacca A (2023) Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability* 15(17):12983
8. Bandura A (2001) Social cognitive theory: An agentic perspective. *Ann Rev Psychol* 52:1–26
9. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 13(3):319–340
10. Deci EL, Ryan RM (2000) The what and why of goal pursuits: Human needs and the self-determination of behavior. *Psychol Inq* 11(4):227–268
11. EL-Deghaidy H, Almasabi Z, Elkholy A, Mansour N (2025) Mapping STEM gender disparities: A scoping review of the MENA region. *J STEM Educ Res* 8(4):531–560
12. Elmahjub E (2023) Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI, vol 36. *Philosophy & Technology*. Article 73
13. European Commission (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence. *Artificial Intelligence Act*
14. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E (2018) AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Mind Mach* 28(4):689–707

15. Friedman B, Hendry DG (2019) Value sensitive design: Shaping technology with moral imagination. MIT Press
16. Holmes W, Bialik M, Fadel C (2019) Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign
17. International Organization for Standardization (2023) ISO/IEC 42001:2023 - Information technology - Artificial intelligence - Management system
18. Kalim U, Kanwar A, Sha J, Xu L, Chen X, Rasool A (2025) Barriers to AI adoption for women in higher education: A systematic review of the Asian context. *Smart Learn Environ*, 12, Article 38.
19. Kalim U, Kanwar A, Xu L, Xu H, Chen X, Irfan S, Rasool A (2026) Female gender bias in artificial intelligence applications for education: A systematic review of regional disparities and equity implications. *AI Soc* 41(4):2943–2959
20. National Institute of Standards and Technology (2023) AI Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce
21. Organisation for Economic Co-operation and Development (2019) Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.
22. Selwyn N (2019) Should robots replace teachers? AI and the future of education. *Polity*
23. Sen A (1999) Development as freedom. Oxford University Press
24. Shneiderman B (2020) Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int J Hum Comput Interact* 36(6):495–504
25. Topal A (2019) Economic reforms and women's empowerment in Saudi Arabia. *Women's Stud Int Forum* 76:102253
26. UNESCO (2021) Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization
27. UNESCO (2026) Saudi Arabia: 2026 Global Education Monitoring Report country case study. United Nations Educational, Scientific and Cultural Organization
28. United Nations (2015) Transforming our world: The 2030 agenda for sustainable development (A/RES/70/1). United Nations General Assembly
29. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: Toward a unified view. *MIS Q* 27(3):425–478
30. Warschauer M (2003) Technology and social inclusion: Rethinking the digital divide. MIT Press
31. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F (2019) Systematic review of research on artificial intelligence applications in higher education - Where are the educators? *Int J Educational Technol High Educ* 16(1):39
32. Zimmerman MA (2000) Empowerment theory: Psychological, organizational, and community levels of analysis. In: Rappaport J, Seidman E (eds) *Handbook of community psychology*. Springer, pp 43–63

Figures

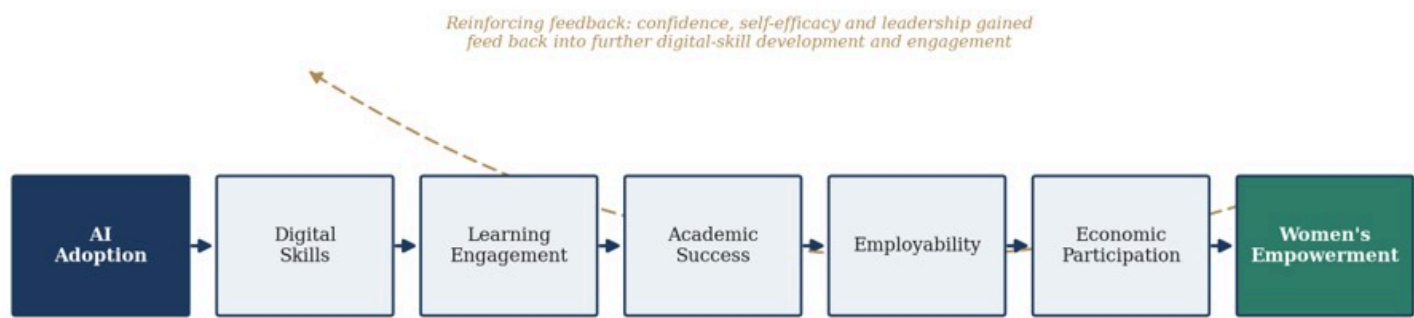


Figure 1

Conceptual Model Linking AI Adoption to Women's Empowerment Outcomes

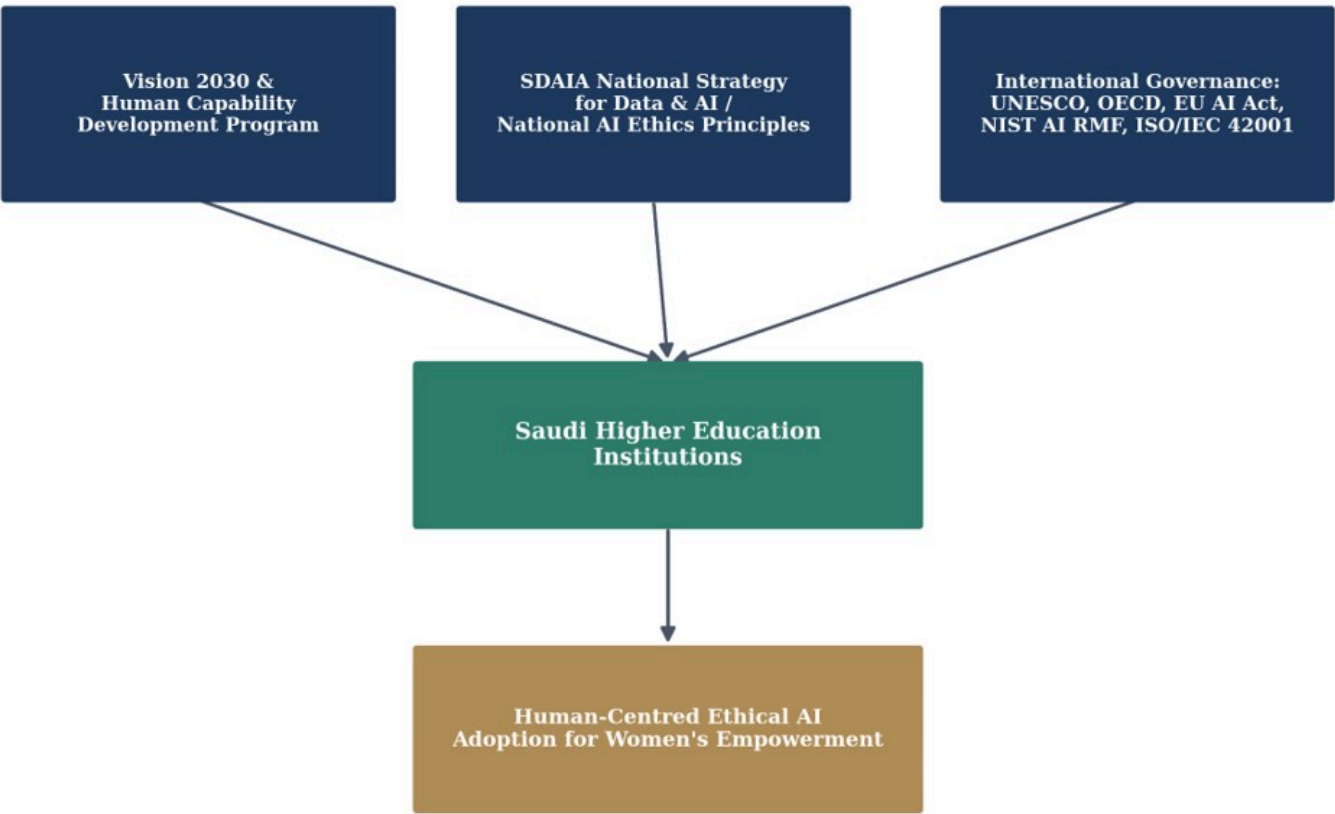


Figure 3. Policy alignment map: convergence of national and international governance instruments upon Saudi higher education institutions

Figure 2

Figure 3. Policy Alignment Map: Convergence of National and International Governance Instruments upon Saudi Higher Education Institutions



Figure 3

Figure 2. The Human-Centred Ethical AI Framework for Women’s Empowerment in Saudi Higher Education (Six Layers)

Continuous monitoring feeds back into readiness assessment and governance

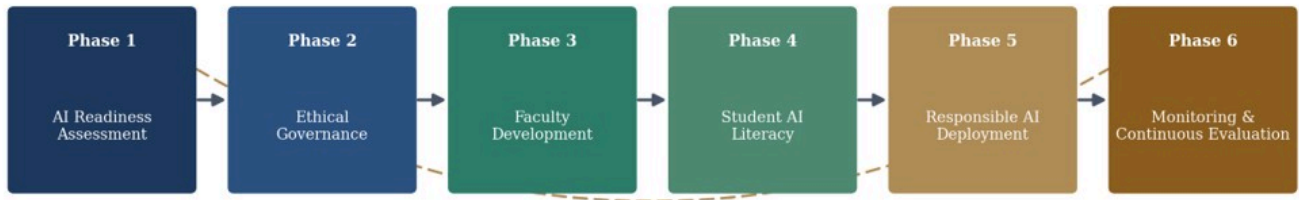


Figure 4. Phased implementation roadmap for human-centred AI adoption

Figure 4

Figure 4. Phased Implementation Roadmap for Human-Centred AI Adoption