مؤسسة قطر
Qatar Foundation

وايز
WISE

الأبحاث والسياسات
RESEARCH & POLICY

FORTIFYING EDUCATION IN THE AGE OF DISINFORMATION

A Strategic Framework for AI in Higher Education in Qatar

November 2025

Lead Authors
Dr. Georgios Mikros
Dr. Wajdi Zaghouani
Dr. Marc Owen Jones

Co-authors
Selma Talha-Jebril
Sopiko Beriashvili

Northwestern
QATAR

جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Contents

Glossary of Acronyms

AI – Artificial Intelligence

API – Application Programming Interface

BERT – Bidirectional Encoder Representations from Transformers

CHSS – College of Humanities and Social Sciences

COVID-19 – Coronavirus Disease 2019

CRA – Qatar's Communications Regulatory Authority

CRAAP – Currency, Relevance, Authority, Accuracy, Purpose (evaluation test)

EC – Education City

EDMO – European Digital Media Observatory

EU – European Union

GCC – Gulf Cooperation Council

GDPR – General Data Protection Regulation

GPT – Generative Pre-trained Transformer

HBKU – Hamad Bin Khalifa University

HE – Higher Education

IT – Information Technology

ITU – International Telecommunication Union

KPI – Key Performance Indicator

LLM – Large Language Model

LMS – Learning Management System

MBZUAI – Mohamed bin Zayed University of Artificial Intelligence

MCIT – Ministry of Communications and Information Technology (Qatar)

MENA – Middle East and North Africa

MIL – Media and Information Literacy

MOEHE – Ministry of Education and Higher Education (Qatar)

MOI – Ministry of Interior (Qatar)

MOPH – Ministry of Public Health (Qatar)

NDLIRS – National Digital Literacy & Information Resilience Strategy

NEMCDL – National Educator Micro-Credential in Digital Literacy

NEOM – (Saudi Arabian mega-city project)

NLP – Natural Language Processing

NU-Q – Northwestern University in Qatar

OSINT – Open-Source Intelligence

QA – Quality Assurance

QCAI – Qatar Center for AI

QCRI – Qatar Computing Research Institute

QF – Qatar Foundation

QIRO – Qatar Information Resilience Observatory

QNRF – Qatar National Research Fund

QU – Qatar University

ROI – Return on Investment

SDAIA – Saudi Data & AI Authority

STEM – Science, Technology, Engineering, and Mathematics

TAMUQ – Texas A&M University at Qatar

UAE – United Arab Emirates

UN – United Nations

UNESCO – United Nations Educational, Scientific, and Cultural Organization

WEF – World Economic Forum

WISE – World Innovation Summit for Education

Foreword

Reflective of our shared commitment to providing unparalleled opportunities for inquiry and discovery, Hamad Bin Khalifa University (HBKU), WISE, and Northwestern University in Qatar (NU-Q) equally understand the profound, and possibly disruptive, role AI is now playing in higher education.

Much beyond streamlining administrative tasks, AI promises a future of highly innovative learning experiences in which pedagogy is finely attuned to individual requirements. It raises the prospect of greater inclusiveness in academic settings, as well as completely new approaches to student empowerment and achievement.

Despite recognizing the tremendous potential, HBKU, NU-Q, and WISE are also acutely aware of the societal and ethical dilemmas technologies like AI can pose. AI's inability to understand biases and inaccuracies automatically risks the spread of disinformation and misinformation where they are needed least. Universities' mandate includes a moral obligation to ensure AI does not propel exclusion and misconduct where academic freedom, student wellbeing, and unfettered access to credible sources of information must remain the norm.

While we acknowledge that AI's role in higher education will continue to grow, we are adamant that the solutions proposed by this report will enable faculty, students, and staff to contend with unreliable and misinformed AI-generated content. Now is not the time for higher education to stand still. This report lays solid foundations for a future in which we work with AI to tackle the very real threat of a distorted information environment while continuing to provide safe and equitable learning for all.

Dr. Ahmed M. Hasnah
President
Hamad Bin Khalifa University

In an age when artificial intelligence is transforming how knowledge is produced, circulated, and trusted, education must act as both compass and anchor. This report, Fortifying Education in the Age of AI and Disinformation—a collaboration among the World Innovation Summit for Education (WISE), Hamad Bin Khalifa University (HBKU), and Northwestern University in Qatar—embodies that dual mission.

At its core, the study tackles one of the defining challenges of our time: safeguarding education systems from the distortions of AI-generated misinformation while harnessing the same technologies to advance learning, critical thinking, and social cohesion. It does so with intellectual rigor and institutional collaboration, linking empirical inquiry at HBKU with applied innovation at Northwestern Qatar, and culminating in a national framework for resilience. The report's tripartite structure—spanning case study, digital toolkit, and policy strategy—mirrors Qatar's broader ambition to connect knowledge creation to societal well-being.

In Arabic, haqiqah means "truth" or "reality." The Haqiqatar strategy proposed in this report draws on that spirit, affirming Qatar's leadership in shaping an educational ecosystem that empowers learners to engage critically, ethically, and responsibly with emerging technologies.

At Northwestern Qatar, we are proud to contribute to this dialogue. The future of education will depend not only on our capacity to innovate with technology, but on our commitment to preserve the human values—curiosity, empathy, and integrity—that define true learning.

Marwan M. Kraidy
Dean and CEO
Professor of Communication
Anthony Shadid Chair in Global Media, Politics and Culture
Northwestern University in Qatar

Editors' Note

At the World Innovation Summit for Education (WISE), we believe that the future of teaching and learning depends not only on innovation but on trust: trust in information, in institutions, and knowledge itself. During the WISE11 Summit (2023) one urgent question emerged: how can higher education build resilience against the twin challenges of generative artificial intelligence and digital disinformation? This question revealed a profound research gap, one that speaks directly to the credibility, ethics, and mission of universities in the 21st century.

In response, WISE launched a dedicated 12-month research initiative, "Fortifying Higher Education in the Age of AI and Disinformation," in collaboration with Hamad Bin Khalifa University (HBKU) and Northwestern University in Qatar (NU-Q). This partnership brought together technologists, educators, and researchers to examine how universities can responsibly adopt AI while strengthening resilience against the spread of false or harmful information. This report represents the culmination of that effort. It provides not just an analysis of the risks of AI and disinformation pose to higher education but a forward-looking strategy for action rooted in evidence, collaboration, and human-centric values. Through three interconnected sub-studies examining AI adoption in university classrooms, strategies for fostering healthier online dialogue by addressing toxicity and bias in multilingual academic settings, and finally, strengthening national information resilience by introducing the Haqiqatar strategy, a human-centered roadmap for building societal resilience against disinformation through critical literacy and experiential learning.

WISE's leadership in convening and guiding this work reaffirms its commitment to exploring critical issues at the intersection of education, policy, and emerging technologies. By initiating this inquiry within Qatar's Education City, WISE has contributed to a dialogue of global importance on how education systems can develop a coordinated response to the challenges of AI and disinformation. WISE remains dedicated to advancing the work by engaging the diverse voices of those who shape and are shaped by the future of education in an AI-driven world.

This report invites policymakers, educators, and innovators to see beyond the dichotomy of opportunity versus risk. It offers a blueprint for building institutions that not only integrate AI responsibly but also foster informed, critical, and resilient learners. Those capable of discerning truth in an opaque digital era.

At its core, this endeavor seeks to ensure that as technology evolves, education remains humanity's strongest safeguard against misinformation and a powerful engine for progress.

Selma Talha-Jebril
WISE, Director of Research and Policy

Sopiko Beriashvili
WISE, Research and Policy Lead

EXECUTIVE SUMMARY

Unified Strategy for an Age of AI and Disinformation

The advent of powerful Generative Artificial Intelligence (Gen-AI) and the proliferation of sophisticated digital disinformation represent two sides of the same coin, a profound, systemic challenge to the core mission of higher education. Together, they are reshaping how knowledge is created, validated, and shared, creating both unprecedented opportunities and significant risks. Institutions worldwide are grappling with a dual imperative: how to harness AI as a transformative tool for learning and research while simultaneously building resilience against the digital threats that undermine academic integrity, social cohesion, and informed civic discourse.

This study presents the findings of *Fortifying Higher Education in the Age of AI and Disinformation*, a collaborative research initiative between the World Innovation Summit for Education (WISE), led by Selma Talha Jebril (Director of Research and Policy) and Sopiko Beriashvili (Research and Policy Lead); the College of Humanities and Social Sciences (CHSS) at Hamad Bin Khalifa University (HBKU), represented by Dr. George Mikros; and Northwestern University, Qatar (NU-Q), represented by Dr. Wajdi Zaghouani and Dr. Marc Owen — commissioned under Qatar Foundation (QF). The study provides an in-depth analysis of how AI is reshaping higher education, with particular focus on addressing AI-driven disinformation and enhancing educational resilience within Qatar's Education City. This initiative employs three integrated research streams: examining student and faculty perceptions of AI in classroom settings, developing AI-powered tools to mitigate toxic online behavior, and designing role-playing workshops to build disinformation awareness. Together, these streams offer both empirical insights and practical strategies for responsibly integrating AI into Qatar's higher education landscape.

The research moves beyond reactive responses to offer a proactive, human-centric vision for navigating this transformative era. Drawing on a comprehensive case study of AI adoption at HBKU, the development of culturally attuned tools for fostering healthy online discourse, and a proposed national strategy for building information resilience, this report synthesizes evidence-based recommendations across institutional, technological, and societal dimensions. These recommendations point to a clear opportunity: Education City is home to branch campuses of leading international universities, a homegrown research university (HBKU), start-up incubators, technology parks, and cultural institutions. As an educational and research hub, Education City is uniquely positioned to pioneer an integrated model that reaffirms the university's role as both a guardian of trusted knowledge and a cultivator of critical, digitally literate citizens capable of thriving in an age of AI and disinformation.

The Core Problem: A Triad of Strategic Gaps

Our comprehensive analysis reveals three interconnected gaps that define the current landscape. Addressing these requires a coordinated, ecosystem-wide response.

1. **The Adoption and Policy Gap:** The first sub-study, based on an in-depth case study at HBKU and NU-Q, confirms that AI adoption is not a future event; it is a present reality driven by pragmatic, bottom-up student use. While over 75% of students use AI weekly or daily for academic tasks, faculty and institutions are struggling to keep pace. This has created a significant policy vacuum, leaving educators unprepared, assessment methods outdated, and academic integrity vulnerable. The data reveals a community that is not resistant to AI but is explicitly asking for clear guidance, training, and a structured framework to manage this transformation responsibly.

2. **The Discourse and Safety Gap:** The second sub-study addresses the digital spaces where learning now occurs. In a multicultural and multilingual environment like Education City, online discussion forums are essential but are also fertile ground for misinterpretation, toxicity, and conflict. Our research shows that generic, English-centric AI moderation tools consistently fail to grasp the nuances of Arabic dialects, code-switching, and cultural context. This creates a safety and fairness gap, where legitimate academic debate can be curtailed while targeted harm goes undetected. Closing this gap requires a "human-in-the-loop" approach, supported by culturally attuned technology and pedagogy designed to foster healthy online discourse.
3. **The Resilience and Literacy Gap:** The third sub-study, *Haqiqatar*, zooms out to the national level, recognizing that Qatar has been subject to the sustained, state-level disinformation campaigns. This creates a critical literacy gap, as many citizens, including students, adults, and vulnerable populations, lack the sophisticated media and information literacy skills needed to identify manipulation, resist emotional targeting, and avoid spreading falsehoods. This is not merely an educational issue; it is a matter of social cohesion and national security, demanding a coordinated strategy to build "information resilience" across society.

Unified Strategic Response: Three Pillars for Action

To close these gaps, this report proposes a unified strategy built on three mutually reinforcing pillars. This framework integrates the key recommendations from all three sub-studies into a coherent plan for action.

Pillar 1: Establish Principled Governance and Human-Centric Policies. The immediate priority is to move from ambiguity to clarity. This requires establishing a unified, cross-campus AI Governance Framework for Education City. Led by a dedicated task force, this framework must create an AI Ethics Charter that commits to fairness, transparency, and accountability. In addition to providing guidance on responsible use of ai for students and faculty It must also provide a practical Policy Toolkit for faculty, with model syllabus statements, guidance on redesigning assessments, and clear rules for citing AI. Crucially, this governance must extend to digital learning spaces by formalizing a human-in-the-loop model for online moderation, ensuring that trained individuals, not opaque algorithms, make all consequential decisions. This internal governance will provide the foundation for advocating for greater platform transparency and accountability at the national level while supporting ethical and responsible use of AI.

Pillar 2: Build Capacity Through Integrated Education and Development. A policy is only as effective as the people who implement it. This pillar focuses on building human capacity across the entire educational ecosystem.

- **For Faculty and Staff:** Launch a coordinated, multi-tiered professional development program that moves beyond basic tool training. It must include advanced pedagogical support for integrating AI into the curriculum, redesigning assessments to foster higher-order thinking, and practical skills for de-escalating conflict and moderating online discussions effectively.
- **For Students:** Embed critical AI and media literacy into the core curriculum for every student, regardless of their major. This education must go beyond simple checklists to include interactive, gamified, and role-playing exercises that build durable resilience against manipulation. The digital literacy platform and workshop models developed in sub-studies 2 and 3 provide a ready-to-deploy foundation for this effort.
- **For the Nation:** Extend this educational mission beyond the campus walls by creating lifelong learning pathways for adults and vulnerable populations. By partnering with community organizations, libraries, and workplaces, Education City can champion a whole-of-society approach to building information resilience.

Pillar 3: Develop Sovereign Capabilities through Collaborative Research and Innovation. To lead effectively, Education City must move from being a consumer of external AI technologies to a key contributor to their ethical and culturally relevant development. This requires fostering a vibrant, cross-campus research ecosystem.

- **Establish a Collaborative Hub:** Launch an interdisciplinary "AI in Education" research initiative, building on existing strengths at Qatar Computing Research Institute (QCRI), HBKU, WISE and partner universities. This hub would serve as the engine for the proposed Qatar Information Resilience Observatory (QIRO).
- **Focus on Strategic Niches:** Prioritize research that addresses local needs and global gaps. This includes developing and refining Arabic-native AI tools (leveraging the curated toxicity dataset from sub-study 2), investigating the cognitive and social impacts of AI on learning, and conducting applied research to monitor and counter disinformation threats targeting Qatar and the region. This work will ensure that Education City's strategy is continuously informed by evidence and remains at the cutting edge of innovation.

The challenges of AI and disinformation are not separate issues to be managed in silos; they are deeply intertwined facets of our new digital reality. Responding effectively requires a holistic and unified strategy that aligns policy, pedagogy, and technology around a shared, human-centric vision.

Qatar Education City, with its unique ecosystem of international universities, research capabilities, and alignment with Qatar's national vision, is perfectly positioned to lead this charge. By closing the gaps in policy, discourse, and literacy, it can do more than protect its own academic mission. It can pioneer a model for the responsible integration of AI, serving as a benchmark for the region and the world. This model can demonstrate that technological advancement and the cultivation of thoughtful, ethical, and resilient human beings are not competing priorities but mutually reinforcing goals. The moment to lead is now.

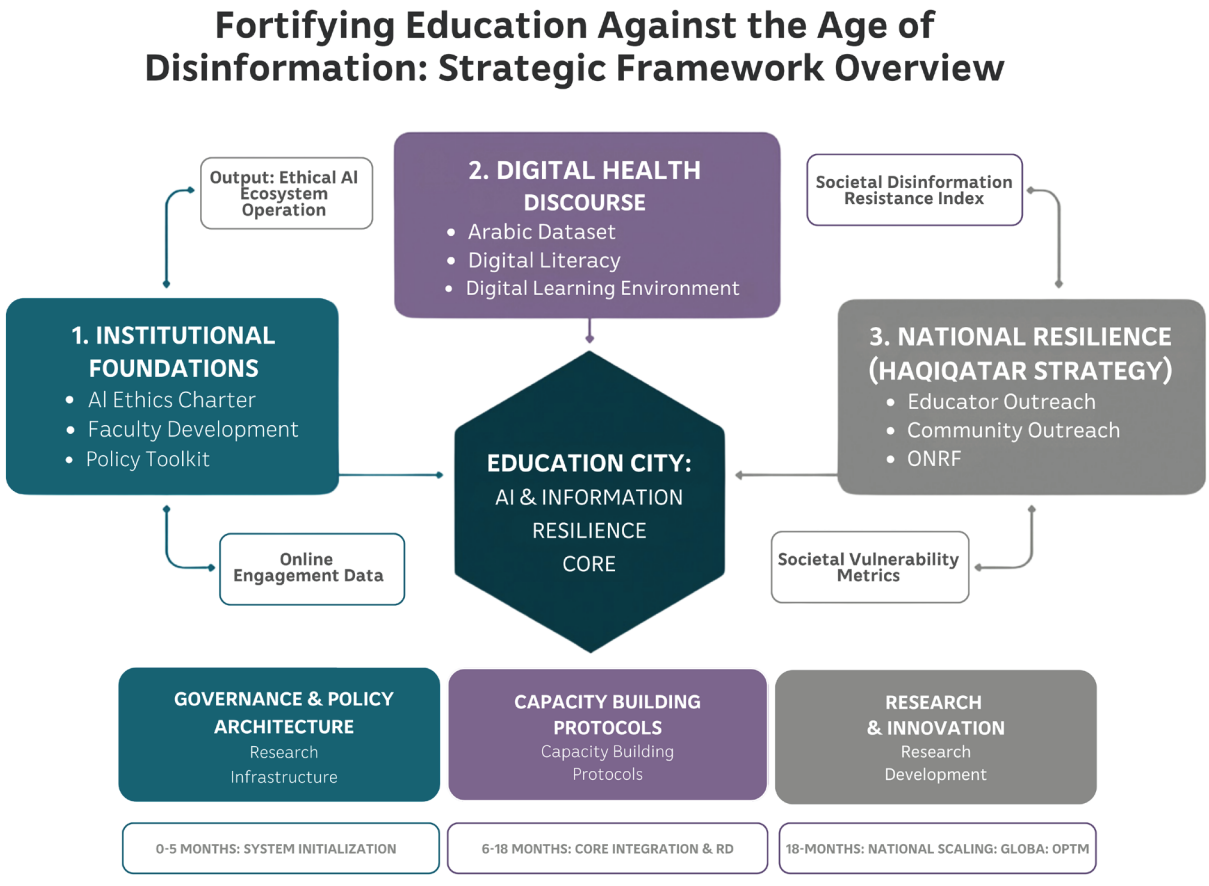


Figure 1.1: Strategic Framework for Fortifying Education in the Age of AI and Disinformation: A visual summary of the three interconnected pillars—Institutional Foundations, Healthy Digital Discourse, and National Information Resilience—supported by governance, capacity building, and research.

Introduction

I. Background and Context

The advent of powerful and widely accessible Generative Artificial Intelligence (Gen-AI) represents one of the most significant technological shifts in modern history, with profound implications for education, information ecosystems, and democratic discourse. Tools capable of generating sophisticated text, images, and analysis have moved from experimental laboratories to everyday use, fundamentally altering how knowledge is created, validated, shared, and, critically, how it can be manipulated. This transformation arrives at a moment when the global information environment is already under strain. The proliferation of disinformation has created what many scholars describe as an "epistemic crisis"—a breakdown in our collective ability to agree on basic facts —and has been defined by the World Economic Forum as one of the most serious short-term threats to global stability.

For higher education institutions, this presents a dual challenge. Universities must harness AI as a transformative tool for learning and research while simultaneously building resilience against the digital threats that undermine academic integrity, social cohesion, and informed citizenship. This is not a choice between embracing or rejecting technology; AI adoption is already happening, driven by students, faculty, and the broader society. The question is whether institutions will shape this transformation intentionally or merely react to it. Globally, institutions are struggling to keep pace. A mid-2023 UNESCO survey revealed that fewer than 10% of universities had established formal policies on generative AI, highlighting a significant governance vacuum (UNESCO, 2023). However, a clear upward trend has been observed in more recent UNESCO survey, which indicates that nearly two-thirds of higher education institutions have now developed or are in the process of developing guidance on AI use (UNESCO, 2025).

In the Middle East and North Africa (MENA) region, the transformation is equally urgent. Governments in Qatar, the UAE, and Saudi Arabia have launched ambitious national AI strategies, viewing technological leadership as essential for economic diversification and global competitiveness. Amidst this dynamic landscape, Qatar has carved out a distinctive niche characterized by a deep, ecosystem-focused integration of AI strategy with its long-term vision for human capital development. Its strategy is built upon the unique infrastructure of Education City, leveraging a cluster of universities as a crucible for both research and talent cultivation, with a key differentiator being its leadership in Arabic language technologies.

It is within this context of immense promise and urgent uncertainty that this research initiative was conceived. Funded and led by the World Innovation Summit for Education (WISE) and conducted through Hamad Bin Khalifa University (HBKU) and Northwestern University Qatar (NUQ), this report seeks to move beyond reactive responses to offer a proactive, evidence-based, and human-centric framework for navigating the challenges of AI and disinformation in educational settings. Rather than treating AI adoption, online discourse quality, and information resilience as separate challenges, this study recognizes them as deeply interconnected facets of a single, systemic transformation. Our core philosophy is that an effective strategy must be explicitly designed to combine the strengths of people and technology in a **"human-in-the-loop" system**, where technology provides scale and data, but trained people make consequential judgments. Through this partnership, WISE and its partners aim to provide Education City, and by extension Qatar and the broader region, with a strategic blueprint for responding to this transformation in a way that preserves the core mission of higher education while harnessing technology for human flourishing.

The report is organized into three main parts, which together form a comprehensive ecosystem approach, moving from understanding the problem, to designing solutions, to implementing them at scale.

- **Part 1: Institutional Foundations** presents an in-depth case study of AI adoption at HBKU, establishing the empirical foundation for our recommendations by examining how students and faculty perceive, use, and are governed in their interactions with AI.

- **Part 2: Fostering Healthy Digital Discourse** translates these insights into action by developing practical tools, including a curated Arabic toxicity dataset and a bilingual digital literacy platform, to address the critical challenge of maintaining respectful and informed online academic interactions.
- **Part 3: Building National Resilience** scales these lessons to the policy level, proposing the *Haqiqatar* framework—a ten-year national strategy for building information resilience and media literacy that aligns with Qatar's long-term education and innovation goals.

Taken together, these three subprojects ensure that empirical insights inform intervention design, and that both feed into a coherent, evidence-based policy vision for building ethical, informed, and resilient digital learning environments.

II. Purpose and Objectives of the Report

Rather than treating AI adoption in higher education settings, online discourse quality, and information resilience as separate challenges, this study recognizes them as deeply interconnected facets of a single, systemic transformation. Through this partnership, WISE and HBKU aim to provide Education City, and, by extension, Qatar and the broader MENA region, with a strategic blueprint for responding to this transformation in a way that preserves the core mission of higher education while harnessing technology for human flourishing.

Figure 1.2: Education City in Qatar

The study comprises three complementary sub-studies, each addressing a distinct but interrelated dimension that can be seen in the following table:

Table 1.1: Three Integrated Sub-studies at a Glance

Three Integrated Subprojects at a Glance			
Substudy	**What**	**How**	**Deliverables**
1. Institutional Foundations (Education City)	AI governance frameworks for higher education	Case study at HBKU examining student and faculty AI use and concerns	• AI governance framework • Policy toolkit • Faculty development program
2. Healthy Digital Discourse (Education City + regional)	Tools to combat online toxicity in educational spaces	Arabic toxicity dataset creation + bilingual digital literacy platform	• 10,000-post Arabic dataset • AI moderation tools • Digital literacy platform
3. National Information Resilience (National - Qatar)	Nationwide media literacy and disinformation resilience	"Haqiqatar" 10-year national strategy based on EC lessons	• National strategy document • Educator training program (NEMCDL) • Implementation roadmap

Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

Integration and Coherence

While each sub-study can stand independently, its actual value emerges through their integration. Under WISE's leadership, integrating these interconnected sub-studies into a comprehensive initiative allows for a holistic understanding of how misinformation and AI intersect within higher education. This integrated framing ensures the study speaks to multiple audiences: students, faculty, institutional leaders, researchers, and policymakers, each of whom plays a role in shaping informed and resilient educational ecosystems.

The **HBKU case study (sub-study 1)** establishes the empirical foundation by uncovering how AI tools are used, perceived, and governed within higher education settings. It generates a grounded understanding of both opportunities and challenges faced by faculty and students in navigating AI adoption responsibly.

The **digital discourse project (sub-study 2)** translates these insights into action by developing practical tools and interventions that address one of the key challenges identified in sub-study 1: the quality and health of online academic interactions. By fostering respectful, informed digital engagement, it bridges the gap between institutional practice and individual digital behavior.

The **national strategy (sub-study 3)** then scales these lessons to the policy level, extending institutional best practices to the broader societal context. The proposed *Haqiqatar* framework provides a roadmap for national information resilience and media literacy that aligns with Qatar's long-term education and innovation goals.

Taken together, these three sub-studies form a comprehensive ecosystem approach, moving from understanding the problem to designing solutions to implementing them at scale. Their integration ensures that empirical insights inform intervention design, and that both feed into a coherent, evidence-based policy vision for building ethical, informed, and resilient digital learning environments.

III. Methodology

This research initiative employed a mixed-methods approach, combining quantitative and qualitative research techniques tailored to each sub-study's objectives while maintaining methodological coherence across the entire study.

Overall Research Design

The project was structured as a sequential, integrated design where each phase builds upon the previous one:

- **Phase 1 (sub-study 1):** Empirical data collection and analysis to establish the evidence base
- **Phase 2 (sub-study 2):** Tool and resource development informed by Phase 1 findings
- **Phase 3 (sub-study 3):** Policy framework development, scaling insights from Phases 1 and 2

Sub-study 1: Institutional Case Study Methodology

- **Data Source**: Survey data from the AI-EDAPT (Artificial Intelligence for Educational Adaptation, Personalization, and Transformation) research project, funded by a Hamad Bin Khalifa University Signature Research Grant. This attribution is essential for maintaining academic integrity and acknowledging the foundational work of the AI-EDAPT research team.
- **Sample**: Faculty (N=17) and students (N=48) across three HBKU colleges (College of Humanities and Social Sciences – CHSS, College of Law - CL, College of Science and Engineering - CSE). This multi-college sample ensures representation across humanities, law, and STEM disciplines, providing a comprehensive view of AI adoption patterns.
- **Instruments**: Structured survey questionnaires covering: - Current AI usage patterns and frequency - Perceptions of AI benefits and risks - Concerns about academic integrity and assessment - Training and support needs - Institutional policy preferences
- **Analysis**: Quantitative analysis using descriptive and inferential statistics to identify patterns, correlations, and significant differences across demographic groups and disciplines.
- **Case Study Rationale**: HBKU serves as a reliable proxy for Education City because of its research-intensive, multidisciplinary nature and deep integration into the Qatar Foundation ecosystem. The patterns observed at HBKU appear to be indicative of, though not necessarily identical to, broader Education City dynamics.

Sub-study 2: Dataset Development and Tool Creation Methodology

- **Literature Review**: Comprehensive review of existing AI tools for detecting harmful language and promoting constructive dialogue, with a focus on Arabic-language contexts.
- **Dataset Creation**: - Collection of 10,000 Arabic social media posts exhibiting various forms of online toxic behavior (hate speech, harassment, bullying, etc.) - Professional annotation using a structured ontology developed through expert consultation - Quality assurance protocols to ensure inter-annotator reliability - Baseline AI model development for toxicity classification.
- **Platform Development**: - Iterative design process informed by pedagogical research on digital literacy - Bilingual (Arabic/English) interface to serve diverse Education City population - Integration of evidence-based strategies for addressing online toxicity - User testing and refinement with Education City stakeholders.

Sub-study 3: National Strategy Development Methodology

- **Policy Analysis**: Examination of successful digital literacy and information resilience strategies from comparable contexts (UNESCO frameworks, EU initiatives, Singapore's models).
- **Stakeholder Consultation**: Engagement with Qatari educational, media, and policy stakeholders to ensure contextual relevance and implementability.
- **Framework Development**: Creation of a comprehensive policy framework including: - Guiding principles aligned with Qatar National Vision 2030 - Seven policy pillars addressing different societal sectors - Implementation timelines and governance structures - Key performance indicators for evaluation
- **Workshop Design**: Development and pilot testing of role-playing-based disinformation awareness workshops incorporating gamification principles and cultural context.

Methodological Coherence

Across all three sub-studies, the research maintained several unifying principles:

- **Evidence-informed**: All recommendations grounded in empirical data or established best practices
- **Context-relevant**: Attention to Qatar's cultural, linguistic, and institutional context
- **Participatory**: Stakeholder engagement throughout the research process
- **Actionable**: Focus on producing practical, implementable outcomes
- **Ethical**: Adherence to research ethics standards, particularly regarding data privacy and human subjects

IV. Scope and Limitations

Scope

This research initiative focuses specifically on:

- **Geographic Scope**:
  - Institutional analysis: Education City, Qatar (with HBKU as primary case study)
  - National strategy: State of Qatar
  - -Regional context: MENA region for comparative insights
- **Sectoral Scope**:
  - Primary focus: Higher education institutions
  - Secondary focus: National education systems and information ecosystems
  - Cross-sectoral connections: Media, technology platforms, policy/governance
- **Thematic Scope**:
  - AI adoption and integration in education
  - Online discourse quality and digital citizenship –
  - Information resilience and media literacy
  - Governance and policy frameworks
- **Temporal Scope**:
  - Data collection period: Spring Semester 2025
  - Strategy horizon: 10-year framework (2026-2035 for national strategy)

Limitations

Sub-study 1 Limitations:

- **Generalization**: The findings are drawn from a single institution (HBKU) and cannot be statistically generalized to the entire Education City population or beyond. However, we argue that HBKU's characteristics, as a large, research-intensive, multidisciplinary university at the heart of Qatar Foundation, make it a highly relevant microcosm whose patterns are likely indicative of broader Education City dynamics.
- **Timing**: Survey data represents a snapshot at a particular moment in AI's rapid evolution. Perceptions and usage patterns will continue to evolve as the technology and its applications mature.
- **Self-reporting**: Survey data relies on participant self-reporting, which may be subject to social desirability bias or incomplete recall.

Sub-study 2 Limitations:

- **Language focus**: While the platform is bilingual, the toxicity dataset focuses primarily on Arabic-language content. Patterns may differ in other linguistic contexts.
- **Platform specificity**: Data drawn from specific social media platforms may not fully represent toxicity patterns across all digital spaces.
- **Dynamic nature**: Online discourse norms and toxic behavior patterns evolve rapidly, requiring ongoing dataset updates and platform refinement.

Sub-study 3 Limitations:

- **Implementation gap**: As a proposed framework, the actual implementation will depend on political will, resource allocation, and stakeholder buy-in, which are beyond the researchers' control.
- **Policy evolution:** The rapidly changing AI and information landscape means some recommendations may require adjustment as technologies and threats evolve.
- **Cross-cultural transfe**r: While the strategy is specifically designed for Qatar, any adaptation to other contexts would require careful consideration of local conditions.

General Limitations:

- **Resource constraints**: Scope limited by available funding and timeline
- **Rapid change**: AI technology evolving faster than research and policy can track
- **Complexity**: Interconnected challenges of AI and disinformation resist simple solutions
- **AI topic fatigue**: At the time of data collection, artificial intelligence had become a ubiquitous topic across media, professional, and academic contexts, potentially leading to survey fatigue among potential participants.

Despite these limitations, the research provides valuable insights, practical tools, and strategic frameworks that can inform decision-making and guide action for Education City, Qatar, and potentially the broader region.

Structure of This Report

Following this introduction, the report is organized into three main parts, each corresponding to one sub-study, followed by a unified conclusion:

- **Part 1**: Institutional Foundations – Education City Strategic Report on AI
- **Part 2**: Fostering Healthy Digital Discourse – From Data to Literacy
- **Part 3**: Building National Information Resilience – The Haqiqatar Strategy
- **Unified Conclusion** – Integrated implementation framework and call to action

The three components of this initiative represent distinct yet mutually reinforcing strategies for strengthening education in the era of AI and disinformation. Part 1 establishes the governance structures and institutional policies necessary for the responsible adoption of AI, examining how students and faculty at HBKU perceive and utilize AI in their academic work. Part 2 demonstrates how AI tools can actively combat online toxicity and promote healthier digital discourse within educational platforms. Part 3 complements these technological interventions with a human-centered approach, proposing a national framework for building citizen resilience against disinformation through experiential learning and role-playing scenarios. While each part addresses a different scale and approach (institutional, technological, and societal), they converge on a single vision: educational institutions as guardians of knowledge integrity and cultivators of critical digital citizenship.

The synergies among these three initiatives are substantial. The institutional AI policies developed in Part 1 create an ethical framework within which the technological interventions of Part 2 can operate transparently and accountably. The digital literacy programs proposed in Part 3 prepare students and citizens to engage productively with both the AI tools and discourse moderation systems explored in Parts 1 and 2. The HBKU case study in Part 1 provides empirical validation for the comprehensive approaches proposed in Parts 2 and 3. This integration is not merely a matter of organizational convenience. It reflects the reality that AI adoption, online discourse quality, and information resilience are inextricably linked challenges that require coordinated solutions.

Each part can be read independently, but together they form a comprehensive strategy for fortifying education in the age of AI and disinformation.

PART 1: INSTITUTIONAL FOUNDATIONS

Education City Strategic Report on AI: From Policy to Practice

This section presents the empirical foundation of our research: an in-depth case study of AI adoption at Hamad Bin Khalifa University (HBKU). By examining how students and faculty actually perceive and use AI in their academic work, we establish the evidence base for understanding both the opportunities and challenges facing Education City institutions. The insights from this analysis directly inform the practical tools developed in Part 2 and the national policy framework proposed in Part 3.

1 The AI Landscape in Higher Education

1.1 Global Trends, Opportunities, and Challenges

The integration of artificial intelligence into higher education is a global phenomenon, characterized by rapid adoption, significant investment, and a collective effort to navigate its disruptive potential. The global AI-in-education market is projected to grow over 30% annually, driven by the demand for personalized learning and the digital transformation accelerated by the recent pandemic (Grand View Research, 2024). This technological surge is compelling universities to move AI from the periphery of their strategic plans to the very center.

1.1.1 Key Opportunities:

The promise of AI in higher education is vast. For students, AI-powered adaptive learning platforms and intelligent tutoring systems offer the potential for highly personalized educational experiences, tailoring content and pacing to individual needs (Bao, 2024). This can make learning more flexible, accessible, and effective, particularly for diverse student populations. For institutions, AI presents a pathway to greater administrative efficiency. Chatbots can handle routine student inquiries, predictive analytics can identify at-risk students for early intervention, and automation can streamline processes from admissions to grant management, freeing human staff to focus on higher-value tasks like mentorship and strategic planning. In the realm of research, AI acts as a force multiplier, enabling scholars to analyze vast datasets, accelerate discovery, and foster innovation across disciplines (Zawacki-Richter & Latchem, 2018).

1.1.2 Dominant Challenges:

Alongside these opportunities, a set of formidable challenges has emerged. The most immediate concern for many educators is academic integrity. Generative AI's ability to produce human-like text has called traditional assessment methods into question, creating an urgent need to redesign assignments and cultivate a culture of ethical AI use (Gruenhagen et al., 2024).

For students, a growing concern centers on cognitive offload, the risk that over-reliance on AI for writing, problem-solving, and analysis may erode their own critical thinking capacities and deep learning. When AI tools provide ready-made answers and polished outputs, students may bypass the cognitive struggle essential for developing genuine understanding and intellectual independence. This creates a paradox: tools designed to enhance learning may inadvertently undermine the very skills higher education seeks to cultivate (Lodge & Thompson, 2024).

Beyond the classroom, data privacy and algorithmic bias are critical ethical hurdles. AI systems trained on historical data can perpetuate or even amplify existing societal biases, potentially disadvantaging underrepresented student groups. Ensuring fairness, transparency, and accountability in AI algorithms is a paramount concern (Dai & Li, 2024).

Furthermore, a significant global digital divide persists, meaning that access to the infrastructure and literacy required for AI-enhanced education is far from universal. This creates a risk of exacerbating educational inequalities between and within nations. Finally, institutional change management presents a significant obstacle. Faculty and staff often exhibit a mix of enthusiasm and resistance, stemming from concerns about workload, job security, and the preservation of pedagogical quality. Overcoming this inertia requires thoughtful leadership, inclusive dialogue, and substantial investment in professional development (Farrokhnia et al., 2024).

1.1.3 Pervasive Policy Vacuum:

Perhaps the most defining characteristic of the current moment is the "vacuum of guidance." The pace of AI's development has outstripped the ability of most institutions to create comprehensive policies. The dominant trend, however, is a move away from prohibition and toward cautious integration. Universities are forming AI task forces, drafting ethical principles, and beginning the difficult work of creating governance frameworks that can balance innovation with responsibility (Chan, 2024). This global struggle to "catch up" underscores the critical need for proactive, evidence-informed strategic planning, a challenge that Education City is well-positioned to address.

1.2 Regional Context: The MENA Region and Qatar's Position

The Middle East and North Africa (MENA) region is actively engaging with the AI revolution, with several nations viewing AI adoption as a strategic imperative for economic diversification and global influence. The region's engagement, however, is not monolithic; it reflects a spectrum of national priorities, investment capacities, and existing educational infrastructures. AI's contribution to the MENA economy is projected to reach $320 billion by 2030, and the higher education sector is seen as a critical engine for cultivating the human capital needed to achieve this growth (PwC Middle East, n.d.).

1.2.1 The Regional Leaders: UAE, Saudi Arabia, and Qatar

A trio of Gulf nations has emerged as "advanced adopters" in the regional AI landscape, each pursuing ambitious, state-led strategies.

The United Arab Emirates (UAE) has distinguished itself with early and decisive action, appointing the world's first Minister of State for AI in 2017 and launching its National AI Strategy 2031. A flagship initiative is the establishment of the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), the world's first graduate-level, research-based AI university. This move signals the UAE's strategy of rapidly attracting top-tier global talent to build a world-class research ecosystem.

Saudi Arabia has centralized its efforts under the Saudi Data & AI Authority (SDAIA), established in 2019. Aligned with its Vision 2030, the Kingdom is making massive investments in AI to transform public services and drive mega-projects like the futuristic city of NEOM. Saudi Arabia leverages its scale and hosts high-profile events like the Global AI Summit to position itself as a major hub for AI development and policy discourse.

1.2.2 Qatar's Distinctive Approach

Amidst this dynamic landscape, Qatar has carved out a distinctive niche characterized by a deep integration of AI strategy with its long-term vision for human capital development. Qatar's National AI Strategy, launched in 2019, was co-developed by the government and academic experts at HBKU's Qatar Computing Research Institute (QCRI), ensuring from the outset that the nation's roadmap was grounded in research excellence (Qatar Computing Research Institute, 2019).

While the UAE focuses on attracting global talent and Saudi Arabia on large-scale transformation, Qatar's approach is arguably more ecosystem-focused. Its strategy is built upon the unique infrastructure of Education City, leveraging a cluster of world-class universities as a crucible for both research and talent cultivation. Qatar's comparative advantage lies in this synergy between policy, research, and education (Aboulnaga et al., 2021).

A key differentiator for Qatar is its leadership in Arabic language technologies. Through QCRI's initiatives, such as the Fanar large language model, Qatar is making strategic investments to ensure that the Arabic language and its cultural nuances are not left behind in the AI era (Fanar Team, 2025). Fanar's six specialized use cases include Taleem (Arabic for "education"), a purpose-built application specifically designed for educators, demonstrating Qatar's commitment to integrating AI into education in culturally and linguistically authentic ways. This dual focus on cultural adaptation and practical educational application represents a vital contribution to the entire Arabic-speaking world (Qatar Foundation, 2025).

Furthermore, Qatar's strategy places a strong emphasis on ethics and international collaboration (Ministry of Communications and Information Technology Qatar, n.d.). By championing responsible AI and actively participating in global dialogues through platforms like WISE, Qatar is positioning itself not just as a consumer or producer of AI, but as a thoughtful contributor to its ethical governance. Within this regional context, Education City is not merely an educational campus; it is the central pillar of Qatar's national strategy to build a sustainable, knowledge-based economy driven by human-centric artificial intelligence.

Table 1.1: Comparative table of national AI strategies

Comparative Table of National AI Strategies

	UAE	Saudi Arabia	Qatar
Launch Year	2017	2019	2019
Document	Strategy for AI	National Strategy for Data & AI	National AI Strategy
Lead Governance Body	UAE AI Office/Ministry of AI	SDAIA (Saudi Data & AI Authority)	Ministry of Communications & IT / QCRI (ecosystem)
Flagship Initiatives	Smart Dubai, AI Lab, AI in Government	NEOM AI, National Data Bank, AI Academy	Qatar Center for AI (QCAI), Arabic NLP via QCRI
Key Strategic Focus	Government services, economy, talent	Data economy, public sector AI, scaling talent	National capability, Arabic AI, education & research

Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

2 HBKU Case Study: Perceptions and Adoption of AI

2.1 Introduction to the HBKU Case Study

To ground our strategic recommendations in the lived experiences of an academic community, this report centers on a detailed case study of Hamad Bin Khalifa University (HBKU). As a research-intensive, multidisciplinary university established by Qatar Foundation, HBKU represents a microcosm of the opportunities and challenges inherent in Education City's broader ecosystem. The findings presented in this chapter are drawn from a survey of 17 faculty and 48 students across three of HBKU's colleges (Humanities and Social Sciences, Law, and Science and Engineering), conducted as part of the AI-EDAPT research project funded by the 2nd Cycle of HBKU's Signature Research Grant Program.

This in-depth analysis provides a data-driven snapshot of the on-the-ground realities of AI adoption. It reveals a community that is actively engaging with AI but is also grappling with its implications for teaching, learning, and academic values. The survey data paints a picture of two distinct but interconnected user groups: a student body that has rapidly and pragmatically embraced AI as a tool for academic productivity, and a faculty that is more measured, blending cautious optimism with significant concerns about ethics and pedagogy (Crompton & Burke, 2023). By examining their respective practices, perceptions, and anxieties, we can identify critical points of tension and alignment that must be addressed in any city-wide strategy. This case study, therefore, serves not as a definitive census of Education City but as an illustrative and powerful diagnostic tool, revealing the human dynamics at the heart of this technological transformation.

2.2 Community Perceptions and Practices: A Tale of Two User Groups

The 2025 survey data from HBKU reveals a fascinating and complex picture of AI adoption, characterized by differing patterns of use and priority concerns between faculty and students, and aligns with global trends showing that students are often early adopters of educational technologies, particularly those that can enhance their academic productivity (Kasneci et al., 2023).

2.2.1 Faculty: Cautious Optimism and Practical Application

The HBKU faculty demonstrates a high degree of awareness and engagement with AI. The vast majority are either "very familiar" (29%) or "somewhat familiar" (65%) with generative AI tools. This familiarity translates into practical application in their professional lives. Over half of the faculty respondents use AI to create course materials, and a significant number leverage it for research-related tasks, particularly for writing and editing assistance. There is a strong consensus that AI will become an integral part of higher education, with over 88% agreeing that this is definite or probable. Many also see its potential for efficiency, with 50% believing it can "significantly" save time on routine tasks.

This optimism, however, is tempered by profound worries. While faculty members recognize the potential benefits of AI technologies, they express significant concerns about their impact on academic integrity and student learning outcomes (Lodge & Thompson, 2024). When asked about integrating AI into higher education, faculty anxieties are clear and potent. The fear of plagiarism is paramount, with 93.75% expressing concern. This finding reflects broader global concerns about generative AI's impact on traditional assessment methods (Eke, 2023). This concern is not merely about rule enforcement; it represents more profound questions about what we consider knowledge, the very purpose of higher education itself, and how to maintain academic standards while adapting to new technological realities (Gruenhagen et al., 2024).

The second most important concern is about student over-reliance on AI (87.5%) and a potential loss of creativity (87.5%). These concerns directly impact their classroom management, with faculty split between encouraging cautious use, setting strict ethical guidelines, or prohibiting AI altogether. This careful stance is further reflected in a perceived lack of institutional support; a large majority of faculty feel they do not receive adequate training on how to use AI and are not fully aware of their institution's policies.

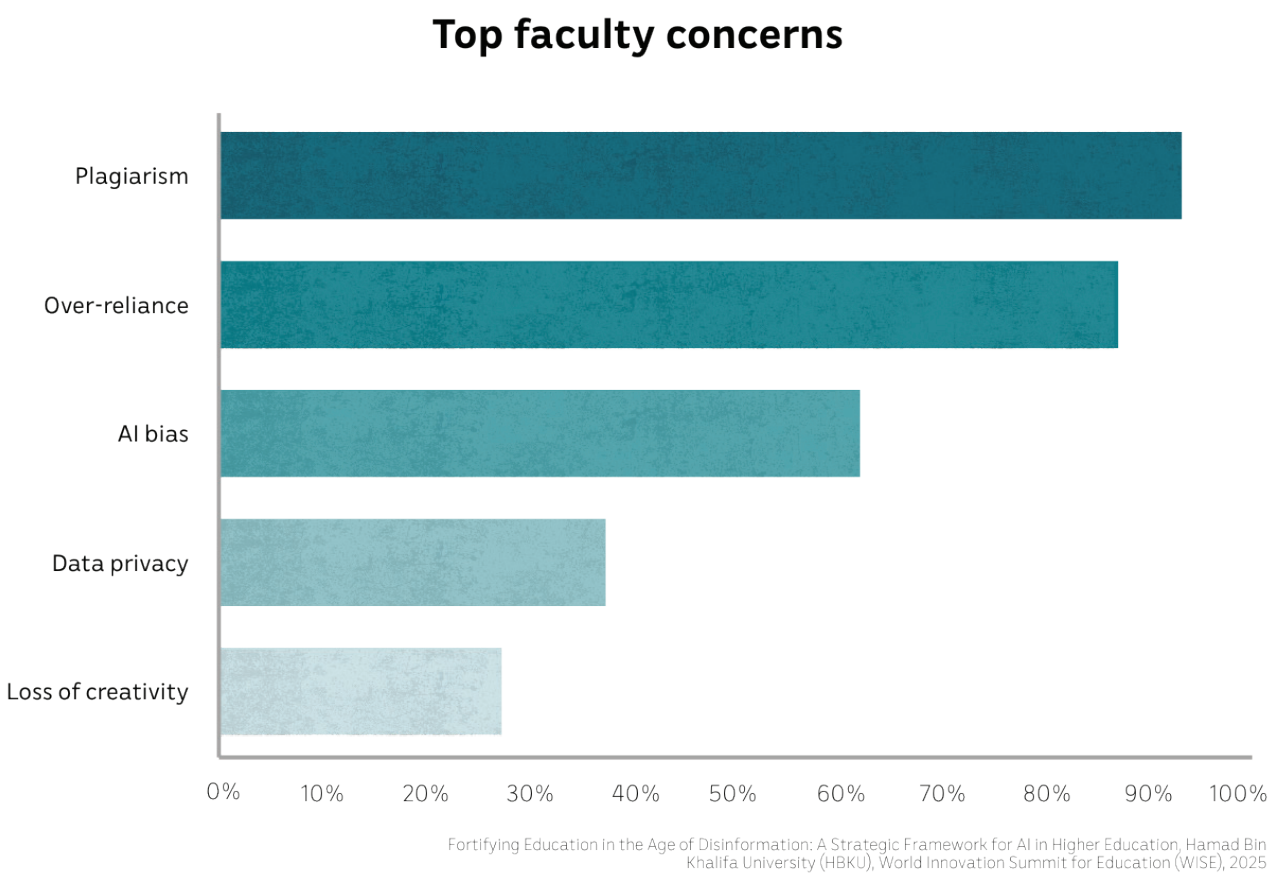


Figure 2.1: Top faculty concerns

Perhaps more significantly, the survey reveals a substantial preparedness gap among faculty. A large majority of faculty respondents reported feeling they do not receive adequate training on how to use AI in their teaching effectively. They also expressed limited awareness of their institution's existing AI policies. The explicit requests for better institutional policies, structured training, and technical support indicate that faculty members want to engage constructively with AI technologies but feel inadequately equipped to do so effectively (Al-Zyoud, 2020). This finding aligns with research on faculty professional development needs in the digital age, which emphasizes the importance of ongoing, comprehensive training programs rather than one-time workshops (Billiot, 2023).

2.2.2 Students: Widespread Adoption and Utilitarian Use

The student body at HBKU has embraced generative AI with remarkable speed and intensity. Over 75% consider themselves "very familiar" with AI technologies, and an equal percentage report using them either daily or weekly. ChatGPT is nearly ubiquitous, with 97.5% of student users reporting experience with it. This pattern reflects the versatility of modern generative AI systems and their ability to support diverse learning activities (Bozkurt, 2024)

Their use of AI is highly pragmatic and geared towards academic productivity. The most common applications are brainstorming ideas for projects (88%), summarizing research papers (73%), and writing and grammar checking (73%). This indicates that students primarily view AI as a powerful assistant that can help them manage their academic workload more efficiently.

While students are enthusiastic users, they are not naive about the risks. Their primary concern mirrors that of the faculty: over-reliance on AI (82%). However, their other top concerns diverge, focusing more on personal and ethical dimensions. Students are worried about the loss of identity and authenticity in their own work (56%), AI bias and fairness (56%), and data privacy and security (51%). This demonstrates a sophisticated awareness of the ethical complexities of these technologies. Their view on academic integrity further complicates this. At the same time, most express concern about plagiarism. Notably, 28% admit to having used AI in a way they later realized might have been an academic violation, pointing to a need for clearer guidelines.

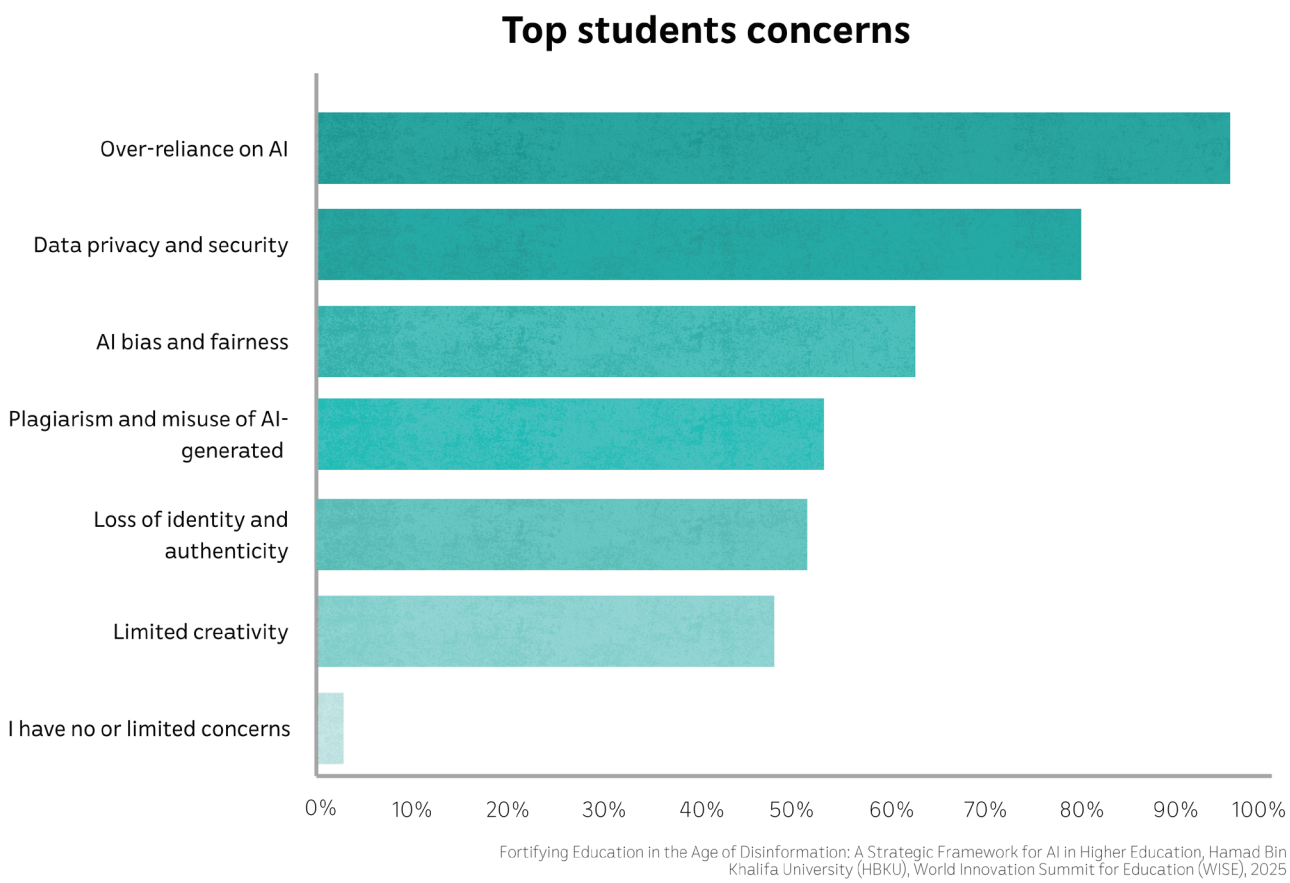


Figure 2.2: Top students' concerns

2.3 Synthesizing the Findings: Alignment, Divergence, and the Call for Action

The survey data from HBKU's faculty and students, when viewed together, provide a nuanced narrative of AI adoption within a modern university. It is a story of shared recognition of AI's inevitable role, but with differing perspectives on the most urgent risks—this synthesis of alignment and divergence points toward a clear mandate for institutional action.

2.3.1 Shared Perspectives: A Consensus on Over-Reliance

The most striking point of convergence between faculty and students is the shared anxiety about over-reliance. It is the number one concern for students and a very close second for faculty. This consensus is critical. It indicates that both educators and learners recognize that the uncritical use of AI poses a fundamental threat to the development of essential cognitive skills, such as critical thinking, problem-solving, and deep learning. This shared understanding provides a powerful foundation for developing pedagogical strategies and policies. It suggests that initiatives aimed at fostering critical AI literacy and promoting mindful, reflective use of these tools will likely be met with broad support from the entire academic community. Both groups understand that the goal is not simply to use AI, but to use it in a way that augments, rather than replaces, human intellect.

2.3.2 Divergent Priorities: Integrity vs. Authenticity

While both groups are concerned, their primary fears diverge in telling ways. For faculty, the most pressing threat is to academic integrity, with plagiarism being their foremost concern. This perspective is rooted in their role as assessors and guardians of academic standards. They are focused on the validity of student work and the fairness of the evaluation process.

Students, while also concerned about rules, frame their anxieties more in terms of personal authenticity and ethics. Their top concerns, after over-reliance, include the loss of their unique voice, the fairness of AI algorithms, and the privacy of their data. This reflects their position as creators and individuals navigating a complex digital world. They are worried not just about breaking rules, but about what using these tools means for their own identity as thinkers and the ethical implications of the systems they are using.

Notably, students themselves articulate these concerns with remarkable insight. Students warned that **"too much reliance on AI may affect critical thinking and problem-solving skills"** and emphasized the need for AI to **"work with us, not against us"** through critical engagement. Rather than resisting institutional oversight, students are actively requesting it. Multiple respondents called for clear guidelines, arguing that **"every institute should establish clear guidelines on the use of AI tools"** so that students can **"understand the appropriate extent to which these tools can be used"** while preserving their intellectual development and authenticity. As one student explained:

> "Students do not understand the long-term effects of not actually learning, thinking, or struggling with concepts. This is the key issue in my view."

This divergence is not a conflict, but a difference in emphasis. It highlights the need for a multi-faceted institutional response that addresses both the procedural issue of plagiarism and the more profound, more philosophical questions of originality, identity, and algorithmic justice.

Faculty Focus: Academic Integrity	Student Focus: Personal Authenticity
93.75% concerned about plagiarism.	56% concerned about the loss of identity in their work.
A comprehensive strategy must address both.	

Info Box 2.1: Divergent Priorities

2.3.3 A Unified Mandate for Action

Ultimately, the data from both surveys point to a single, unambiguous conclusion: the HBKU community, as a proxy for Education City, is calling for a more structured and supportive institutional framework for AI. The rapid, bottom-up adoption by students has created a reality that the top-down institutional response has yet to meet fully.

The faculty are explicitly asking for clear policies and training. Students, through their nuanced concerns and admissions of potential misuse, are implicitly asking for clearer guidelines and an education that prepares them to use these powerful tools ethically and effectively. The demand is not for prohibition but for guidance. The entire community recognizes that AI is here to stay, and they are looking to the institution to lead the way in navigating this new frontier responsibly. This unified mandate for action is the central takeaway from the HBKU case study and forms the basis for the discussion and strategic recommendations that follow.

3 Discussion: From Case Study to Strategy for Education City

3.1 Key Insights from the HBKU Experience: A Microcosm of Education City

The findings from the Hamad Bin Khalifa University case study provide more than just a snapshot of one institution; they offer a powerful lens through which to understand the systemic dynamics of AI adoption across Education City. The tensions, patterns, and needs identified at HBKU are likely being replicated in various forms at each of the partner universities. These issues align with broader global trends in educational technology adoption, where student-led innovation often outpaces institutional policy development (Chan, 2023).

The data allows us to move from abstract discussions about AI to a concrete diagnosis of the human factors shaping its integration, providing a solid foundation for a city-wide strategy.

3.1.1 The Inevitable Reality of Bottom-Up Adoption

The most significant insight is the confirmation that AI adoption is not waiting for institutional permission. It is happening now, driven primarily by students in a bottom-up, pragmatic fashion. The student survey shows near-universal familiarity and frequent, utilitarian use of AI tools to manage academic workloads. This is not a trend that can be managed through prohibition; it is a fundamental shift in the toolkit students bring to their education.

This student-led reality is outpacing the institutional response. While students have rapidly integrated AI into their daily study habits, faculty and administrators are still in a phase of cautious deliberation. This creates a critical disconnect. The university's formal structures for teaching, assessment, and academic integrity were not designed for an environment where students have constant access to powerful generative tools. The HBKU data strongly

suggests that any successful AI strategy for Education City must start by acknowledging this reality. It cannot be a top-down plan that ignores existing user behavior. Instead, it must be a strategy of engagement, one that meets students and faculty where they are and provides the necessary frameworks to channel this organic adoption toward productive and ethical ends.

3.1.2 A Community Primed for, and in Need of, Guidance

A second key insight is that the academic community is not resistant to AI, but it is deeply concerned about its responsible implementation. The survey data does not show a community in denial. On the contrary, both faculty and students display a sophisticated understanding of AI's dual nature. Faculty see its potential for efficiency, and students leverage it for productivity. Crucially, both groups share a core anxiety about the erosion of critical thinking skills through over-reliance on certain methods.

This is an optimistic finding. It means that the primary challenge is not convincing people that AI is important but rather providing them with the tools and knowledge to use it well. The explicit call from faculty for clearer policies and more training is a direct invitation for institutional leadership to take action. They are not asking for AI to be banned; they are asking for guidance on how to manage it (Barus, 2025). Similarly, students' concerns about authenticity and bias indicate a desire to be ethical actors, if only the rules of engagement were clearer. This collective desire for a structured approach represents a significant opportunity. Education City's leadership is not pushing against a closed door; it is being asked to open one and lead the way through. The community is primed for a thoughtful, well-structured AI strategy (Mahrishi et al., 2025).

3.2 Addressing the Core Challenges

The insights from the HBKU case study illuminate three critical, interconnected challenges that Education City must address to formulate a successful AI strategy: The Policy Gap, the Pedagogical Challenge, and the Preparedness Gap. Global research confirms these are universal challenges: UNESCO (2023) found that fewer than 10% of universities worldwide have formal AI policies, while studies consistently identify faculty unpreparedness and pedagogical uncertainty as primary barriers to effective AI integration (Farrokhnia et al., 2024).

Importantly, these institutional challenges directly enable the broader goals of this report. Without clear policies (Policy Gap), institutions cannot effectively govern AI's role in both creating and detecting disinformation. Without curriculum redesign (Pedagogical Challenge), students lack the critical evaluation skills essential for information resilience. Without faculty training (Preparedness Gap), educators cannot teach the media literacy competencies needed to navigate AI-shaped information ecosystems. The institutional foundations explored in this section are therefore essential prerequisites for the digital discourse tools (Part 2) and national resilience strategy (Part 3) that follow.

3.2.1 The Policy Gap: An Urgent Need for Clarity and Consistency

The HBKU data reveals a state of policy ambiguity. A significant portion of faculty were unaware of existing institutional policies regarding AI, and students admitted to potentially violating academic rules, suggesting that the guidelines are either unclear or not effectively communicated. This "policy gap" creates uncertainty and risk. Research indicates that only 20% of universities have established comprehensive AI governance frameworks, leaving the majority of institutions vulnerable to various risks, including violations of international data protection standards (such as GDPR (General Data Protection Regulation)), algorithmic bias, and accessibility failures (Ruffalo Noel Levitz, 2024). This policy vacuum creates uncertainty and inconsistency that undermines both academic integrity and educational effectiveness. For faculty and for students, it leads to inconsistent classroom rules, where an action permitted in one course is grounds for disciplinary action in another. For students specifically, it creates a confusing ethical landscape where the line between a helpful tool and an act of academic dishonesty is blurred. Specifically, it creates a confusing ethical landscape where the line between a helpful tool and an act of academic dishonesty is blurred.

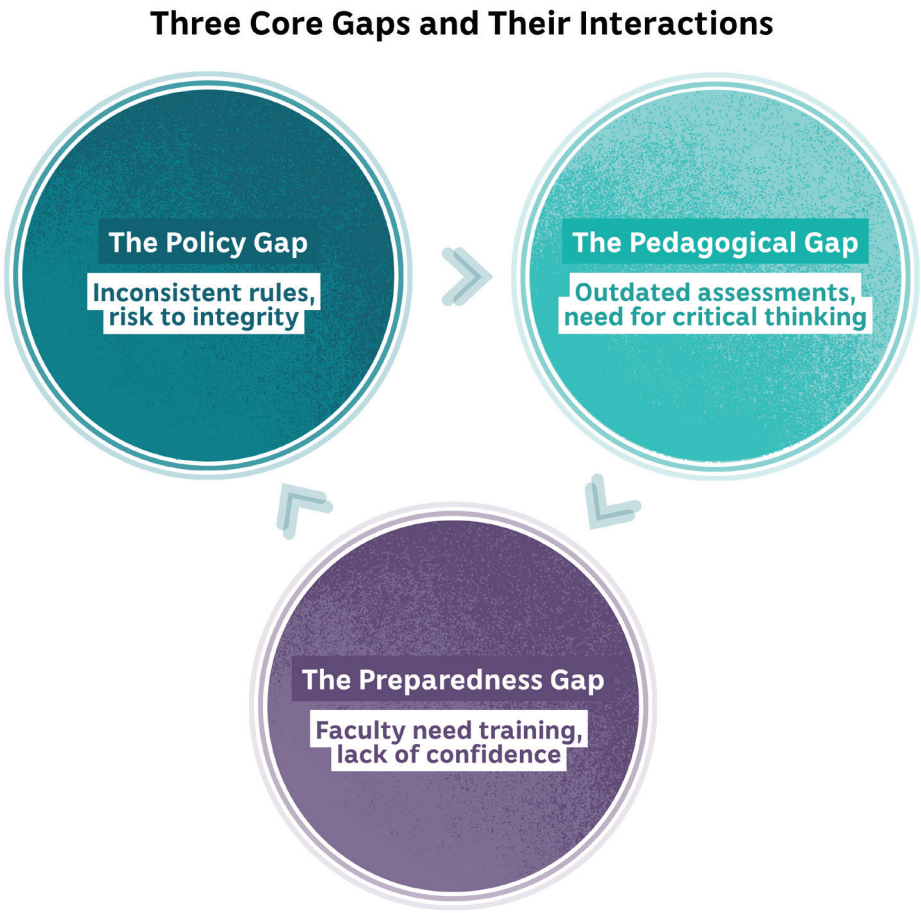
This is the most urgent challenge to address. Without a clear, consistent, and well-communicated framework for the acceptable use of AI, Education City cannot ensure fairness or uphold academic standards. The risk is not just individual cases of plagiarism, but a systemic erosion of trust in the integrity of learning and teaching, and ultimately, in the value of the degrees it offers. This gap extends beyond academic integrity to the standards of online behavior. As detailed in the sub-study 2 (below section 2), *Healthy Online Discourse*, a lack of clear policy also leaves online learning environments vulnerable to toxicity, which AI tools can both create and help mitigate.

3.2.2 The Pedagogical Challenge: Moving Beyond Fear to Redesign

The faculty's dominant concern with plagiarism, while valid, highlights a deeper pedagogical challenge. An exclusive focus on preventing cheating is a defensive posture. It treats AI as a threat to be policed rather than a new element of the learning environment to be integrated. The true pedagogical challenge is to move from a mindset of "AI-proofing" assignments to one of "AI-enriching" learning (Southworth, 2023).

This requires a fundamental rethinking of curriculum and assessment. If an assignment can be completed entirely by a chatbot, it may not be assessing the higher-order skills of critical thinking, synthesis, and creativity that are the hallmark of a university education. The HBKU data, with its emphasis on student over-reliance, suggests that both faculty and students understand this risk.

The challenge, therefore, is for educators to design learning experiences that leverage AI as a tool while demanding uniquely human contributions. This might involve tasks that require students to critique AI-generated content, use AI for data analysis, but then build a novel argument, or apply ethical reasoning to complex, real-world scenarios that AI cannot solve alone. This is a significant undertaking that requires creativity, time, and institutional support. It is, however, the most important long-term response to AI's presence in education.



Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

Figure 3.1: Three Core Gaps and their Interactions

3.2.3 The Preparedness Gap: Equipping Faculty for a New Era

The final and perhaps most foundational challenge is one of readiness. Research on professional development for AI integration emphasizes the need for comprehensive, ongoing training programs that address both technical skills and pedagogical strategies (Randhawa & Jackson, 2020). The HBKU faculty survey sends a clear message: educators do not feel adequately prepared to teach in the age of AI. They explicitly requested more training, workshops, technical support, and ethical resources. This "preparedness gap" is a major barrier to realizing AI's potential benefits.

Without confidence and competence, faculty are likely to adopt one of two extremes: either ignore AI and hope it goes away or ban it from their classrooms out of uncertainty. Neither approach serves students well. Ignoring AI leaves students without guidance, while banning it denies them the opportunity to learn how to use a technology that will be ubiquitous in their professional lives.

This gap is not limited to generative AI. The rise of sophisticated disinformation, as explored in sub-study 3 (below) by *Haqiqiqatar*, places a new burden on educators to teach media and information literacy. Closing the preparedness gap, therefore, means equipping faculty not only to teach *with* AI but also to teach *about* the digital world AI is creating. This requires comprehensive digital literacy development for both educators and students, encompassing AI literacy (understanding and using AI tools), media and information literacy (combating disinformation), and broader digital citizenship competencies. This requires a concerted, sustained investment in professional development that addresses both pedagogical techniques and critical digital literacy. Building this capacity is a prerequisite for successfully addressing the policy and pedagogical challenges. An empowered and well-supported faculty will be the most effective engine for driving thoughtful and innovative AI integration across Education City thoughtful and innovative AI integration across Education City.

3.3 The Opportunity for Education City: Leading a Human-Centric AI Transformation

The challenges laid bare by the HBKU case study, which include the policy vacuum, the pedagogical imperative, and the preparedness gap, are not insurmountable obstacles. Instead, they represent a defining opportunity for Education City to assert its leadership in the region and beyond. By addressing these issues head-on with a coordinated and forward-thinking strategy, Education City can move beyond a reactive stance and pioneer a model for AI integration that is innovative, ethical, and fundamentally human-centric.

3.3.1 From Ad-Hoc Adoption to Intentional Design

The current state, as reflected in the data, is one of ad-hoc, uncoordinated adoption. The opportunity now is to transition to a phase of intentional design. This means creating a cohesive ecosystem where AI is not just a tool that students use on the side, but an integrated component of the educational experience, guided by clear principles and aimed at enhancing specific learning outcomes (Almaraz-Menéndez & Maz-Machado, 2022). Qatar Foundation's unique structure, which brings together a diverse collection of world-class institutions, provides the ideal environment to shape and advance such a model. The ability to develop initiatives collaboratively, share best practices across campuses, and leverage centralized resources is a significant competitive advantage.

3.3.2 Championing a Human-Centric Model

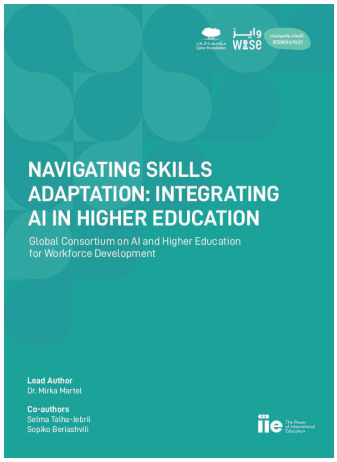
In a global landscape often dominated by purely technological or commercial approaches to AI, Education City has the opportunity to champion a different path, one that places human values and human development at its core. The shared concern for preserving critical thinking and authenticity, evident in both the faculty and student surveys, provides a powerful mandate for this approach.

A human-centric model means prioritizing the development of uniquely human skills: creativity, ethical reasoning, cross-cultural collaboration, and complex problem-solving. It means designing AI systems that augment, rather than replace, the vital role of the educator. It also means actively building resilience to the negative externalities of the digital age. This includes fostering healthy online discourse, as detailed in sub-study 2, and equipping the community with media literacy skills to combat disinformation, the focus of sub-study 3. By ensuring that technology serves pedagogy, not the other way around, Education City can offer a compelling vision for the future of higher education, one that resonates with the values of Qatar's National Vision 2030.

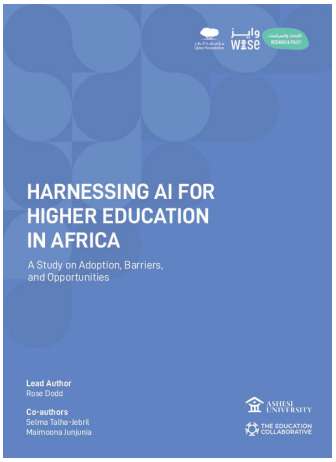
3.3.3 Becoming a Regional and Global Benchmark

By successfully navigating its own AI transformation, Education City can become a benchmark for other institutions in the MENA region and the Global South. Many universities and university networks are facing the same challenges, but may lack the resources or concentration of expertise available in Education City.

The solutions developed through this study, from the unified AI ethics charter to the cross-campus faculty development programs and innovative assessment models, can be packaged and shared as models of best practice. One of the key objectives of this WISE research study is to identify transferable insights from this Qatar-focused study that can inform or inspire implementation in other contexts. The WISE team plans to leverage the findings to engage other research and higher education partners, including the **WISE & IIE Consortium on AI, Higher Education, and Workforce Development** and the **WISE & Education Collaborative Study on AI Adoption in African Universities**.



WISE and the Institute of International Education (IIE) partnered with seven universities across the globe to explore how AI is transforming higher education for workforce development. The study offers policy recommendations and case studies on curriculum design, faculty training, and industry collaboration.



WISE and The Education Collaborative at Ashesi University (Ghana) partnered with Mohammed VI Polytechnic University (UM6P/Morocco) and AfriLabs (Nigeria) to map the adoption of AI in African higher education institutions, examining both the opportunities and the systemic barriers that shape its integration.

The tangible outputs of this entire project, including the curated Arabic toxicity dataset from sub-study two and the disinformation workshop toolkit from sub-study 3, are valuable assets that can be disseminated regionally. Through platforms like WISE, Education City can share its learnings, host crucial dialogues, and help shape a regional consensus on the responsible use of AI in education. This is not just an opportunity for institutional advancement; it is an opportunity to make a significant contribution to the global conversation and to help ensure that the AI revolution in education is equitable, inclusive, and beneficial for all. The findings of this report should therefore be seen as a call to action, an invitation for Education City to seize this moment and lead the way.

4 A Strategic Framework for Human-Centric AI Integration

Based on a comprehensive analysis of the global AI landscape and the specific, granular insights derived from the Hamad Bin Khalifa University (HBKU) case study, this chapter presents a reinforced set of strategic recommendations for Qatar Foundation and the leadership of Education City's partner universities. The recommendations are structured around four key pillars, forming the foundation of a comprehensive, human-centric AI strategy. The objective is to guide the transition from the current state of ad-hoc, bottom-up adoption to a future of intentional, coordinated, and innovative AI integration, positioning Education City as a global leader.

The HBKU survey data provides a clear mandate for action, revealing a community that is not resistant to AI but is actively seeking guidance, structure, and a clear vision for the future. Simultaneously, our extensive desktop research of the global higher education landscape underscores the urgency of this task. Universities worldwide are grappling with a significant "governance vacuum", struggling to balance innovation with profound challenges to academic integrity, equity, and pedagogy. This broader context provides not only a cautionary tale but also a rich repository of promising practices from leading institutions. By synthesizing our local, empirical data with global benchmarks, this framework offers a robust and evidence-based roadmap for navigating the complexities of AI in higher education.

Key Findings Summary

The evidence base for these recommendations draws from both local empirical data and global research findings, as derived from the desktop research:

Table 4.1: Key findings table that led the development of the strategic recommendations.

Key Findings Leading the Development of the Strategic Recommendations		
Stakeholder Group	**Key Finding**	**Implication**
Faculty	Worried about plagiarism	Need for a clear academic integrity policies
Faculty	Identify the student's overreliance on AI as the primary risk	Shared concern with students about balance
Students	Concerned about over-reliance on AI	Students seek guidance on responsible use
Students	Near-ubiquitous adoption for academic tasks	Bottom-up adoption already occurring
Both Groups	Explicitly requesting institutional policies	Clear mandate for governance framework
Both Groups	Seeking structured training and support	Need for comprehensive development programs

Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

4.1 The Four Pillars of a Coordinated AI Strategy

The proposed strategy is built on four interconnected pillars designed to create a holistic and sustainable AI ecosystem within Education City:

1. **Unified Governance and Clear Policies** to ensure ethical use, academic integrity, and consistent experience across campuses.
2. **Comprehensive Faculty and Staff Development** to build confidence, foster pedagogical innovation, and manage organizational change effectively.
3. **Integrated Curriculum and Student Literacy** to cultivate critical, responsible, and media-savvy AI users.
4. **Collaborative Research and Innovation** to establish Education City as a leader in educational AI research and development.

4.2 Pillar 1: Unified Governance and Clear Policies

The most immediate priority, confirmed by both the HBKU case study and global analysis, is to close the policy gap discussed above (section 3.2.1 The Policy Gap: An Urgent Need for Clarity and Consistency).

The HBKU survey revealed a strong desire from both faculty and students for clear institutional guidance. This local need is mirrored by a global "governance vacuum," as mentioned in the relevant UNESCO report (UNESCO, 2023). This presents Education City with an opportunity to leapfrog its peers by establishing a proactive and comprehensive governance framework.

An effective governance structure is the bedrock of a successful AI strategy. It provides the clarity needed to manage risks, the consistency to ensure fairness, and the transparency to build trust across the academic community. Without a clear framework, institutions risk a chaotic and inequitable adoption of AI, where individual instructors and students are left to navigate complex ethical and pedagogical challenges on their own. This can lead to inconsistent standards, increased risks to academic integrity, and a failure to harness the full potential of AI for learning and research.

Recommendation: Establish a Cross-Campus AI Task Force to Develop a Unified Governance Framework

Qatar Foundation should immediately convene a high-level Education City AI Task Force. This body should be more than an advisory committee; it should be an empowered working group with a clear mandate to develop and oversee the implementation of a unified AI governance framework within a 12–18-month timeframe. Its membership must be diverse, including senior academic leaders, faculty from various disciplines, IT and data security professionals, legal experts, and, crucially, student representatives to ensure that the policies are grounded in the lived experiences of the entire community.

To avoid reinventing the wheel and to align with global best practices, the task force should benchmark its work against the frameworks of leading institutions. As noted in our desktop research, universities like Stanford University, MIT, and the University of Toronto have pioneered policies emphasizing transparency, ethical review, and algorithmic fairness. The task force should analyze these models and adapt them to the unique context of Education City.

Table 4.2: Global Context and Best Practices

Global Context and Best Practices		
Institution	**Key Innovation**	**Relevance to Education City**
Stanford University	Public ethics review board for AI projects	Model for transparent governance
Massachusetts Institute of Technology (MIT)	Regular AI ethics audits with dedicated oversight	Framework for ongoing policy refinement
University of Toronto	Comprehensive ethics framework addressing bias and fairness	Template for inclusive AI policies
Carnegie Mellon University	Generative AI Teaching as Research (GAITAR) Initiative	Example of pedagogical innovation program
Various Universities (eg. Brandeis University, Texas Tech University, among others)	Integration of AI competency in tenure/promotion criteria	Strategy for incentivizing faculty engagement

Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

Key Deliverables

The three key deliverables for this task force should be:

1. **An Education City AI Ethics Charter**: This is not merely a symbolic document but a foundational text that articulates a shared set of principles for responsible AI use. It should be a public commitment to values such as fairness, transparency, accountability, and human oversight. For example, the principle of transparency would mean that any use of AI in administrative or academic decisions that significantly affect students (e.g., admissions, grading, or disciplinary actions) must be clearly disclosed. The principle of fairness would require regular audits of AI tools to ensure they are not perpetuating or amplifying biases against any group of students.
2. **A Shared Policy Toolkit for Academic Integrity**: This practical toolkit is essential to address the primary concern of faculty (93.75%), who are worried about plagiarism. It should offer a spectrum of policy options that instructors can adapt, ranging from "zero-tolerance" policies for specific courses to more permissive policies that encourage students to use AI as a tool for learning. The toolkit should include:
    a. **Model Syllabus Statements**: Clear and adaptable language that instructors can include in their syllabi to define the permissible use of AI in their courses.
    b. **Citation Guidelines:** A standardized format for how students should cite the use of generative AI in their work, promoting transparency and academic honesty.
    c. **Assessment Redesign Guidance**: Practical strategies for creating assignments that are more resilient to AI misuse and that focus on assessing higher-order thinking skills. This could include a shift towards more in-class activities, oral presentations, project-based learning, and assessments that require students to critique or analyze AI-generated content.
3. **Guidelines for Data Governance and Privacy:** In an era of data-driven education, clear protocols for the ethical use of student data are paramount. The task force must establish guidelines that are aligned with international standards like GDPR, ensuring that student data is collected and used responsibly. This includes establishing clear policies on data ownership, consent, and the use of data in AI-powered learning analytics tools. The goal is to leverage the power of data to enhance learning while protecting the privacy and rights of students.

Relationship to Broader National Infrastructure: This Education City AI Task Force is distinct from, but complementary to, the Qatar Information Resilience Observatory (QIRO) proposed in Part 3 of this report. While the Task Force focuses on institutional governance of AI within Education City (policies, pedagogy, academic integrity), QIRO operates at the national level as a research and monitoring body that detects and analyzes disinformation threats targeting Qatar. The Task Force will shape how AI is used in teaching and learning; QIRO will research information threats and provide evidence to inform national policy. Education City institutions, particularly QCRI and HBKU, will play a key role in both: leading the Task Force for institutional governance and hosting/staffing QIRO for national research. This dual role positions Education City as both a model practitioner of responsible AI integration and a research leader in information resilience.

4.3 Pillar 2: Comprehensive Faculty and Staff Development

The HBKU survey revealed a critical "preparedness gap," with faculty explicitly requesting more training and support. This is not a localized issue. Our desktop research confirms that organizational resistance and faculty unfamiliarity with AI are among the most significant barriers to successful integration globally. Overcoming this requires moving beyond ad-hoc workshops to a sustained, multi-tiered professional development program that addresses both the technical skills and the pedagogical mindset needed to thrive in an AI-enhanced educational landscape.

A successful AI strategy is contingent upon empowering educators. Faculty and staff are the key agents of change, and their confidence and competence are essential for any institutional initiative to succeed. To foster buy-in and combat the resistance identified in our desktop research, the program must be framed as a supportive measure designed to enhance, not replace, the educator's role. It should be a collaborative effort, developed with input from faculty to ensure that it meets their real-world needs.

Recommendation: Launch a Coordinated, Multi-Tiered Professional Development Program on AI in Education

Education City has established a teaching and learning support infrastructure that provides a strong foundation for this initiative. HBKU's Center for Faculty Advancement & Success in Teaching (CFAST), launched by the Office of the Provost, currently offers workshops on teaching and learning best practices, custom on-demand training, individual consultations, and classroom observations. Texas A&M Qatar's Center for Teaching and Learning (CTL), established in 2017, focuses on student learning support, engineering enrichment, and teaching innovation. VCUarts Qatar provides teaching and learning support through its Library, including a dedicated Head of Teaching, Learning, and Strategic Initiatives. Importantly, these institutions already collaborate on cross-campus faculty development. In September 2024, VCUarts Qatar organized an Education City-wide faculty panel discussion on "Generative AI in Teaching and Learning" that attracted more than 130 faculty participants from across all EC universities, demonstrating both the infrastructure for collaboration and strong faculty interest in AI-related pedagogy.

Building on this foundation and the demonstrated success of cross-campus collaboration, this recommendation calls for establishing a formal EC Faculty Development Consortium for AI in Education. This consortium should be led jointly by CFAST, CTL, and other teaching and learning units across Education City, working in close consultation with the proposed AI Task Force. The consortium's mandate is to design, coordinate, and implement a comprehensive, multi-tiered professional development program focused on AI integration in pedagogy, making it accessible to all Education City faculty, regardless of their home institution. While existing centers provide valuable general pedagogical support, this initiative requires substantial expansion in three critical areas: (1) developing AI-specific training modules and resources, (2) ensuring cross-campus consistency and accessibility, and (3) establishing sustained, multi-tiered pathways for ongoing development rather than one-off workshops.

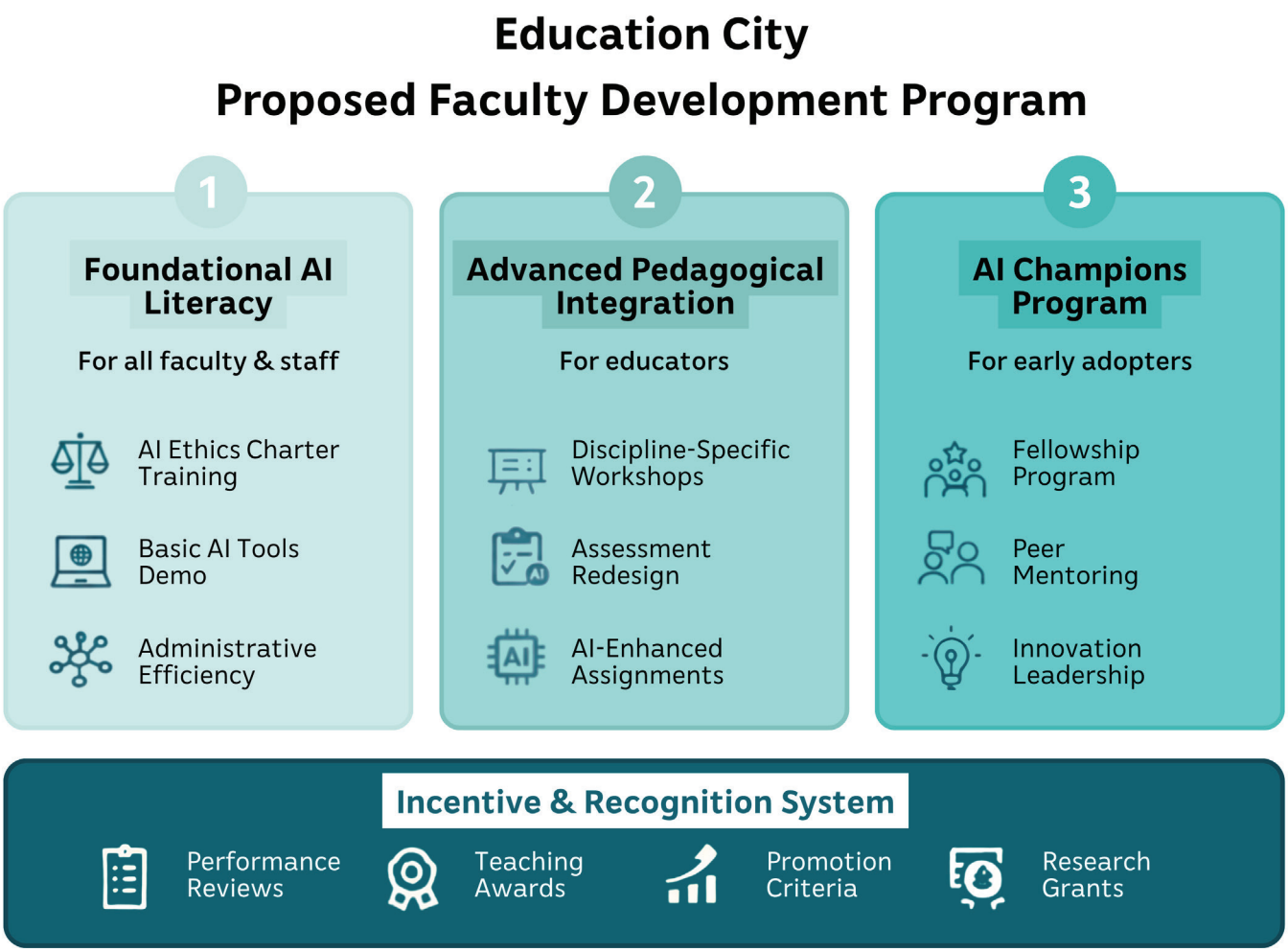


Figure 4.1: Faculty Development Program for AI in Education

Implementation Structure

The EC Faculty Development Consortium for AI in Education should operate under the following structure:

- **Governance**: Joint steering committee with representatives from CFAST, CTL, VCUarts Qatar Library, and other EC teaching and learning units, reporting to the AI Task Force.
- **Coordination**: Dedicated program coordinator position (rotating among institutions or jointly funded) to ensure consistency, manage shared resources, and facilitate cross-campus communication.
- **Resource Sharing**: Unified platform for sharing training materials, workshop recordings, best practice guides, and assessment tools across all EC institutions.
- **Faculty Network**: Establish an EC-wide "AI in Education Faculty Network" by building on the 130+ participants from VCUarts Qatar's September 2024 panel discussion "Generative AI in Teaching and Learning" to foster ongoing peer learning and collaboration. The network should facilitate monthly virtual meetups, quarterly in-person symposia rotating across EC campuses, a shared repository of teaching resources, and peer mentorship opportunities. VCUarts Qatar Library should coordinate initial network activities until formal governance is established under the EC Faculty Development Consortium.
- **Evaluation**: Common assessment framework to measure program effectiveness and continuously improve based on faculty feedback and learning outcomes.

Faculty Professional Development Program Components

1. Tier 1: Foundational AI Literacy (For All Faculty and Staff): This foundational tier ensures that every member of the academic community has a baseline understanding of AI and the institution's policies. It should be a mandatory yet engaging introduction that demystifies AI, explains the Education City AI Ethics Charter, and showcases practical examples of how AI tools can enhance productivity and teaching. For example, workshops could demonstrate how to use AI for tasks like generating course materials, creating formative quizzes, or providing initial feedback on student writing. Delivery should leverage both in-person and online formats to maximize accessibility. CFAST and CTL can alternate hosting these foundational workshops, with recordings made available to all EC faculty through a shared platform. Example workshop titles: "AI Fundamentals for Educators," "Understanding the Education City AI Ethics Charter," "AI Tools for Course Design and Assessment."

2. Tier 2: Advanced Pedagogical Integration (For Educators): This more advanced, discipline-specific tier should focus on the core pedagogical challenge of redesigning courses and assessments. These hands-on sessions should guide faculty through the process of creating AI-enriched assignments that foster higher-order skills. For instance, a history professor might learn how to design a project where students use an AI to generate a first draft of a historical narrative and then are graded on their ability to critically evaluate, correct, and enrich that draft with primary source evidence. A science professor might explore how to use AI-powered simulations to enhance laboratory learning. These sessions should be discipline-specific, with faculty from different EC institutions co-facilitating to share diverse pedagogical approaches. For example, engineering faculty from TAMUQ might partner with social sciences faculty from HBKU to explore AI integration across STEM and humanities contexts. **Recommended delivery:** Monthly discipline-specific cohorts meeting over a semester, combining workshops with peer collaboration time.

3. Tier 3: AI Champions Program (For Early Adopters): To accelerate the diffusion of innovation, this program would identify and cultivate faculty leaders in educational AI. This fellowship or certificate program would provide these "AI Champions" with advanced training, resources, and a platform to serve as peer mentors and innovation leaders within their own departments. They would become the go-to experts for their colleagues, helping to build a grassroots network of support and expertise. AI Champions should receive a micro-credential or certificate co-branded by their home institution and the EC Faculty Development Consortium, recognizing their expertise. They should be given opportunities to lead workshops for their peers, serve as consultants for course redesign projects, and contribute to the ongoing development of the program. **Annual cohort size:** 15-20 faculty across Education City.

Incentive Structures

Furthermore, as suggested by global best practices in change management, Education City should consider tying AI competency to incentive structures. This is a powerful lever for signaling institutional commitment and motivating engagement. This could include:

- Recognizing AI-related pedagogical innovation in annual performance reviews and teaching excellence awards.
- Updating promotion and tenure criteria to explicitly value contributions to digital innovation and educational technology.
- Providing small grants or course releases for faculty who wish to redesign their courses to integrate AI in innovative ways.

By investing in its human capital, Education City can ensure that AI is integrated into the curriculum in a thoughtful, ethical, and impactful manner, ultimately enhancing the quality of teaching and learning across all its institutions.

Conclusion: A Coordinated Response to a Shared Challenge

The preparedness gap identified in the HBKU survey is not unique to any single institution. It is an Education City-wide challenge that demands an Education City-wide response. By formalizing the collaboration already demonstrated through successful events like the September 2024 AI panel and by establishing a sustainable consortium model, Education City can transform faculty preparation from an ad-hoc, institution-specific effort into a coordinated, high-quality program that positions EC as a regional and global leader in AI-enhanced pedagogy. This investment in human capital is not merely about adopting new technologies; it is about empowering educators to shape how AI is used in the service of deeper learning, critical thinking, and human flourishing.

4.4 Pillar 3: Integrated Curriculum and Student Literacy

While the HBKU survey shows that students are adept, pragmatic users of AI, it also reveals a shared concern (voiced by 82% of students) about over-reliance on these tools. Students themselves are seeking a balance between leveraging AI for efficiency and preserving the authenticity of their work and their critical thinking skills. This aligns with global research on AI literacy, which emphasizes the need for comprehensive approaches that address not just technical skills but also ethical considerations and critical evaluation. The goal is to intentionally evolve students from being mere consumers of AI to becoming critical, ethical, and innovative partners with it.

Simply teaching students how to use AI is insufficient. A truly future-proof education must equip them to understand how AI works, how it fails, and how it shapes their world. This deeper literacy is essential for them to navigate a future where AI is ubiquitous, to ask critical questions about its outputs, and to use it in ways that are both innovative and responsible.

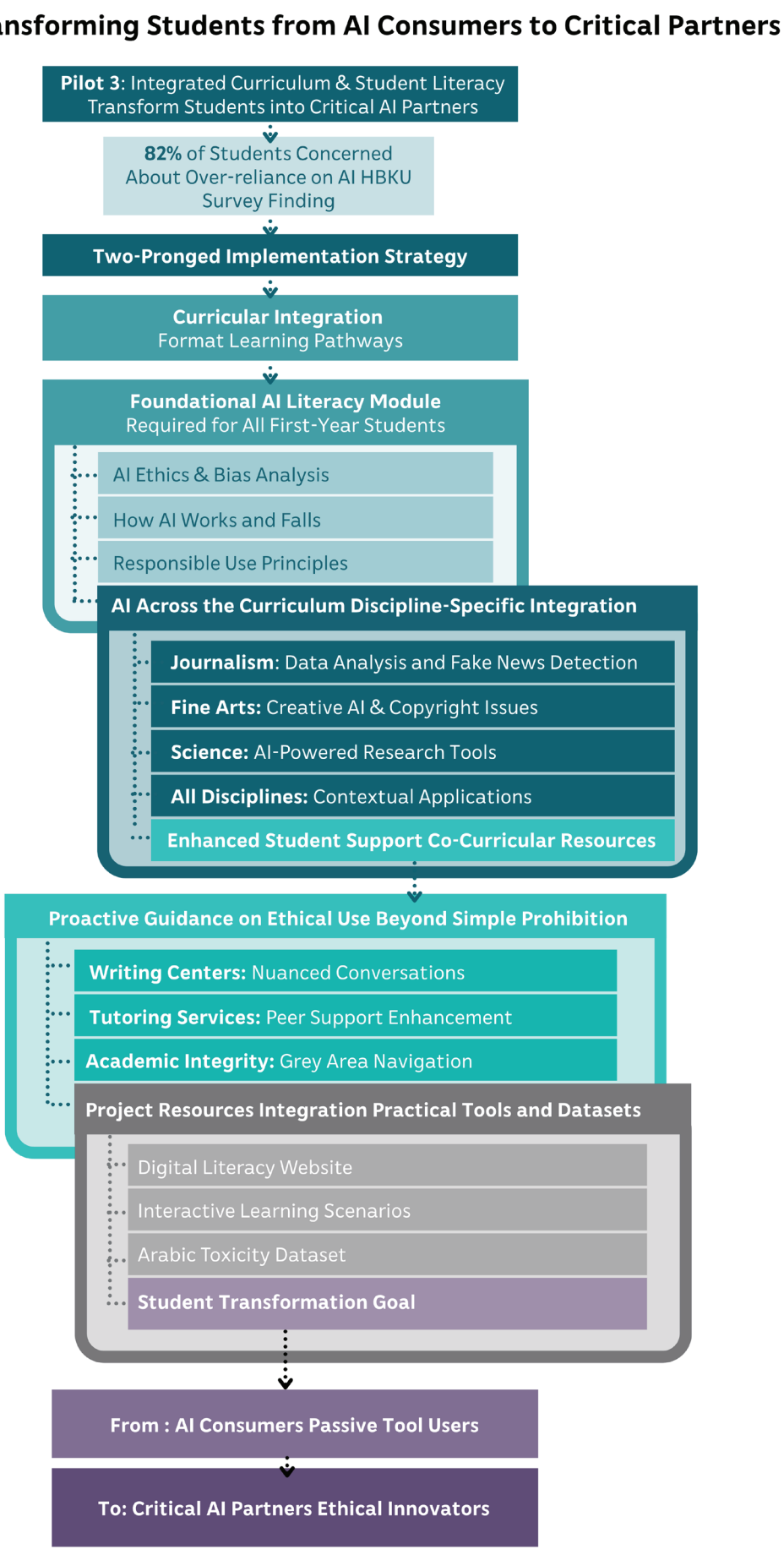
Recommendation: Integrate Critical AI Literacy into the Core Curriculum and Enhance Student Support Systems

Education City should ensure that every graduate, regardless of their major, possesses a foundational understanding of artificial intelligence and its societal implications. This requires a two-pronged approach focusing on both formal curriculum and co-curricular support.

Implementation Strategy

1. **Curricular Integration:**
    a. **Develop a Foundational AI Literacy Module:** This should be a common learning module, ideally required for all first-year students, that goes beyond a simple "how-to" guide. It should cover the basics of how AI models like LLMs and SLMs work, the ethical challenges inherent in their design and use (including bias, privacy, and disinformation), and the principles of responsible use as outlined in the Education City AI Ethics Charter. For example, the module could include interactive case studies where students analyze biased outputs from an AI and trace the potential sources of that bias in the training data. This provides a crucial foundation for critical thinking.
    b. **Promote "AI Across the Curriculum":** The goal is to make AI literacy a shared responsibility across all disciplines, ensuring it is contextualized within each student's field of study. This approach makes learning more relevant and impactful. For instance, a journalism course could integrate modules on using AI for data analysis in investigative reporting while also teaching students how to detect AI-generated "fake news." A fine arts course could explore the use of generative AI for creative brainstorming while also engaging in a critical discussion about AI, creativity, and copyright. This disciplinary context is key to fostering deep and lasting understanding.

2. **Enhanced Student Support and Resources:**
    a. **Proactive Guidance on Ethical Use:** Academic support services, such as writing centers and tutoring services, must be trained to move beyond a simple "don't cheat" message. They should be equipped to have nuanced conversations with students about how to use AI as a powerful tool for brainstorming, outlining, and revising their work without compromising academic integrity. They can become partners in helping students navigate the grey areas, fostering a culture of ethical and effective use.
    b. **Leverage Project Resources for Teaching:** The curriculum should be enriched with the practical tools developed by the broader Fortifying Higher Education research study team (WISE, HBKU, and NU-Q). The Digital Literacy Website and interactive scenarios from sub-study two can be directly embedded into first-year seminars or specific courses to provide hands-on, practical experience in navigating online discourse in AI-mediated environments. The curated Arabic toxicity dataset can become a core resource for capstone projects in computer science, linguistics, or Middle Eastern studies, creating a direct and valuable link between the research and teaching missions of the university. In collaboration with NU-Q, the WISE team will take the lead in piloting some of these resources in 2026 with the Education City community.



Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

Figure 4.2: Transforming Students from AI Consumers to Critical Partners - Pillar 3 Strategic Framework

4.5 Pillar 4: Collaborative Research and Innovation

Education City's concentration of world-class research talent and resources, as highlighted in our desktop analysis of the MENA region, provides a unique opportunity to become a leader in the study and development of educational AI. Qatar's national AI strategy, co-developed with HBKU's Qatar Computing Research Institute (QCRI), already emphasizes an ecosystem-focused approach that leverages this concentration as a crucible for research and talent cultivation. QCRI has established the campus as a recognized center of excellence through pioneering work in Arabic natural language processing, including the Fanar platform discussed earlier, alongside critical advances in machine learning, social computing, and data analytics.

The challenge now is to expand this leadership across all Education City institutions, transforming existing strengths into a truly synergistic research ecosystem. This means breaking down silos between universities and disciplines to tackle complex, cross-cutting challenges collaboratively. A successful strategy will create a virtuous cycle: local educational challenges inform a global research agenda, while global breakthroughs are adapted and applied to enrich the local environment. Such an approach moves the community beyond being merely consumers of AI technologies to becoming key contributors to their thoughtful development and ethical application.

The recommendations that follow aim to achieve this vision by building upon QCRI's strong foundation, ensuring their expertise and innovations catalyze broader interdisciplinary engagement. By creating mechanisms for all universities to contribute to and benefit from a shared research ecosystem, Education City can generate valuable knowledge and tools that enhance local education quality while making meaningful contributions to the global community.

Recommendation: Foster Collaborative, Cross-Campus Research Initiatives on AI in Education

Qatar Foundation should actively incentivize and support interdisciplinary, cross institutional research that addresses key challenges in AI and education. This will solidify Education City's position as a regional and global innovation hub, attracting talent and investment.

Key Implementation Actions

1. **Launch an "AI in Education" Seed Grant Program:** This dedicated fund would be a powerful catalyst for collaboration. It should be designed to support high-risk, high-reward projects that bring together faculty and students from multiple Education City universities. Priority should be given to interdisciplinary proposals that:
    a. **Investigate the impact of AI on learning outcomes:** Rigorous studies that move beyond anecdotal evidence to measure how different AI tools and pedagogical approaches affect student learning, engagement, and equity.
    b. **Explore new pedagogical models:** Research into how AI can enable new forms of teaching and learning, such as personalized learning at scale, collaborative problem-solving with AI partners, and AI-driven assessment for learning.
    c. **Develop culturally adapted AI tools:** A critical niche for Education City is the development of AI tools that are attuned to the linguistic and cultural context of the Arabic-speaking world. This includes not only language models but also educational content and interfaces that are relevant and effective for the region.
2. **Establish Thematic Research Groups:** To build critical mass and foster sustained collaboration, QF should facilitate the formation of cross-campus research groups focused on strategic niche areas. Drawing on existing strengths, potential themes could include:
    a. **Arabic Language AI in Education:** This group would build on and expand QCRI's pioneering work, including the Fanar model and its extensive research in Arabic NLP, to create a world-leading hub for Arabic language educational technologies.

b. **AI Ethics and Policy in the MENA Context:** This group could become a vital regional resource, developing influential research and policy recommendations that are tailored to the specific ethical, cultural, and legal contexts of the Middle East and North Africa.
c. **The Science of AI-Enhanced Learning:** This interdisciplinary group, which brings together cognitive scientists, education researchers, and computer scientists, will investigate the fundamental cognitive and social impacts of AI on students, providing the evidence base for the next generation of educational technologies.

By fostering this collaborative research ecosystem, Education City can ensure that its approach to AI in education is not only innovative but also evidence-based, ethical, and responsive to the needs of its community and the wider region.

Table 4.2: Implementation Framework and Timeline.

Phased Implementation Approach

Pillar	Phase 1 0-6 months	Phase 2 6-12 months	Phase 3 12-18 months	Phase 4 18+ months
Governance & Policy	Establish **AI Task Force**	Develop **AI Ethics Charter**	Implement policy toolkit	Monitor and refine policies
Faculty Development	Launch Tier 1 training	Roll out Tier 2 workshops	Establish **AI Champions Program**	Integrate into promotion criteria
Student Literacy	Design foundational module	Pilot AI literacy curriculum	Full curriculum integration	Assess and improve outcomes
Research & Innovation	Launch seed grant program	Form thematic research groups	Publish initial research findings	Establish EC as regional AI hub

Pillar	Resource Allocation Framework: Primary Resources Needed	Resource Allocation Framework: Estimated Investment Level	Resource Allocation Framework: Expected ROI	Sucess Metrics and Evaluation Framework: Short-term metrics 6-12 months	Sucess Metrics and Evaluation Framework: Medium-term metrics 1-2 years	Sucess Metrics and Evaluation Framework: Long-term metrics 3+ years
Governance & Policy	Task force time, legal expertise, policy development	High	**High** Foundational for all other initiatives	Task force established, charter drafted	Policies implemented across all universities	Consistent, effective governance demonstrated
Faculty Development	Training programs, release time, incentive funds	Medium	**Very high** Multiplier effect across all teaching	Training participation rates	Faculty confidence and competency assessments	Integration of AI in teaching practices
Student Literacy	Curriculum development, support service	High	**High** Pepares graduates for AI-enabled workforce	Curriculum modules developed	Student AI literacy assessments	Graduate preparedness for AI enabled careers
Research & Innovation	Seed grants, research infrastructure, collaboration	Medium	**Very high** Positions EC as a global leader	Seed grants awarded, research groups formed	Research publications and innovations	Recognition as regional AI education leader

Risk Mitigation Strategies

Pillar	Specific Risk	Mitigation Strategy	Success Indicator
Academic Integrity	Increased plagiarism through AI misuse	Clear policies, redesigned assessments, student education	Maintained or improved academic standards
Faculty Resistence	Reluctance to adopt new technologies	Inclusive development process, incentive alignment, peer support	High participation in training programs
Student Over Reliance	Diminished critical thinking skills	Balanced curriculum, ethical use training, assessment	Demonstrated critical thinking in AI mediated tasks
Implementation Gaps	Inconsistent adoption across universities	Unified governance, shared resources, regular monitoring	Consistent policy implementation across EC

Fortifying Education in the Age of Disinformation: A Strategic Framework for AI in Higher Education, Hamad Bin Khalifa University (HBKU), World Innovation Summit for Education (WISE), 2025

5 Conclusion: Seizing the Moment for Human-Centric Leadership

The evidence presented in this study section, anchored by the case study of Hamad Bin Khalifa University, leads to an unequivocal conclusion: artificial intelligence is already an integral, yet unofficial, part of the academic fabric of Education City. The era of debating whether AI will impact higher education is over. The pressing question now is *how* institutions will lead and shape this transformation.

The current landscape is defined by a strategic gap between the rapid, pragmatic, bottom-up adoption of AI by students and a more cautious, deliberative institutional response. This is not a shortcoming but a call to action. It reveals an academic community that is not resistant to change but is actively seeking guidance, structure, and a clear vision for the future. The shared concerns over preserving critical thinking and academic integrity provide a powerful, unified mandate for proactive leadership.

The four pillars outlined in our recommendations (Governance and Policy, Faculty and Staff Development, Curriculum and Student Literacy, and Research and Collaboration) offer a comprehensive roadmap. This is not merely a technical implementation plan; it is a strategy for reaffirming the core mission of higher education in the 21st century. It is about equipping students with the skills to be discerning digital citizens, empowering faculty to be innovative pedagogical leaders, and ensuring that technology serves humanistic goals.

By embracing this challenge, Education City has a unique opportunity to move beyond a reactive posture. It can pioneer a human-centric model of AI integration that becomes a benchmark for the region and the world, proving that technological innovation and the cultivation of thoughtful, ethical, and creative human beings are not mutually exclusive but mutually reinforcing goals.

A fundamental pillar of the proposed human-centric model is the cultivation of a healthy digital discourse environment. As generative AI tools accelerate the volume and velocity of online communication, the institutional gaps in policy and preparedness detailed in our analysis are brought into sharper relief, presenting a significant challenge to the quality of academic engagement. Addressing this challenge effectively requires a strategic shift from abstract principles to concrete, applied interventions. It necessitates the development of practical tools and pedagogical frameworks that empower both faculty and students to engage in online discourse constructively and ethically. In response to this imperative, the subsequent part of this report details an evidence-based strategy and a set of purpose-built resources explicitly designed for fostering a resilient and healthy digital discourse ecosystem within Education City.

PART 2: FOSTERING HEALTHY DIGITAL DISCOURSE

From Data to Literacy: Using AI Insights to Promote Healthy Online Discussions

This section addresses a critical challenge amplified by AI adoption: maintaining healthy, constructive online discourse in educational settings. Building on the institutional insights from Part 1 (above), we develop practical, evidence-based tools, including a curated Arabic toxicity dataset and a bilingual digital literacy platform, to help Education City foster respectful digital citizenship. These resources provide ready-to-deploy interventions for one of the key concerns identified by the HBKU community.

6 Introduction

6.1 Context and Rationale

The institutional analysis in Part 1 established the urgent need for clear governance and pedagogical support in an era of rapid AI adoption. A critical dimension of this challenge is maintaining the quality and safety of academic discourse as it migrates to digital spaces. Class debates that once lived only within physical seminar rooms now unfold in learning management systems and messaging channels, creating new opportunities for access and reflection but also amplifying potent hazards. The absence of nonverbal cues, combined with the cultural and linguistic diversity of Education City, means that emotionally charged topics can escalate with alarming speed. The normalization of generative AI intensifies these dynamics, increasing the volume and velocity of online interaction and making it far easier for misinterpretations to proliferate.

The central challenge is not *whether* to regulate online discourse, but *how* to do so in a way that is educationally sound, fundamentally fair across languages and identities, and credible to both students and faculty. Generic, English-centric moderation tools consistently fail in Education City's complex multilingual environment. An effective strategy must therefore combine the strengths of people and technology, using AI to support—but never replace—the reasoned judgment of trained educators. This section translates that core philosophy into a detailed, operational plan for Education City.

6.2 Purpose and Objectives

The purpose of this report is to provide senior education leaders with an actionable blueprint for fostering and maintaining healthy online discourse across our academic digital spaces. The aim is not to sanitize disagreement, but rather to ensure that rigorous and challenging debate remains compatible with the fundamental principles of learning, dignity, and safety. Four primary objectives anchor this blueprint:

1. **Establish a shared governance framework** that sets common, unambiguous expectations and ensures a trained person makes all significant moderation decisions.
2. **Build faculty and staff capacity** through a concise, scenario-based training program focused on practical skills like de-escalation and transparent communication.

3. **Embed student digital citizenship** directly into the curriculum, focusing on prevention by allowing students to practice effective response strategies in low-stakes environments.
4. **Develop customized, Arabic-native AI tools** to ensure our approach is based on our own community's data and cultural context, rather than relying on generic external tools.

These objectives are designed to support effective executive oversight and accountability. Success will be measured by clear outcomes, including: a tangible reduction of repeat toxic incidents in pilot courses, the widespread adoption of digital literacy activities, clear evidence that our custom-built tools are becoming more accurate over time, and a high completion rate of the faculty training program. In this model, measurement is not an afterthought; it is the engine of our operating model.

6.3 Methodology

The recommendations outlined in this section of the report are not based on abstract theory alone; they draw on four integrated streams of practical and scholarly work:

Research Review
We conducted a comprehensive analysis of recent scholarship on AI-assisted content moderation, tracing its evolution from traditional classifiers to modern transformer models and large language models. This review revealed persistent challenges that even the most advanced AI systems face: they consistently struggle to correctly interpret sarcasm, coded speech, and legitimate scholarly critique. These enduring failure modes demonstrate why expert human judgment remains essential, particularly in educational contexts where nuanced discussion must be protected.

Arabic Dataset Development
We built a specialized dataset of 10,000 Arabic tweets, each carefully annotated by multiple native speakers. The classification system goes beyond simple binary judgments—it distinguishes between general offensive content and identity-based hate speech, while also capturing critical contextual elements like emotional tone, sarcasm, humor, and the likely effect on readers. When annotators disagreed, we conducted formal adjudication sessions to reach consensus. This rigorous process produced data of sufficient quality for both training AI models and creating explainable case studies for faculty development programs.

AI Model Testing and Benchmarking
We established baseline performance metrics by comparing classical machine learning approaches against advanced, Arabic-specialized transformer encoders. The results show that modern transformer models achieve significantly stronger performance on offensive-content detection tasks, providing a practical and effective starting point for technical implementation within Education City's sandbox environment.

However, our testing also confirmed that hate-speech categories are much rarer and more context-dependent than general offensive content. Because of this complexity, our design philosophy mandates conservative detection thresholds and requires human review before any significant sanctions are imposed. This principle of human oversight is explicitly reflected throughout our policy recommendations.

Digital Literacy Website
We developed a bilingual digital-literacy resource (hosted on the WISE website) with three interactive components designed to build practical skills:

- An **explorer tool** that clearly defines and illustrates different types of online harm
- A **branching decision tree** that allows users to practice constructive responses to challenging scenarios
- A **safety-settings simulator** that transforms abstract privacy concepts into visible, reversible actions users can understand and control

This project is designed to be grounded in the operational realities of our institution. As a crucial next step, these tools and recommendations will be refined based on direct feedback from instructors, staff, and students to ensure they are practical, user-friendly, and address the real-world constraints of campus life.

6.4 Scope and Limitations

Our Scope: Official Academic Spaces

This strategy focuses exclusively on the online spaces owned or overseen by Education City. This includes Learning Management System (LMS) discussion boards, official course-related messaging channels, and other university-hosted forums. We will not monitor students' personal social media accounts. Our goal is to intervene strategically where the university has a clear educational mission and a duty of care to its community.

Our Guiding Principles

The strategy operates under three important principles:

1. **People Make All Final Decisions.** We firmly reject fully automated decisions. While technology can help flag potential issues, all consequential actions (for example, grade penalties or formal conduct findings) will be made by trained staff who must provide clear reasons and inform students of their right to appeal.
2. **Our Tools Will Continuously Improve.** Our initial AI models were built using social media data, which differs from classroom discussions. Therefore, we have a built-in process (the "sandbox") to safely adapt and refine these tools over time using our own de-identified data. This ensures that they become more accurate and effective for our specific environment.
3. **Fairness Must Be Actively Measured.** A tool can be accurate on average but still be unfair to specific groups. We commit to actively measuring our tools' performance to ensure they work equally well across different dialects and for students from all backgrounds. The rules for when content is flagged will be set carefully and deliberately to align with Education City's values and risk tolerance.

Within these clear boundaries, this report offers a practical and sustainable path to healthier online discourse. It is a system built on clear governance, effective training, a curriculum that prioritizes prevention, and a process for continuously improving our Arabic-specific tools to ensure they remain accountable to the community they serve.

7 Landscape for Healthy Online Discourse in Higher Education

7.1 Defining Healthy Online Discourse in Higher Education

For this report, **healthy online discourse** is an exchange that is intellectually challenging and socially respectful, where academic goals are advanced without causing personal harm.

- **Academically Rigorous:** Discussions are evidence-based and challenge ideas, not people. They are clearly tied to the course's learning objectives.
- **Respectful and Safe:** The exchange is free from personal attacks, harassment, and biased language. While vigorous disagreement is encouraged, demeaning or hurtful comments are unacceptable.
- **Inclusive and Equitable:** All participants have the opportunity to contribute, and diverse perspectives are welcomed and valued, not dismissed due to a person's identity or background.

7.2 Global Trends, Opportunities, and Challenges

Across the global higher education sector, learning now unfolds in text-first, always-on digital environments. This fundamental shift, accelerated by the recent pandemic and now cemented by ubiquitous mobile access, has redefined how academic communities form and how intellectual disagreement is expressed. At the very same time, assistive and generative AI has become a routine, almost mundane, part of study and teaching. These two powerful forces, digitally mediated conversation and AI at everyone's fingertips, are not separate trends; they are intertwined factors that shape the new landscape for online discourse. Universities that fail to see their connection will inevitably remain in a reactive, crisis-management mode. Institutions that look at them together, however, can design digital discourse ecosystems that are safer, fairer, and more educationally purposeful.

From a governance perspective, the defining global pattern is a widespread vacuum of clear guidance that institutions are now scrambling to fill. In the realm of online discourse, this lack of proactive policy has allowed ad-hoc, instructor-by-instructor rules to become the de facto standard. This approach cannot cope with the realities of cross-course digital communities or with the speed at which contentious posts can spread. In practice and as discussed in sub-study one, this produces two visible pathologies on campuses worldwide. First, students experience bewilderingly inconsistent standards, where behavior accepted in one course becomes a sanctionable offense in another. Second, faculty spend an outsized portion of their time and emotional energy adjudicating ambiguous edge cases without a shared playbook for tone, context, or escalation. The predictable consequence is uneven enforcement and a corrosive loss of trust among the very students who most need academic forums to feel safe enough to participate in challenging conversations.

Yet these same global forces are creating opportunities that did not exist even a few years ago. AI has matured from primitive keyword filters into sophisticated contextual models that can triage enormous volumes of text, cluster incidents by pattern, and surface the small handful of high-risk items that genuinely warrant human attention. When paired with clear institutional policy and carefully calibrated thresholds, these systems can significantly reduce moderator overload, identify repeat patterns that might merit a course-level intervention, and provide consistent "second opinions" that make outcomes more predictable for everyone. For students, assistive AI can be a powerful tool for improving discourse by translating, summarizing, or reformulating a heated post into more constructive language. This approach scaffolds better communication rather than merely policing bad behavior. For institutions, new instrumentation in the Learning Management System can finally connect policy to measurement, providing anonymized signals that show where discussions are thriving, where norms are degrading, and which specific interventions, such as better prompt design or targeted TA nudges, improve the climate.

Still, the challenges facing any institution are real and must be acknowledged at the executive level. The first and most challenging problem is that of context. Harm in academic conversation is rarely a simple matter of profanity. It lives in the subtleties of sarcasm, in coded dog-whistles, in scholarly quotes pulled into new and damaging contexts, and in the fraught territory of identity-laden debate. AI models trained on the blunt data of general social media consistently fail to grasp these subtleties, making them unreliable precisely when fairness and precision are most critical.

The second challenge is ensuring both the fairness and cultural fit of the moderation tools. In general, standard, English-centric AI systems demonstrate significant limitations in multilingual environments like Education City. They frequently misinterpret dialectal Arabic, miscode local references, and misclassify reclaimed language used within scholarly critique.

This technical deficiency creates a tension between core institutional values:

- **Over-blocking** of legitimate academic and dialectal speech can impede the principles of academic freedom and disadvantage members of the Arabic-speaking community.
- **Under-blocking** of targeted harassment and coded insults compromises the institution's commitment to an inclusive and safe learning environment by deterring student participation.

This dilemma highlights the necessity of developing Arabic-native capabilities, as generic tools cannot adequately balance the protection of open inquiry with the duty to maintain a safe and equitable environment for all learners.

The third major challenge is governance at scale. If the decision-support process is opaque, users will experience moderation as arbitrary and unjust, even if the average model accuracy looks impressive on a dashboard. The antidote is not to abandon automation but to make the entire system explicitly human-in-the-loop: machines triage and provide evidence, while trained people, accountable to clear and public policy, decide the significant cases and offer reasons that students can understand and trust.

A final global challenge is the significant capacity gap among educators. Most instructors were never trained to moderate online communities, de-escalate emotionally charged exchanges, or translate an ambiguous model-generated flag into an equitable and educationally sound action. Teaching assistants, who are often the first responders in class forums, receive even less preparation. Around the world, the most successful pattern for addressing this gap is converging on short, practice-heavy micro-credentials that blend case analysis with hands-on scripts for tone-setting, de-escalation, and proper documentation. This kind of practical preparation dramatically reduces variance in outcomes and, just as importantly, helps instructors explain their decisions in language that lowers defensiveness and invites repair. In sum, the global picture is paradoxical but promising. Institutions are scrambling to fill a policy vacuum even as the technical and pedagogical tools to do so have finally arrived. The winning posture is not prohibition but deliberate and thoughtful integration. The task is to make expectations explicit, train the people who lead our classrooms, instrument the LMS to learn from our own practice, and use our own Arabic-native AI as carefully governed decision support rather than as an infallible judge.

7.3 Regional Context: MENA and Qatar's Position

Within the MENA region, higher education's turn toward AI is accelerated by ambitious state-level digital strategies and a compelling demographic reality: a young, online-first population whose academic and social lives are already inseparable from digital platforms. Across the Gulf in particular, national AI programs have moved with remarkable speed from high-level vision statements to the establishment of dedicated institutions and the rollout of significant infrastructure. This is critically important for the challenge of online discourse because the same investments that fund national research labs, cloud capacity, and data platforms also create the capability to build Arabic-native safety systems and to set regional norms for their ethical use in university settings.

The regional leaders have each pursued a distinct strategy. The UAE distinguished itself with early and highly visible initiatives, including the appointment of a ministerial-level portfolio and the creation of a graduate AI university, signaling a clear talent-attraction strategy and an intent to lead in the field. Saudi Arabia's approach emphasizes national scale and centralized platforms, with a powerful data and AI authority driving large-scale programs designed to embed AI across public services and its flagship giga-projects. Amid this dynamic landscape, **Qatar has positioned itself differently: as an ecosystem builder whose primary comparative advantage is the tight coupling of policy, research, and education within a single, collaborative innovation district. This ecosystem orientation is a crucial advantage for managing discourse safety, as it enables shared assets such as datasets, models, and policy guidelines to be designed once and adopted across multiple universities.** This creates a consistent and predictable experience for students who cross-register for courses and generates significant efficiencies for instructors who may teach in more than one institution.

One feature, however, distinguishes Qatar's approach in particular: Arabic language technologies are a strategic national priority, not a mere afterthought. This commitment is far more than cultural symbolism; it is a practical precondition for achieving fairness in any AI-driven system in the region. The immense dialectal variety, morphological richness, and common use of Arabizi spellings confound generic models that are predominantly trained on English and a handful of other high-resource European languages. By investing directly in Arabic Natural Language Processing (NLP) and hosting world-class research units like QCRI with a regional mandate, Qatar creates the conditions for its universities to operate with tools that understand the language of their students. This drastically reduces the rate of false positives on reclaimed or quoted terms and significantly improves the recall rate when identity-based denigration is coded or implied rather than explicit. For an Education City audience, the message is simple: Arabic-native capability is not a boutique research interest; it is the fundamental backbone of safe, credible, and fair moderation in our bilingual classrooms.

The regional policy momentum also extends to a shared emphasis on governance ideals like fairness, transparency, and human oversight, all of which align naturally with core higher-education values. This alignment is helpful because discourse policy in a university must navigate three sensitive imperatives simultaneously:

(1) protecting open inquiry so that difficult debate is not curtailed

(2) protecting students from targeted harm so that the most vulnerable do not bear the entire cost of "free speech"

(3) protecting due process so that enforcement is seen as legitimate and fair.

In the unique context of the Gulf, where multiple universities share students within a single city-wide campus, a shared governance layer is particularly powerful. It allows for common definitions of unacceptable conduct, common escalation tiers, and common appeal routes, all of which lower confusion and build trust as students move between different courses and institutions. It also makes measurement meaningful: when there is one shared language for incidents and outcomes, institutions can compare like with like and improve their practices together.

The opportunity for Education City is to combine this top-down technical capability with bottom-up pedagogy. On the capability side, shared Arabic datasets and fine-tuned transformer models can provide a strong foundation for effective decision support. On the pedagogy side, bilingual digital-literacy resources tailored to Gulf campus life, with scenarios about cross-cultural misreading, the dynamics of group chats, and how to challenge misinformation respectfully, can move prevention upstream. When these elements are integrated into first-year courses and TA training, the result is a discourse culture that is measurably safer without becoming brittle or punitive. Qatar's unique ecosystem strengths make Education City the natural place to pioneer this integrated model for the entire region.

8 Assets and Evidence

8.1 Purpose and Overview

A credible strategy for healthier online discourse requires more than just rules; it demands shared tools that turn principles into practice. This section presents the practical, evidence-based toolkit we developed. While created for Education City, any university can adapt these assets to fit its specific context.

Our toolkit includes four core assets:

- **Clear Annotation Guidelines:** We created a detailed guide for identifying harmful content with cultural sensitivity. It provides clear definitions to distinguish between general offensive language and targeted hate speech, ensuring that human judgments are fair and consistent.

- **A Specialized Arabic Dataset:** We built a unique dataset of 10,000 annotated Arabic social media posts. This high-quality resource is the essential foundation for building our own culturally aware AI tools and creating realistic training examples for faculty.
- **AI Model Testing and Benchmarks:** We produced clear evidence showing that specialized Arabic AI models are significantly more effective at detecting harmful content than generic ones. This proves the value of our tailored approach and provides a strong basis for using AI as a decision-support tool.
- **A Bilingual Digital Literacy Platform:** We developed an interactive learning platform (hosted on the WISE website) with hands-on modules that help students practice how to respond to difficult online situations. It turns abstract safety concepts into practical, real-world skills.

These assets are designed to work together in a continuous improvement cycle. The guidelines define what we are looking for, the dataset provides the real-world evidence, the benchmarks prove our approach is effective, and the platform turns this knowledge into practical skills for our students. This creates a complete loop, moving from definition to measurement, and finally to education.

8.2 Annotation Guidelines: Ontology, Procedures, and Quality Assurance

At the core of our system is a carefully crafted three-way distinction between clean, offensive, and hate speech. This is not an arbitrary set of labels; it is a classification system designed for the specific needs of an academic environment. Offensive content includes insults, vulgarity, or demeaning language directed at individuals without reference to a broader group of identity. In contrast, hate speech is defined as attacks or demeaning content that is explicitly tied to a protected or socially salient identity. Our guidelines identify several such categories, including religion or sect; origin, ethnicity, or nationality; ideology, politics, or sports affiliation; gender; social class or profession; and physical appearance or disability. This separation is deliberately conservative. It allows for proportional institutional responses and minimizes the risk of overreach in academic settings where the quotation of sensitive texts, robust critique of ideas, and the use of reclaimed language are common and legitimate pedagogical practices.

To make human judgments both explainable and consistent, the guidelines add a layer of secondary contextual dimensions that annotators record alongside the primary classification. These include:

- Emotions and their intensity (e.g., anger, fear, disgust; rated as weak, moderate, or strong).
- The presence of sarcasm and humor flags a type of tone that often confuses automated systems and human readers alike.
- The perceived effect on readers (positive, neutral, or negative) is a dimension that is particularly useful for student reflection exercises.
- Factuality (distinguishing a verifiable claim from a rumor or an uncheckable opinion).
- The presence of spam (repetitive or promotional material).

Our procedures are designed for rigor and consistency. Annotators are instructed to read the full post, including all hashtags and emojis, before assigning a primary class, identifying any group targets, and applying the secondary labels. An explicit “unsure and comment” option is provided to prevent forced certainty in ambiguous cases; this response automatically triggers a review by a senior adjudicator. All training materials include bilingual definitions, a bank of edge-case examples (such as quoted slurs in scholarly critique or coded references in hashtags), and clear decision trees for ambiguous situations. Quality assurance is maintained through multi-annotator agreement on every item, periodic calibration sessions to ensure the team is aligned, and a fully documented adjudication process for resolving disagreements. The ultimate goal of this meticulous process is to produce labels that are simultaneously reliable enough for model training and transparent enough for human review and classroom instruction.

The strategic value of this classification system is that it maps directly onto both policy and pedagogy. Institutional policies can reference these clear definitions to create escalation paths that differ appropriately for general offensive content versus targeted hate speech. In the classroom, instructors can convert the secondary labels into powerful reflection prompts for students (e.g., "How might this post be read by someone from a different cultural background?") and into clear rubric criteria for assessing civility, the use of evidence, and the demonstration of empathy in online discussions. The same set of annotated examples can thus serve two critical purposes: as moderation practice for staff and as case-based learning for students.

8.3 Arabic Toxicity Dataset (10,000 Posts)

To ground our work in data, we built a high-quality dataset containing 10,000 Arabic social media posts that include examples of potentially harmful content. The posts were carefully selected to cover a broad range of topics and dialects. The data was then cleaned to remove duplicates and posts that were too short for reliable interpretation. Each of the 10,000 posts was independently labeled by multiple trained native speakers, with a final expert review to ensure consistency. Each post was assigned a primary category (clean, offensive, or hate speech) as well as labels for important context, such as tone, sarcasm, and emotion.

This unique dataset supports four key institutional functions:

- **Developing Custom AI Models:** The dataset is used to **fine-tune** powerful AI models. Fine-tuning is the process of adapting a general model to understand the specific nuances of the Arabic language and our academic context.
- **Informing Institutional Policy:** The data provides the evidence needed to set **intervention thresholds** (the rules for when content should be flagged) and to design an efficient **triage system** for our human reviewers.
- **Enhancing Instructional Design:** It offers a rich bank of real-world, culturally relevant examples for use in faculty training, workshops, and student assignments.
- **Ensuring Transparency:** It serves as the basis for **'dataset cards'** and **'model cards.'** These are simple, plain-language documents that explain what our data and AI tools do, how they should be used, and what their known limitations are.

8.4 AI Model Testing and Benchmarking

Our tests show that modern, Arabic-specialized AI models significantly outperform older, more generic methods, **correctly detecting 79% of harmful content** in our analysis. As expected, performance is lower on the much rarer hate-speech categories. This is a known challenge in AI caused by **class imbalance**, a situation where the system has far fewer examples of hate speech to learn from compared to other types of content.

The implication of these findings for executives is straightforward and directly informs our recommended strategy. We should use automation confidently to perform an initial **triage**, or sorting, of high-volume content. However, our policy must require meticulous human review for all consequential decisions, especially in cases involving student identity, sarcasm, or scholarly quotation.

To ensure full transparency, the dataset is accompanied by a 'dataset card' that clearly documents its purpose, scope, development process, known limitations, and the ethical considerations that guided its creation.

8.5 Bilingual Digital-Literacy Platform: From Evidence to Practice

Lasting changes in behavior in digital spaces do not come from reading a policy document; they come when learners are given the chance to practice making better decisions. Our bilingual digital-literacy platform is designed around short, interactive activities that fit naturally into first-year seminars, writing and communication courses, media literacy modules, or TA training. The pedagogical approach is guided by established learning principles, including constructivism (where learners build understanding by doing), social-cognitive modeling (through the observation of effective responses), and experiential learning (via a cycle of action, feedback, and reflection).

The module provides a clear, interactive map of the landscape of online harms. To make them easier to understand, it groups them into **three key categories:**

- **Targeted Harm to Individuals:** This includes direct attacks like **hate speech, harassment,** and **cyberbullying.**
- **Misuse of Identity and Information:** This covers the unauthorized sharing of private data (**doxing**) and deceptive practices like **impersonation.**
- **Disruptive and Deceptive Conduct:** This includes actions that derail productive conversation, such as **trolling, spreading misinformation,** and deliberately **excluding others.**

This module builds a shared vocabulary across the community and reduces confusion about what constitutes a policy violation.

- **Response Decision Tree:** This tool presents users with realistic, campus-based scenarios that branch according to their choices, providing immediate feedback on the likely social and academic consequences of each path. Prompts require users to justify their selections, and these justifications can be graded by instructors using concise rubrics. The primary aim is to rehearse the crucial habit of pausing, considering, and responding rather than simply reacting in the heat of the moment.
- **Safety Settings Simulator:** This is a hands-on sandbox environment where learners can toggle common privacy and security controls on mock social media profiles and see a "before and after" panel visualizing exactly what changes are made. This exercise makes the often-abstract concept of digital hygiene tangible, memorable, and quickly transferable to their real accounts.

The platform is fully bilingual and has been designed to accommodate Arabic dialects and the common practice of code-switching. The scenarios reflect situations that are highly relevant to campus life: cross-cultural misreading in group work, identity-laden debates in discussion boards, rumor cascades around sensitive events, and the fine line between scholarly critique and personal denigration. Instructors can parameterize the difficulty of the scenarios by adding or removing contextual cues, such as sarcasm, or by shifting from clearly unacceptable behavior to more ambiguous cases that require careful deliberation. Without storing any personal data, the platform can capture aggregate patterns of user choices and misconceptions. These patterns provide invaluable data that feeds a continuous improvement cycle, informing the creation of new instructional scenarios, the clarification of policy FAQs, and the development of targeted guidance for teaching assistants.

9 Discussion: From Evidence to Strategy

9.1 Synthesis of the Evidence Base

The evidence from our research points to three key realities that shape our recommendations:

1. **Learning Has Moved Online.** The discussion forums on our learning platforms are no longer optional extras; they are central to how students learn, form their identities, and build a sense of community.
2. **AI is Powerful but Limited.** Modern AI tools can help sort through vast amounts of text, but they consistently fail to understand the nuance of academic debate, including irony and sarcasm. They are least reliable in the complex situations that matter most to a university.
3. **We Now Have Concrete Tools to Act.** We have moved beyond theory to create a full toolkit, including a rigorous system for classifying harmful content, a specialized Arabic dataset, and a hands-on digital literacy platform for students. Given these realities, the question is no longer whether to act, but how to design a response that is both principled and practical.

   Our analysis leads to several clear conclusions that are the foundation of our strategy:

- **Automation can assist, but people must decide.** AI is useful as a filter, but it cannot serve as a judge. All important decisions must be made by trained staff.
- **Measurement must be proactive, not reactive.** We must build systems to measure the health of our online discourse from the start, rather than waiting for a crisis to occur.
- **Prevention is a matter of teaching, not just rules.** Both students and faculty need to practice making better choices in low-stakes environments before high-stakes incidents happen.
- **Fairness is about the entire system, not just the technology.** An AI tool's accuracy alone does not guarantee fairness. Our policies, the training we provide, and the appeal processes we offer are all equally important in ensuring the community sees our decisions as legitimate and just.

9.2 A Diagnostic of Gaps: Policy, Pedagogy, Preparedness & Infrastructure

In most universities, the current status quo yields four predictable and damaging institutional gaps. The first is a **Policy Gap**, which manifests as widespread incoherence. Without a common addendum for learning-management systems and other institutionally sponsored channels, individual instructors are forced to improvise their own rules. This leads to diverging escalation paths and leaves students to navigate a confusing and inconsistent set of standards as they move from one course to another.

The second is a **Pedagogical Gap**, a form of systemic underinvestment in essential skills. When digital citizenship and clear discourse norms are absent from the formal curriculum, enforcement of any policy becomes reactive and brittle. In this environment, missteps are punished but are rarely transformed into meaningful learning opportunities for the student or the broader community.

The third is a **Preparedness Gap**, which represents a significant capacity constraint. Faculty and teaching assistants, who serve as the first responders in our online forums, typically receive little to no formal preparation for the complex tasks of de-escalating conflict, documenting reasoned decisions, or distinguishing between personal offense and identity-based hate in contexts where quotation, critique, and satire are common.

The fourth is an **Infrastructure Gap**, where the digital tools used by the institution are not designed to support healthy discourse. Learning Management Systems may lack the necessary features for effective moderation, and the reliance on generic, English-centric AI models creates significant fairness and accuracy problems in our multilingual context.

Each of these gaps carries predictable and harmful consequences for the institution. Policy incoherence erodes trust and encourages "forum shopping," where participants gravitate to the spaces that least constrain them rather than those that best support genuine learning. Pedagogical underinvestment shifts an unsustainable burden onto our student conduct offices and deans, consuming valuable time and resources without ever improving the underlying campus climate. Finally, the preparedness gap fuels inconsistency in outcomes: similar cases receive vastly different treatments, and the frequent absence of clear, well-documented reason-giving invites the corrosive inference that decisions are arbitrary or biased. None of these outcomes is inevitable. They persist only because institutions have not yet redesigned their governance, teaching, and support systems for the world they actually inhabit, a world where the default medium of scholarly exchange is digital, multilingual, and socially complex.

9.3 Tensions and Trade-offs in Designing Remedies

Any effective strategy must balance competing priorities. Our approach is designed to navigate four critical trade-offs:

Open Debate vs. Student Safety

We must protect open inquiry and spirited debate, but we also have a non-negotiable duty to protect students from targeted, identity-based harassment.

- **Our Approach:** Our framework manages this by distinguishing between general "offense" and targeted "hate speech." This allows for proportional responses, ensuring we don't overreach in academic discussions where critique and controversial topics are necessary.

Automation Speed vs. Human Accuracy Automated AI can flag thousands of posts instantly, but human experts are slower, more expensive, yet far more accurate with complex context. Setting the AI filter too sensitively overwhelms our human reviewers; setting it too loosely allows harmful content to persist, eroding trust.

- **Our Approach:** We treat the sensitivity of our AI filters not as a technical decision, but as an **executive choice about institutional risk tolerance.** These settings will be reviewed quarterly by leadership to ensure they align with our values.

Central Policy vs. Local Flexibility To be fair and legitimate, we need a single, consistent university-wide policy. However, individual instructors need the flexibility to adapt the rules to their specific courses and disciplines.

- **Our Approach:** Our policy has two layers. Core principles like definitions of harm and appeal rights are universal across the institution. However, individual instructors are encouraged to adapt examples, and discussion prompts to fit their unique teaching goals.

Quantitative Data vs. Qualitative Insight While leaders rightly want clear data dashboards (e.g., number of incidents), the most meaningful changes in campus climate—like shifts in tone or increased empathy—are qualitative and harder to measure.

- **Our Approach:** We will use a **balanced approach to measurement.** We will track data dashboards while supplementing them with qualitative insights from rubric scores, student feedback, and reviews of how difficult cases were handled.

9.4 Human-in-the-Loop as Institutional Posture

To close the four gaps we've identified, our strategy is built on a core principle: **human-in-the-loop by design, not by exception.** This means that trained people, supported by technology, make all final decisions.

In practice, our automated systems are assigned a few specific and limited tasks. The system's primary role is to **triage** content, sorting it by urgency to reduce reviewer fatigue and highlight high-risk items for immediate attention. It also **summarizes** incident patterns over time, allowing leadership to shift from reactive case management to proactive interventions. Finally, it **assists in explanation** by helping reviewers generate clear, consistent rationales for their decisions, grounded in our shared policies.

Our human decision-makers, in turn, commit to several key responsibilities. They provide clear, written reasons for all significant actions in plain language. They practice **de-escalation first**, reserving formal sanctions for severe or repeated issues. Crucially, they are trained to treat borderline incidents as **teachable moments**, using them as an opportunity for student reflection and learning.

This approach aligns our governance with our educational mission. It assigns to technology the tasks it does well, like managing volume and ensuring consistency, while entrusting to our people the uniquely human skills of protecting fairness, understanding context, and upholding dignity. While a common concern is that this human-centric system will be too slow, **speed comes from clarity.** When our definitions, policies, and escalation paths are clear and our staff is well-trained, the time required to reach a fair decision falls dramatically. The goal is not to eliminate human judgment, but to make it more consistent, transparent, and just.

10 Strategic Recommendations

To address the institutional gaps diagnosed earlier, our strategic response is built on four mutually reinforcing pillars of action. Each pillar is designed to provide a direct and practical solution to one of the core challenges:

1. **Unified Governance and Policy** will aim to close the **Policy Gap** by establishing clear, predictable, and fair rules for the entire institution.
2. **Faculty and Staff Development** will aim to close the **Preparedness Gap** by equipping educators with the skills and confidence to manage online discussions effectively.
3. **Curriculum and Student Literacy** will aim to close the **Pedagogical Gap** by proactively teaching students the principles of healthy digital citizenship.
4. **A Research & Collaboration Sandbox** will aim to close the **Infrastructure Gap** by building Arabic-native, context-aware tools that are fit for our specific needs.

The recommendations that follow detail the rationale and key design choices for each of these pillars.

10.1 Pillar 1: Governance and Policy

The first pillar addresses the **Policy Gap** by creating a transparent, institution-wide governance layer. Its purpose is to enable timely, consistent, and fair action across all university-sponsored digital learning spaces. The cornerstone of this pillar is a concise **Policy Addendum** that applies to all learning platforms. This document will define categories of harm with clear examples, establish straightforward escalation tiers, and set standards for response times. To ensure consistency, it will be supported by an **Incident Classification Framework** and practical **Operational Decision Trees** that provide first-hour guidance for instructors and TAs. All automated systems will be accompanied by plain-language **Model Cards** that explain their capabilities, limits, and where human review is mandatory.

This governance model is built on the principles of due process and legitimacy, with formal appeal routes and standardized templates for any sanctions. While core definitions and rights will be uniform, the policy is designed for **disciplined local adaptation**, allowing individual programs to tailor examples to their specific academic contexts. All processes will operate within a defined ethical perimeter that guarantees **privacy by design** through data minimization and de-identification.

10.2 Pillar 2: Faculty and Staff Development

This pillar closes the **Preparedness Gap** by equipping instructors and TAs with the skills to apply policy consistently, intervene early, and make decisions that can withstand review. The core of this effort is a practice-centered **Micro-Credential** for all educators. The program uses a blended format to respect faculty time, combining a brief asynchronous primer with a live, hands-on workshop focused on real-world case studies, de-escalation techniques, and documenting decisions. A separate, streamlined track will be offered for TAs, focusing on their first-responder role.

To ensure long-term success, graduates of this program will form a network of **Discourse Mentors**, serving as a peer support resource for their colleagues. This ongoing development will be integrated into institutional structures, with completion of training linked to performance reviews and teaching assignments where appropriate.

10.3 Pillar 3: Curriculum and Student Literacy

To address the **Pedagogical Gap**, this pillar focuses on **prevention by teaching students how to recognize risky situations and choose constructive responses before conflicts escalate.** The strategy includes a mandatory 30-minute **Orientation Module** for all incoming students, which uses an interactive branching scenario and a privacy-settings simulator to build a shared vocabulary and decision-making habits from day one.

Beyond orientation, these skills will be reinforced through **Gateway Course Integrations**, where reflective assignments and bystander intervention micro-lessons are embedded into the curriculum. Grading rubrics will be designed to reward the quality of argumentation and scholarly conduct, not just agreement with a position, and will teach students the critical distinction between being offended by challenging ideas and being harmed by targeted abuse.

10.4 Pillar 4: Research & Collaboration (Infrastructure & Sandbox)

This final pillar closes the **Infrastructure Gap** by creating a lean, ethically governed research and development loop. Its purpose is to improve our Arabic-native, context-aware AI tools continuously and to generate high-quality teaching materials based on real-world challenges. This "sandbox" will be managed by a small, cross-functional team with clear roles covering the technical, data stewardship, and pedagogical aspects of the project, all under the review of an **Oversight Committee**.

The sandbox operates with strict safeguards. Its data pipeline is consent-aware and uses irreversible de-identification techniques. All AI models are stress-tested on complex cases involving sarcasm and coded language, and are used only for **decision-support (flagging for human review), not for automated blocking.** Fairness is a primary objective, with continuous monitoring for performance equity across different dialects and student groups. This ensures our technology remains accountable to the community it serves.

11 Implementation and Monitoring Framework

11.1 Implementation Plan

This framework provides a unified view of our implementation plan, linking the key activities for each pillar to its success metrics and primary risks. This integrated approach ensures that our actions are measurable and our risks are managed proactively throughout the lifecycle of the initiative.

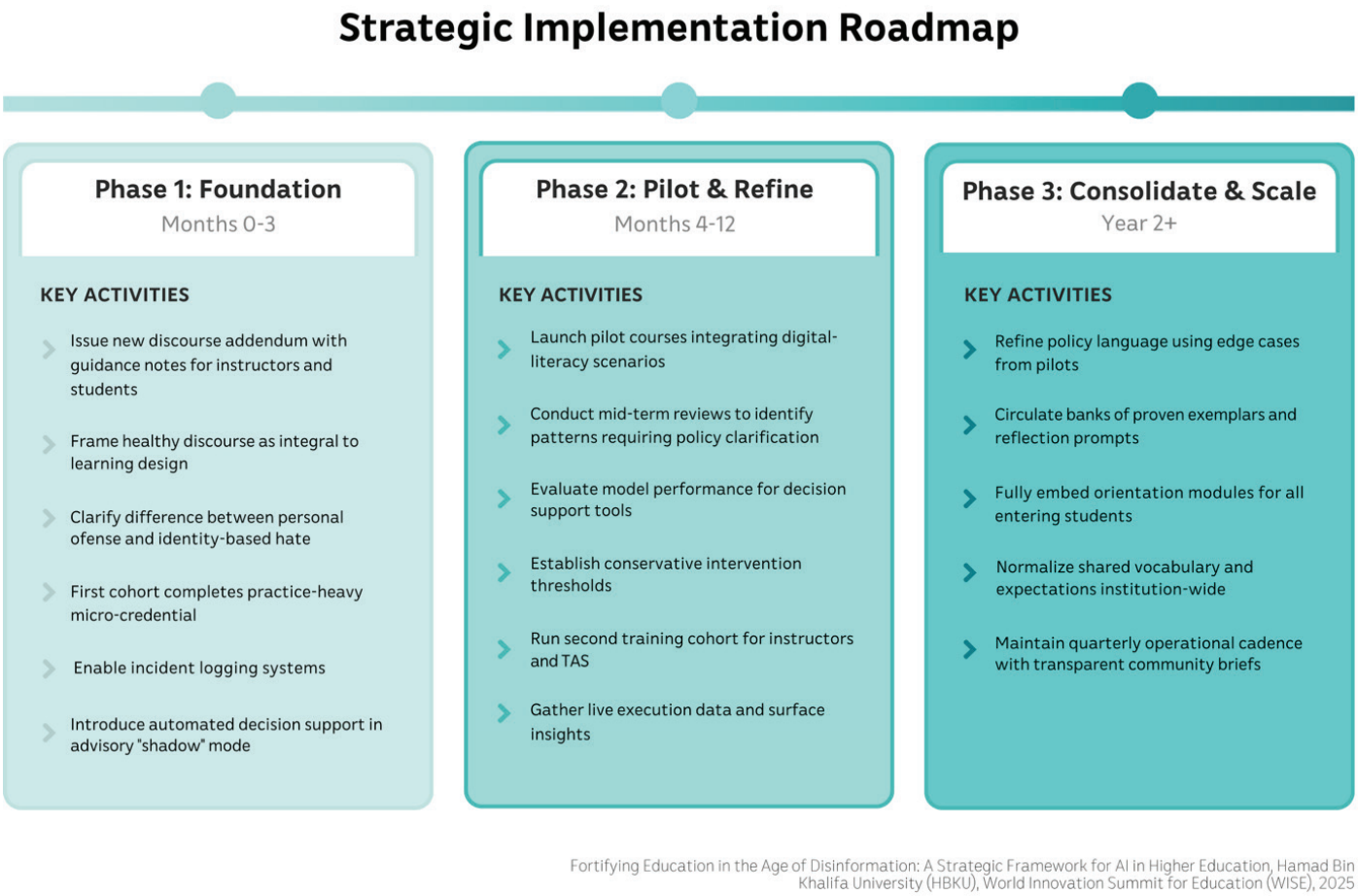


Figure 11.1: Strategic Implementation Roadmap

Below is a detailed breakdown of the activities planned for each phase.

Phase 1: Foundation (Months 0-3)

The focus will be on establishing a common operating language and capturing a clean baseline for measurement. The new discourse addendum will be issued, accompanied by concise and practical guidance notes for both instructors and students. These materials will frame healthy discourse as an integral part of learning design, explain the crucial difference between personal offense and identity-based hate, clarify the rationale for our human-in-the-loop review process, and outline the steps for escalation and appeal. In parallel, the first cohort of instructors and teaching assistants will complete the new practice-heavy micro-credential. The technical preparation will be modest: incident logging will be enabled, and any automated decision support will be introduced in a purely advisory, "shadow" mode.

Phase 2: Pilot & Refine (Months 4-12)

The emphasis will shift to live execution and data gathering. A selected set of pilot courses will integrate scenarios from our digital-literacy platform. Mid-term reviews of the pilot data will surface patterns that require policy clarification. For courses where decision support is being piloted, we will evaluate model performance and recommend conservative intervention thresholds. A second training cohort will also be run during this period.

Phase 3: Consolidate & Scale (Year 2+)

Our strategy will be to consolidate our gains before scaling further. The policy language will be refined using edge cases from the pilots. Banks of proven exemplars and reflection prompts will be circulated to instructors. The orientation modules will be fully embedded for all entering students to normalize our shared vocabulary and expectations from day one. Throughout this entire process, the operational cadence will remain quarterly, with a transparent brief shared with the community at the end of each term.

For this initiative to be successful, the assignment of clear responsibilities across four key areas is important, namely: policy, teaching, technology, and conflict resolution.

- **Policy Lead:** A Policy Lead, likely a senior faculty member (like a committee chair), will be in charge of the official policy document. Their job is to ensure the language is consistent across all colleges and to lead meetings each term to update the policy based on feedback and lessons learned.
- **Pedagogy Lead:** A Pedagogy Lead, working from a central department like a teaching and learning center, will oversee the educational content and curriculum. They will advise faculty on designing assignments and assessments and will manage a shared library of successful examples, scenarios, and grading guides for everyone to use.
- **Technology Lead:** A Technology Lead from IT or a learning technology group will manage the technical tools within our learning platform. This includes integrating any support software and maintaining the public dashboards and technical documentation.
- **Adjudicator:** Finally, an **Adjudicator**, such as an Ombudsperson or a senior student affairs officer, will be appointed to handle any formal complaints or appeals. This person will be independent of the project's daily operations to ensure a fair and impartial process.

A small steering group comprising these four leads will meet quarterly to formally ratify any proposed changes, review the key performance indicators, and commission periodic audits of the system. While this group governs the central framework, course-level moderation, the application of rubrics, and the facilitation of reflective teaching will remain the core responsibility of the instructor of record. Any action that materially affects a student's academic standing, such as a grade penalty, a formal conduct finding, or a restriction on access to university platforms, will require documented human judgment, a written rationale grounded in our shared policy definitions, and a clearly visible appeal route.

11.2 Monitoring and Evaluation

The monitoring and Evaluation (M&E) framework is designed as a system for institutional learning, not for surveillance. It is built on a foundation of settled definitions, simple and sustainable instruments, and a steady and predictable cadence. At the system level, the institution will track the rate of substantiated incidents per hundred learners per term and, more importantly, the repeat-incident rate, which is defined as a second substantiated incident by the same actor in the same term. The time-to-decision will be recorded for all escalations, and a set of adoption indicators will capture the proportion of courses that have integrated at least one digital-literacy activity, as well as the number of instructors and teaching assistants who have completed the micro-credential.

At the classroom level, instructors will apply a concise, shared rubric to a sample of student discussion posts and reflections, scoring them on the dimensions of civility, use of evidence, and perspective-taking. Short, scenario-based knowledge checks will confirm that students understand when to de-escalate a situation themselves and when they should seek help. End-of-term reflections will ask learners to narrate a moment when they changed their intended online response and explain why, providing rich qualitative evidence and seeding new teaching cases for the future.

At the model level, where we deploy decision support, their performance will be tracked through a **clear, non-technical dashboard.** This dashboard will report on three key areas:

- **Overall Effectiveness:** How accurately does the tool identify harmful content?
- **High-Risk Performance:** How well does it handle severe cases, such as hate speech?
- **Fairness and Equity:** Is the tool performing consistently and without bias across different student groups and Arabic dialects?

The trade-offs associated with different intervention thresholds will be presented in non-technical graphics so that academic leaders can weigh the risks of false positives versus false negatives in plain language. Each quarter will close with a short, public brief that documents what improved, what failed, what changed, and what remains under review, ensuring that our progress is traceable and our decisions are always explainable.

11.3 Risk, Legal, and Ethical Assurance

A comprehensive set of safeguards will be embedded in this framework from the very outset. Privacy will be protected by a strict adherence to data minimization, the use of irreversible de-identification techniques, role-based access controls, short data retention windows tied directly to our review cycles, and a documented data lineage from source to report. Due process will be guaranteed through clearly published appeal routes, the requirement of written reasons for all consequential actions, and the formal independence of the final adjudicator from day-to-day moderators.

Fairness will be treated as a systemic property rather than a simple metric. We will actively monitor for subgroup parity, our annotator calibration will be refreshed whenever local labeling occurs, our intervention thresholds will be set conservatively in all identity-linked cases, and all borderline items will be automatically routed for human review. Academic freedom will be preserved by consistently distinguishing robust intellectual critique from personal denigration, and by rewarding, through our rubrics, evidence-based argument and thoughtful perspective-taking rather than mere politeness or conformity. Periodic audits will synthesize our safeguards and outcomes in accessible language so that both students and faculty can see precisely how our principles are being enacted in practice.

12 Conclusion

12.1 What Has Been Established

This sub-study began with a simple observation: the discourse of higher education now lives substantially online. That reality brings new reach and flexibility, but it also creates new forms of harm and misunderstanding that cannot be managed by intuition or by fragmented, course-by-course rules. To address this, we assembled an evidence base that is both technical and pedagogical, including a rigorous system for classifying harmful content, a curated Arabic dataset of 10,000 annotated posts, performance benchmarks that set realistic expectations for automation, and a bilingual digital-literacy platform that turns prevention into practice.

From this foundation, we derived a strategy that treats healthy online discourse as a system grounded in shared policy, human judgment, teachable skills, and measured improvement. Our analysis surfaced the central design principle for this system: automation is valuable as a tool for triage and analysis, but a human-in-the-loop posture is the only credible approach for making consequential decisions in academic contexts.

The four strategic pillars operationalize that principle. Governance and policy provide a common language, proportionality, and due process. Faculty and staff development builds the capacity to de-escalate conflict and explain judgments. Curriculum and student literacy move prevention upstream by turning norms into practiced skills. Finally, research and collaboration maintain a "sandbox" that improves our models, documents fairness, and feeds the classroom with lessons extracted from real-world challenges. Together, these pillars are converted from intention to action by a clear implementation and monitoring framework, designed for a fast start, iterative improvement, and long-term sustainability.

12.2 The Institutional Stance

At its heart, the stance advocated here is both modest and confident. It is modest in what it claims for technology: models assist; people decide. It is confident in what a university can do that a technology platform cannot: make disagreement educative rather than corrosive, protect the dignity of its members without chilling speech, and document its reasons in language that students and faculty can interrogate. This stance is also bilingual and culturally aware by design, recognizing that fairness is inseparable from language and context, and that legitimacy depends as much on how decisions are explained as on what decisions are taken.

Three core commitments anchor this institutional posture. The first is a commitment to explainability over opacity, where every consequential action is accompanied by reasons tied to clear definitions. The second is a belief in proportionality over performative zero-tolerance, recognizing that personal offense and identity-based hate are not the same and that repair should be preferred where it can succeed. The third is a focus on learning over mere control, using measurement to improve our teaching and support systems, not to surveil our community.

12.3 What Success Looks Like

Success will not appear as the absence of conflict; universities are sites of contested ideas by design. Instead, success will manifest in tangible ways. It will look like fewer repeat incidents, because early missteps are converted into learning rather than escalation. It will look like higher-quality participation in our online spaces, because our rubrics will reward the use of evidence and perspective-taking, not just the volume of posts. It will look like shorter, clearer adjudication because our instructors and support staff will share a common vocabulary, decision aids, and de-escalation scripts. It will look like transparent fairness, because our technical work will be framed by non-technical documentation that anyone can read and challenge.

Ultimately, success will feel like predictability with care: students will know what to expect, faculty will know what to do, and leaders will have a clear dashboard to guide their next investment.

12.4 Existence of Risks and How We Manage Them

No strategic approach eliminates risk; it manages it with intention. The primary risks here are familiar: privacy leakage, over-automation, fairness disparities, and the slow drift of governance. These are controlled by design choices already embedded in our plan, including strict data de-identification, clear boundaries that forbid automated sanctions, continuous parity monitoring with conservative thresholds, and a quarterly review cadence that demands accountability.

A more subtle risk is cultural: the potential to confuse civility with conformity. The antidote is to build our assessment around the quality of reasoning, not tone alone. We must value the disciplined practice of explaining and revising one's position even when emotions run high. Our interactive platform scenarios and course rubrics are designed to enact precisely that distinction.

12.5 The Work Ahead

The first year of implementation is less about scale than it is about establishing credibility. The discourse addendum will be published, run, and experiment with the first micro-credential cohorts, instrument learning platforms with minimal friction, and pilot scenarios in a small set of courses. \Baselines will be captured, modest mid-term adjustments will be made as necessary, and the loop will be closed with a public brief showing what changed and why. In the second year, consolidate gains: refining policy with real edge cases, expanding our scenario bank, and embedding an orientation module for all entering students. Beyond year two, the measure of maturity will not be the number of tools we have but the predictability of our rhythm: definitions refreshed, training delivered, and teaching materials updated, quarter after quarter.

Over time, our research and collaboration efforts should pursue two ambitions. First, we must achieve domain alignment by fine-tuning our models on de-identified, institution-specific data, so that our decision aids reflect our own local language and norms. Second, we must produce transferable guidance, publishing patterns, and remedies in a way that peer institutions can reuse. In parallel, our pedagogical work should extend beyond first-year courses, weaving scenario-based practice into programs where public-facing communication is part of a professional identity, such as education, journalism, and the social sciences.

Healthy online discourse is not a luxury; it is a prerequisite for equitable learning in a digital university. It is also a competitive advantage. Institutions that can demonstrate human-centric governance, that can show fair and comprehensible processes, and that can teach students to navigate disagreement with skill will attract talent, retain trust, and set standards for others to follow. The plan in this sub-study is intentionally practical, built on small teams, short cycles, and clear targets. It trusts people with judgment, equips them with tools, and obliges them to explain themselves. If we hold to these commitments, the result will not be a campus without disagreement. It will be a campus where disagreement becomes a source of learning rather than harm, and where technology is used to scale our attention, not to replace it. That is the university's promise in the digital age, and it is within our reach.

Establishing robust governance over institutional digital spaces represents a foundational prerequisite. Yet, the university's mandate as a guardian of knowledge and a cultivator of critical citizenship extends beyond its campus walls. The dynamics of AI-driven manipulation and misinformation, which imperil academic discourse, are the same forces that pose a systemic threat to national security and social cohesion. The institutional capacities, governance frameworks, and pedagogical assets developed in the preceding analysis, therefore, constitute not an end in themselves but the foundation for a more expansive societal mission. Accordingly, the final part of this report scales this strategic vision from the institutional to the national level, proposing a comprehensive framework to build information resilience throughout the state of Qatar.

PART 3: BUILDING NATIONAL RESILIENCE

Haqiqatar: A Blueprint for a National Strategy to Build Information Resilience and Media Literacy

This section scales our focus from institutional strategies to national policy, proposing "Haqiqatar" (Your Truth): a comprehensive ten-year framework for building information resilience across Qatar. Drawing on the evidence from sub-study 1 and the practical tools from sub-study 2, this national strategy demonstrates how Education City's innovations can inform societal transformation. The policy recommendations, implementation frameworks, and governance structures presented in this section represent a whole-of-society approach to the challenges of AI and disinformation.

13 Introduction: The Digital Literacy Imperative

Having established the institutional foundations for AI governance within Education City in Part 1 and developed practical tools for fostering healthy online discourse in Part 2, the following section scales its focus from the university to the nation. The lessons learned and the resources created provide an evidence-based blueprint for a whole-of-society approach to the challenges of AI and disinformation.

In an era where information flows at unprecedented speed, digital literacy has become a crucial competency for civic participation, social cohesion, and national security. While this is a global imperative, the need is particularly acute in Qatar, a nation that faces a unique combination of risk factors. With one of the highest levels of internet and social media penetration in the world, its population is heavily exposed to the digital information ecosystem. Furthermore, since 2017, Qatar has been the target of sustained regional and international disinformation campaigns designed to undermine its credibility, fuel internal tensions, and damage its reputation. This context makes building "information resilience" not just a civic skill but a matter of national security.

This section of the report introduces **Haqiqatar**: a proposed ten-year National Digital Literacy & Information Resilience Strategy (NDLIRS 2026–2035). This comprehensive framework is designed to move beyond incremental fixes to provide a coordinated, systematic, and sustained response. It builds directly on the institutional insights from Part 1 and the digital literacy tools developed in Part 2, scaling these evidence-informed approaches to create nationwide resilience. The following chapters will outline the specific problems and objectives this strategy addresses, its core policy pillars, and a detailed framework for implementation and governance.

13.1 The need for media literacy on disinformation

The gravity of this educational challenge has garnered significant institutional attention. UNESCO Director-General Audrey Azoulay has sounded the alarm about "the intensification of disinformation and hate speech online, which constitutes a major threat to stability and social cohesion." (UNESCO, 2023). UNESCO emphasises that "Media and Information Literacy provides a set of essential skills to address the challenges of the 21st century, including the proliferation of mis- and disinformation and hate speech, the decline of trust in media and digital innovations, notably Artificial Intelligence."(UNESCO, n.d.) The UN has also recognized this urgency, with the UN General Assembly deciding in 2021 to commemorate Global Media and Information Literacy Week, acknowledging that digital divides "can be addressed in part by enhancing people's competencies to seek, receive and share information in the digital realm." (United Nations, n.d.)

The urgency of this challenge is further underscored by recent global research. UNESCO's 2024 opinion poll, conducted across 16 countries, revealed that 85% of citizens are worried about the impact of online disinformation, while 87% believe misinformation has already had a major impact on their country's political life. Perhaps most telling, 88% are calling on governments and regulators to address this problem through social media regulation. These statistics reflect a global population that recognizes the threat but often lacks the tools to address it individually (Henley, 2023).

The problem is compounded by declining public trust in traditional information gatekeepers. As UNESCO Assistant Director-General Tawfik Jelassi noted, "Years of research point to declining public trust in traditional media and institutions. As people increasingly turn to digital content creators for news and information, empowering professional media to lead with MIL practices can rebuild trust, enhance civic dialogue, and help citizens navigate an increasingly complex information landscape (UNESCO, 2025).

Digital literacy encompasses far more than the ability to use technology or apply simple evaluation checklists. Effective digital citizenship requires understanding the complex interplay of psychological, technological, economic, and political factors that shape our information environment. They need to understand disinformation as both an industry and a political construct. **Citizens need to grasp how algorithmic curation influences what they see, how emotional manipulation techniques target their psychological vulnerabilities, and how economic incentives drive content creation and distribution.** This systemic understanding becomes particularly crucial when addressing politically motivated disinformation campaigns that exploit existing cultural tensions and social divisions. Digital literacy education must help learners recognize how their own cognitive biases and emotional responses can be manipulated while developing resilience against sophisticated influence operations.

13.2 Context matters

The importance of digital literacy cannot be separated from questions of cultural context and equity. Disinformation challenges vary significantly across different communities, languages, and political contexts. Approaches that work effectively in one cultural setting may prove counterproductive or irrelevant in another. This reality demands culturally responsive educational approaches that acknowledge diverse epistemological traditions and community-specific vulnerabilities.

Certain populations face heightened risks and require targeted interventions. Misinformation campaigns disproportionately target minorities and older adults, while immigrant communities may encounter language barriers and unfamiliar media landscapes that increase their vulnerability. In the European example, there are increasingly state-backed operations targeting Muslim communities and firing up intercommunal tensions (Szakács & Bognár, 2021).

Ultimately, digital literacy represents a fundamental requirement for informed civic participation and social well-being in the 21st century, regardless of political system. When citizens cannot effectively evaluate information, engage in constructive discourse, or resist manipulation, their ability to make informed decisions about their lives, communities, and societies becomes compromised. Digital literacy education must therefore be understood not merely as a technical skill set but as a crucial component of civic education and social engagement across diverse political contexts.

The stakes of this educational challenge continue to escalate as information technologies become more sophisticated and manipulative techniques more refined. The emergence of deepfake technology, AI-generated content, and increasingly sophisticated social media manipulation tactics demands that digital literacy education evolve rapidly to meet emerging threats. The educational system itself faces significant challenges in addressing this crisis. Many educators lack the necessary digital and pedagogical competencies to enhance media and information literacy among their students, particularly those who are most vulnerable to misinformation. While some progress has been made in improving teachers' digital literacy, substantial deficits remain that require urgent attention. The gravity of this situation has garnered institutional recognition, with the UN publishing a report on "Countering Disinformation" in 2021 and various European bodies emphasizing the collective educational effort needed to address these shortcomings (Cernicova-Buca & Ciurel, 2022)

13.3 Why Policy Intervention is Urgent

The research makes clear that current approaches to digital literacy education are inadequate to meet the scale and complexity of contemporary challenges. The field requires coordinated action across multiple dimensions and institutions: developing rigorous research methodologies, creating adult-focused interventions, establishing pedagogical standards, and building community partnerships that can deliver culturally responsive programming at scale.

Without comprehensive, evidence-based digital literacy education, communities remain vulnerable to increasingly sophisticated misinformation campaigns that threaten public health decision-making, informed civic participation, and social cohesion. The time for incremental approaches has passed; the digital literacy crisis demands immediate, systematic, and sustained educational responses that match the urgency and complexity of our information age.

14 Problems and Objectives

14.1 The problem: digital literacy gaps and disinformation risks in Qatar

There is a global need to strengthen digital literacy, but this need is also contingent on contextual risk factors. Qatar faces a high degree of risk of exposure to disinformation. Firstly, Qatar has one of the highest levels of internet and social media penetration in the world, with residents relying heavily on platforms such as WhatsApp, Instagram, X, and TikTok for news, communication, and civic discussion (World Bank, n.d.). This high level of digital connectivity has brought opportunities for education, innovation, and participation in the global economy, but it has also created vulnerabilities.

Secondly, since the 2017 Gulf crisis, Qatar has been the subject of sustained regional and international disinformation campaigns. These efforts have sought to undermine the state's credibility, fuel tensions between different communities and tribes, and damage Qatar's reputation in the international arena. Narratives targeting Qatar have often spread through coordinated inauthentic networks, foreign media outlets, and social media influence operations, blending rumour, selective framing, and outright fabrication (Jones, 2022).

This context makes digital literacy not only a civic skill but also a matter of national security and resilience. The risks include:

- **Social cohesion**: Disinformation campaigns exploit cultural and linguistic diversity in Qatar's population, amplifying tensions and spreading rumours within and between communities.
- **Public trust**: False narratives erode confidence in institutions, experts, and reliable sources of information.
- **Health and safety**: During global crises such as COVID-19, health misinformation spreads rapidly through private messaging networks.
- **International reputation**: Coordinated disinformation targeting Qatar has influenced perceptions abroad, threatening diplomatic relations, investment climate, and even perceptions of Qataris as a people.

Despite these risks, there remains a significant gap between the complexity of the information environment and the public's ability to navigate it. Many citizens and residents lack the skills to critically evaluate sources, detect manipulation, or regulate their responses to provocative and misleading content. Educational institutions, while investing in digital transformation, are not yet fully equipped to prepare students, teachers, and the wider community for these challenges. Adult populations, including older residents and migrant communities, are especially underserved. Adapting techniques of manipulation and disinformation also means the landscape of literacy needs to be continually updated.

Without targeted and context-specific interventions, Qatar's information environment will remain vulnerable to manipulation by external actors and to the internal spread of rumors that can destabilize social trust.

14.2 Objectives: building national information resilience through literacy training

This proposed national strategy builds directly on institutional insights from sub-study 1 and the digital literacy tools developed in sub-study 2, scaling these evidence-informed approaches to create nationwide information resilience. The objective is to strengthen Qatar's information resilience: the ability of individuals, communities, and institutions to withstand and adapt to mis- and disinformation without undermining social stability or international standing.

Table 14.1: The proposed National Information Resilience Strategy

Objective	Description
Protect social cohesion	Equip citizens and residents from diverse cultural and linguistic backgrounds with the skills to identify manipulation and avoid rumor propagation.
Support national security and reputation	Ensure the public is less vulnerable to influence operations aimed at destabilizing Qatar or damaging its credibility abroad.
Strengthen educator capacity	Train teachers and community educators to deliver digital literacy in ways that are relevant to Qatar's context and multilingual population.
Reach underserved groups	Design programs for adults, migrant workers, and older residents who are disproportionately targeted by misinformation.
Institutionalize monitoring and research	Establish a permanent observatory to track disinformation campaigns against Qatar, evaluate interventions, and provide evidence for policy decisions.
Enhance crisis response	Develop protocols for quickly countering health and security misinformation in Arabic, English, and other key community languages.
Promote transparency and accountability	Secure cooperation from social media platforms operating in Qatar, including access to data on political advertising and coordinated networks.
Prepare for emerging threats	Build capacity to detect synthetic media, deepfakes, and AI-driven disinformation that may be deployed against Qatar in future regional disputes.

By 2035, Qatar should have a population with higher levels of digital and media literacy, reduced vulnerability to external and internal disinformation campaigns, and stronger public trust in credible sources of information. These outcomes will support both domestic stability and Qatar's ability to project resilience and reliability internationally.

15 Background and Evidence: The Critical Imperative of Digital Literacy in the Modern Information Age

In today's information ecosystem, where disinformation is flourishing, digital literacy is no longer a technical add-on but a civic necessity. This framing highlights that literacy is not simply the mechanical operation of technology, but the ability to think critically about information and to participate responsibly in digitally mediated societies.

The threat of disinformation is acute. The World Economic Forum's *Global Risks Report 2025* identifies disinformation as one of the most severe short-term threats to global stability, on par with climate disruption and economic volatility (WEF 2025). The UN General Assembly's 2021 decision to inaugurate Global Media and Information Literacy Week similarly underscores the extent to which international institutions now view digital literacy as essential to bridging inequalities and safeguarding civic life (United Nations, n.d.).

A growing body of research has attempted to map how academia engages with these challenges. A comprehensive review by Boler and colleagues (2025), examining 227 studies of media literacy interventions, reveals both progress and stark limitations (Boler, 2025). Much of the scholarship remains concentrated in library and information sciences, with relatively little input from psychology, public health, or communication studies, even though the phenomena at stake clearly span all of these fields. Interventions overwhelmingly target young people in schools or universities, with nearly half of studies focused on formal education settings, while only a tiny fraction, 7.5%, examined programs for adults outside classrooms. This mismatch is particularly troubling given evidence that older adults, immigrant populations, and workers in precarious information environments are often the most vulnerable to manipulation (Brashier & Schacter, 2020).

The same review highlights the absence of a consistent evaluation. While there are numerous pilot projects and innovative approaches, fewer than one in ten studies included a rigorous assessment of outcomes. Without longitudinal evidence, it is impossible to know whether the protective effects of interventions persist beyond the classroom or the initial moment of exposure. The methodological unevenness also reflects a wider lack of consensus on how to measure "success" in misinformation literacy: should the benchmark be knowledge acquisition, attitudinal change, or demonstrable behavioral resilience? The lack of clarity complicates efforts to scale up effective practices or to compare results across cultural and national settings.

The consequences of these gaps are already visible. Low levels of media and information literacy have had tangible impacts on public health, most dramatically during the COVID-19 pandemic, when misinformation about vaccines spread faster than official guidance. They have also eroded confidence in institutions and journalism. As UNESCO's Assistant Director-General Tawfik Jelassi (UNESCO, 2025) observes, years of declining trust in traditional media have coincided with a shift toward digital content creators, leaving citizens more dependent on influencers whose information practices are rarely guided by professional standards. The political consequences are equally troubling. Head et al. (2018) showed how the very term "fake news" has been weaponized as a propaganda strategy, used to delegitimize media institutions and polarize debate. The erosion of the boundary between credible journalism and manipulative content undermines the capacity for informed decision-making, weakening the foundations of civic participation and public trust.

This is not simply a Western concern. Cultural and political contexts shape how misinformation circulates and how it must be addressed. Mutsvairo and Bebawi (2019) have shown that Western conceptualizations of "fake news" often fail to resonate in Africa or the Middle East, where histories of propaganda, censorship, and political repression produce different vulnerabilities. In Qatar and the Gulf, the challenge is compounded by the multilingual environment, the heavy reliance on transnational media, and the diverse demographics of the population (see, for example, Rizwan and Hillman, 2021). Immigrant workers often encounter disinformation in their languages of origin and can lack the resources to verify claims, while older adults are frequently targeted by scams and conspiracies. These disparities underline the importance of culturally responsive pedagogy, which cannot assume that one model of critical reading or evaluation will apply everywhere.

At the same time, interventions must grapple with the political charge of misinformation. Educators face a dilemma. Some argue for confronting political disinformation head-on, acknowledging that misinformation is often explicitly designed to undermine electoral processes, polarize debate, or scapegoat minorities (Alvarez, 2021). Others recommend more indirect approaches, warning that explicit engagement with politically sensitive topics may provoke backlash or reinforce polarization (Young et al., 2021). What emerges from the research is a need for careful, context-specific strategies that recognize the political stakes without alienating learners.

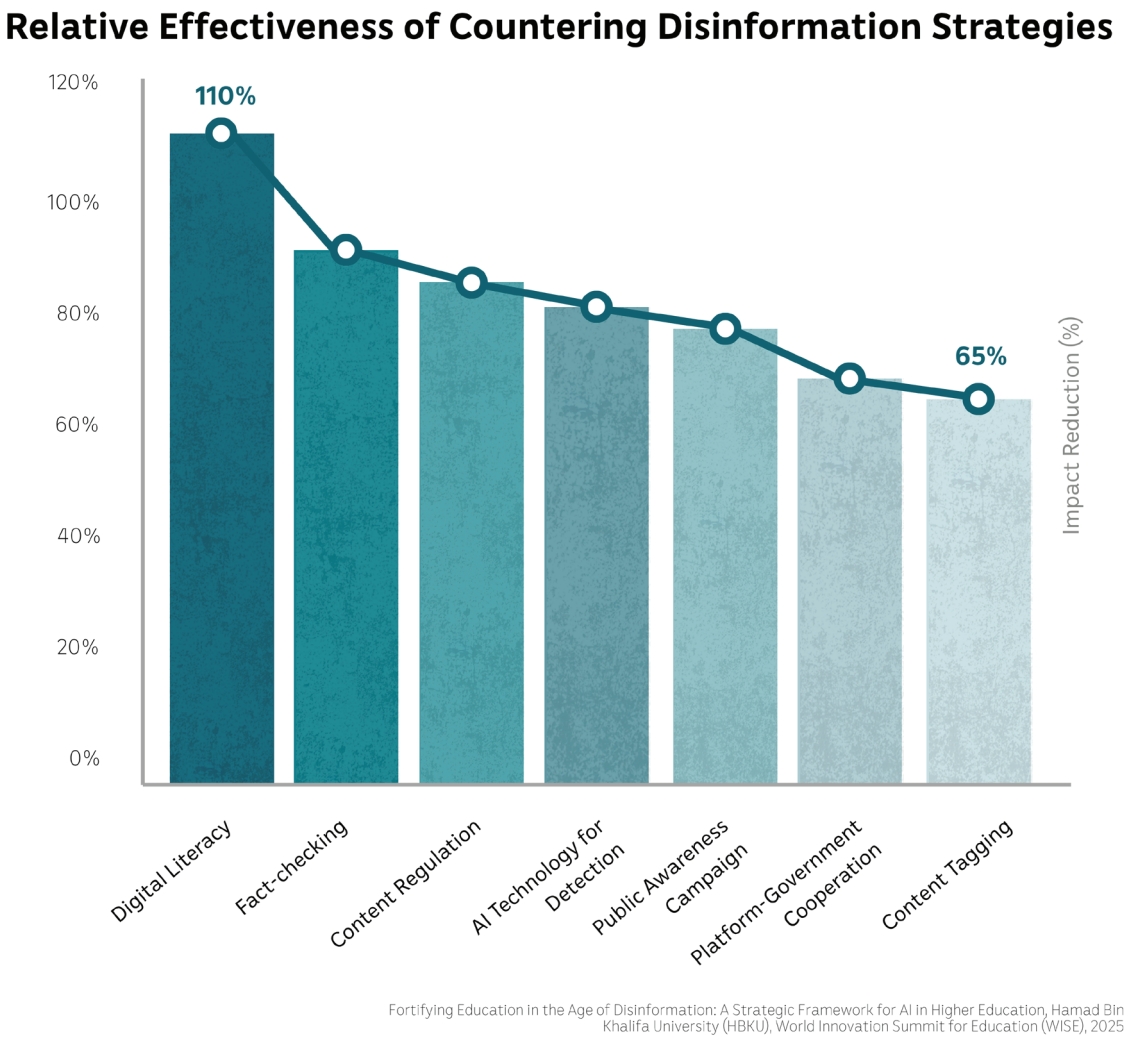


Figure 15.1: Chart showing the effectiveness of digital literacy versus other strategies. Image adapted from Surjatmodjo et al, 2024 under CC license.

The urgency of the challenge is magnified by technological change. The rise of synthetic media - including AI-generated text, deepfakes, and "slopaganda" designed to flood the information space - has intensified the difficulty of distinguishing truth from manipulation. Automated detection systems remain limited in scope, meaning that citizens' own literacy skills constitute the frontline defense. Social media companies are also increasingly taking a back seat in content moderation to protect profits. Pedagogical innovation is therefore vital.

Interactive activities, as well as gamified and role-playing interventions, such as the *Bad News Game* (Maertens et al., 2021) and *Fake It to Make It* (Urban et al., 2018), use inoculation theory to expose learners to weakened forms of disinformation tactics, thereby building cognitive resistance. Inoculation theory is a biological metaphor premised on the belief that people can be 'immunized' against persuasive information attacks in the same way their immune system can be inoculated from harmful illnesses (Compton et al, 2016). Evaluations of these approaches show promising results, with participants better able to recognize manipulation strategies after gameplay. Yet critics point out that most of these interventions remain short-term, culturally narrow, and under-evaluated (Cernicova-Buca & Ciurel, 2022).

Despite these shortcomings, emerging best practices suggest that interactive, narrative-driven, and collaborative learning approaches are consistently more effective than checklist-based or lecture-driven models. They encourage learners to reflect on their own cognitive biases, recognize how emotions are manipulated, and see themselves as active participants in civic life rather than passive consumers of information. Still, such methods remain marginal in the overall field. Boler et al. (2025) found that less than five percent of documented interventions used role-play or collaborative learning, despite their demonstrable effectiveness.

15.1 Qatar: Context is Key

While the global literature demonstrates the urgency of digital literacy, Qatar faces a set of distinctive vulnerabilities that make this challenge particularly acute. Research on Gulf information environments, including Marc Owen Jones (2020; 2022), has documented how disinformation in the region is not episodic but systematic, often linked to regional rivalries, coordinated inauthentic behavior, and state-sponsored influence campaigns.

Qatar has frequently been the target of such operations. During the 2017–2021 Gulf crisis, the country was subjected to large-scale smear campaigns across Twitter, Facebook, and satellite television, which sought to isolate Doha diplomatically and undermine public trust in Qatari institutions. False narratives around alleged terrorism financing, fabricated leaks, and manipulated news stories circulated widely, amplified by coordinated bot networks and sockpuppet accounts. Studies of these campaigns illustrate how disinformation was deployed not only to discredit Qatar internationally but to weaken its domestic resilience and fracture alliances.

As a multilingual society with a highly diverse expatriate population, Qatar also confronts challenges that differ from more homogeneous information environments. Migrant communities consume news in Arabic, Hindi, Urdu, Tagalog, and Nepali, among other languages, often through transnational media ecosystems that are difficult for Qatari authorities to monitor or regulate. This makes the spread of health misinformation, scams, and politically charged disinformation harder to contain.

The 2022 FIFA World Cup further demonstrated Qatar's vulnerability to Jones's reputational attacks. Coordinated campaigns sought to undermine Qatar's legitimacy as host, sometimes through legitimate criticism but also via fabricated stories and manipulated visuals (Jones, 2023). These episodes revealed how reputational challenges can quickly scale into diplomatic crises, amplified by digital platforms.

Case Study: The Closure of TAMUQ

The abrupt closure of Texas A&M's Qatar campus shows how disinformation can upend institutions overnight. A dubious report, amplified by ideologically driven media and lobbyists, claimed Qatar posed nuclear security risks—despite being dismissed as "insanity" by Texas A&M's own president. Yet within days, regents voted to sever a 20-year partnership, disrupting hundreds of lives.

This episode reveals how influence campaigns weaponize falsehoods to punish diplomacy, smear adversaries, and erode international credibility. Qatar's mediation in Gaza made it a target, and TAMUQ became the "lowest hanging fruit" in a broader strategy to delegitimize peace efforts.

Qatar's role as a global media hub, through Al Jazeera and other platforms, compounds these dynamics. On the one hand, it makes the country a central player in shaping global narratives; on the other hand, it increases its exposure to coordinated counter-campaigns. Literacy in this context must therefore address not only individual evaluation skills but also structural knowledge of how propaganda, algorithmic amplification, and geopolitical rivalry intersect in the Gulf.

Most recently, disinformation potentially contributed to the closure of Texas A&M's Qatar campus in 2024. The disinformation campaign was based on a hostile information campaign linked to Qatar's role in mediating a ceasefire in Gaza. For Qatar, these dynamics highlight the need for literacy programs that do more than teach individuals to "spot fake news." They must build systemic resilience by helping citizens, residents, and institutions understand how disinformation is embedded in geopolitical conflict and sustained over time. Qataris must foster 'diligent doubt' and informed skepticism about narratives circulating in the media.

Improving literacy against disinformation should not be a trivial afterthought in education, but a core requirement for civic participation and social cohesion in the twenty-first century. Without it, societies are more vulnerable to health crises, polarization, and influence operations that exploit cultural divisions. The research shows that interventions can work, but the field remains fragmented, Western-centric, and inattentive to adults and marginalized groups. Qatar, as a middle power with a multilingual and diverse population, faces both heightened risks and unique opportunities. By embedding equity, cultural responsiveness, and rigorous evaluation into a national strategy, Qatar can lead internationally in shaping digital literacy as a cornerstone of information resilience.

16 Policy Framework: National Digital Literacy & Information Resilience Strategy (NDLIRS 2026–2035)

Digital Literacy requires concerted coordination and strategic thinking to ensure that media literacy is constantly adapting to the country's needs and the rapidly changing digital landscape. The proposed National Digital Literacy & Information Resilience Strategy (NDLIRS) would provide a coordinated five-year plan (2026–2035) to strengthen Qatar's ability to withstand and respond to misinformation and disinformation. It builds on international best practice but is tailored to Qatar's unique context: a highly connected society with one of the world's highest social media penetration rates, a diverse and multilingual population, and a history of being the target of sustained disinformation campaigns, particularly since 2017.

The framework is not a single program but a long-term, whole-of-society approach. It is designed to align schools, universities, community organizations, government ministries, and media institutions around a common goal: building national information resilience.

16.1 Guiding Principles

- **Equity and Inclusion:** All communities in Qatar must benefit from digital literacy efforts. This means targeted interventions for groups disproportionately affected by disinformation, including older adults, migrant workers, and those with limited literacy. It also means ensuring access in multiple languages, across a variety of platforms, and for different levels of digital competence.
- **Cultural Responsiveness:** Effective programs must resonate with Qatar's social and cultural realities. Educational content should reflect local traditions and values, while being adaptable to the country's multilingual, multicultural environment. Social responsiveness also ensures that digital literacy programming does not simply import foreign models but addresses the specific ways disinformation operates in Qatar and the wider Gulf region.
- **Evidence-Informed Practice:** All initiatives must be guided by research and subject to systematic evaluation. Qatar has an opportunity to set regional standards by embedding monitoring and evaluation in every program. This requires collaboration between universities, research institutes, and government bodies to ensure that interventions are tested, refined, and scaled based on measurable outcomes.

16.2 Policy Pillars Overview

To translate these principles into practice, the NDLIRS is organized around seven interlocking pillars. Each pillar addresses a critical dimension of the problem, and together they create a coherent framework for building Qatar's information resilience.

- **Educator Upskilling and Emerging Threats** - Teachers and community educators will be trained to integrate digital literacy into both formal and informal learning environments. This includes core competencies such as critical evaluation, algorithmic awareness, emotional regulation, and lateral reading. Importantly, training will also address emerging threats by equipping educators to teach recognition of AI-generated content, deepfakes, manipulated images, and synthetic audio. By embedding these topics into the national education system and community programs, Qatar can prepare citizens for both current and future information challenges.
- **Adult & Vulnerable Populations** - Many of Qatar's most at-risk groups lie outside formal education. Targeted interventions will focus on older adults, migrant workers, and multilingual communities. Programs will be delivered through community hubs, libraries, workplaces, and mobile outreach units. Content will be adapted for different literacy levels and delivered in relevant languages, including Arabic, English, Hindi, Nepali, Tagalog, and Urdu.
- **Platform Transparency & Data Access** - Qatar will work with digital platforms to increase accountability and transparency. This includes establishing advertising libraries for political and issue-based content in both Arabic and English and securing access to anonymized data for accredited researchers. Such cooperation is essential for tracking coordinated influence operations and evaluating the impact of interventions.
- **Public Service Media & Information Health** - National awareness campaigns will promote verification practices, rumor-checking, and responsible information sharing. Campaigns will be co-designed with trusted figures, including journalists, healthcare professionals, religious leaders, and social media influencers. Qatar's established media sector, including public broadcasters and global platforms such as Al Jazeera, provides an opportunity to amplify these efforts domestically and internationally.
- **Research & Threat Observatory** - A permanent, interdisciplinary observatory will monitor disinformation campaigns targeting Qatar, conduct applied research, and evaluate interventions. The observatory will publish quarterly briefs, develop early-warning systems, and provide evidence directly to policymakers. It will serve as a hub for collaboration between Qatari universities, think tanks, and international partners.
- **Crisis & Health Misinformation Protocols** - During health emergencies, security incidents, or politically sensitive events, misinformation spreads rapidly. Qatar will develop rapid response protocols to ensure timely, multilingual communication. Pre-prepared rumour rebuttal templates, coordinated spokesperson networks, and partnerships with local media will reduce the impact of harmful narratives.
- **Standards & Quality Assurance** - To ensure that digital literacy programs are effective and scalable, Qatar will introduce a national framework for evaluation and certification. All government-funded initiatives will be required to demonstrate clear learning outcomes, use culturally responsive pedagogy, and include independent evaluation. This quality assurance system will prevent duplication, improve efficiency, and build a consistent evidence base for what works.

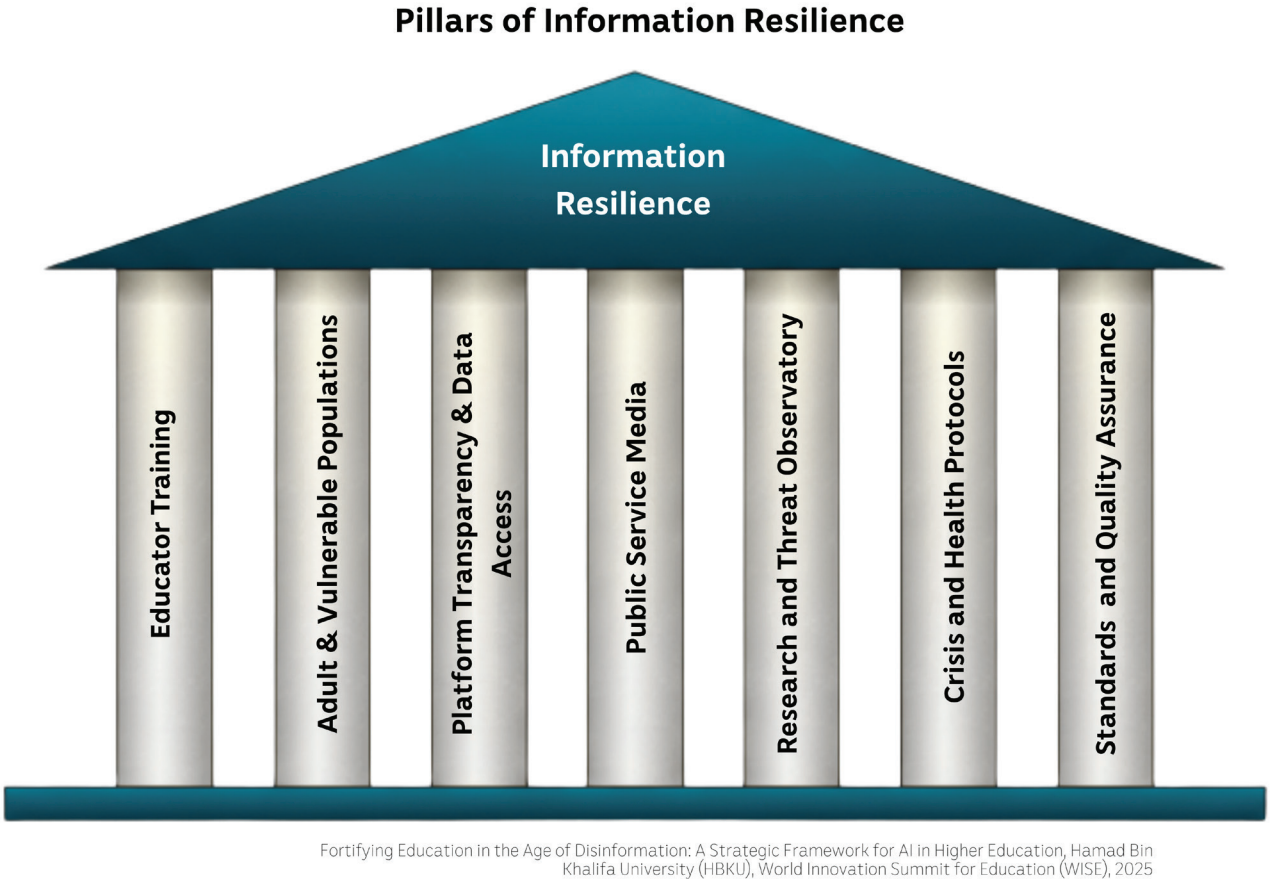


Figure 16.1: The Pillars of Information Resilience.

17 National Digital Literacy & Information Resilience Strategy Policy Pillars and Recommendations

17.1 Educator Upskilling and Gamified learning

17.1.1 The Problem

Educators are central to shaping how societies understand and respond to information. In Qatar, significant investment has gone into digital learning and infrastructure, but evidence shows that digital literacy and pedagogical capacity remain uneven.

A 2025 study of government schools in Qatar found that teachers' digital competency varies considerably, depending in part on the strength of school leadership. Where leadership is weaker, teachers report less confidence in applying digital skills beyond basic tool use(Abu-Tineh et al., 2025). Another study of elementary schools revealed that while some teachers incorporate critical thinking practices, others remain constrained by rote approaches, limiting students' opportunities to develop evaluation and analysis skills(Semmar & Fakhro, 2009). These findings suggest that although many educators can use digital platforms, fewer are equipped to teach critical evaluation, algorithmic awareness, or emotional regulation, skills central to resisting manipulation.

At the workforce level, Qatar has recognized digital skills mismatches and is working to align supply and demand (The Peninsula Newspaper, 2023). Yet training programs remain fragmented, with little systematic attention to how adults outside formal education, especially older residents and migrant workers, develop information resilience.

Emerging threats intensify these challenges. Since 2017, Qatar has faced sustained regional disinformation campaigns aimed at destabilizing society and damaging its international reputation. At the same time, there are concerns about educators' and citizens' preparedness to recognize AI-generated content, deepfakes, and synthetic media (The Peninsula Newspaper, 2023).

Taken together, the evidence indicates that Qatar's education and training systems have laid important groundwork but are not yet equipped to prepare citizens for the complexity of the modern information environment. Without targeted interventions, students and community learners, particularly adults outside formal education, older residents, and migrant communities, will remain underprepared to navigate misinformation, disinformation, and increasingly sophisticated influence operations.

17.1.2 Gamified and Role-Play Approaches

Methods of instruction are important for providing information. Evidence demonstrates that interactivity, gamification, and role-play are among the most effective methods for building resilience to disinformation, yet they remain severely underutilized. A recent scoping review by Boler and colleagues found that fewer than five percent of documented misinformation literacy interventions employed role-play techniques, even though those that did were consistently more effective in producing sustained behavioral change than checklist-based or lecture-driven approaches. (Boler et al., 2025) This underrepresentation is striking given the complexity of the disinformation problem, which is often embedded in political, social, and cultural contexts.

Checklist models, such as the popular CRAAP test (currency, relevance, authority, accuracy, and purpose), have dominated media literacy pedagogy for decades. These tools are simple to apply and easy to scale, but their limitations are increasingly clear. While they may help learners recall criteria for source evaluation and form an important part of media literacy training, they rarely equip individuals to withstand the emotional, social, and political manipulations that make falsehoods compelling in practice (Boler et al., 2025). The research consensus suggests that checklist approaches tend to improve declarative knowledge ("what to look for") but fail to instill procedural resilience ("how to act when targeted"), particularly in high-stakes or affectively charged information environments.

By contrast, gamified and role-play approaches, including classroom simulations, interactive activities, and narrative-based learning, add value through motivation and immersion. A growing body of experimental studies shows that game-based interventions, such as *Bad News* and *Fake It to Make It*, produce protective effects that persist weeks or months after exposure, whereas fact-checking or lecture-based approaches often yield only short-term gains. (Urban et al., 2018) This durability reflects their grounding in inoculation theory, which deliberately exposes learners to "weakened doses" of manipulative strategies in safe environments. By recognizing these techniques in play, learners are better prepared to resist them in real-world contexts. (Appel et al., 2025)

Engagement is another key advantage. Research in political psychology demonstrates that individuals can be more motivated to participate, reflect, and discuss when learning is framed as a simulation or game rather than a top-down instructional exercise. (Clark & Scherpereel, 2024) This is particularly important in politically sensitive contexts, where direct confrontation with polarizing issues can backfire, entrenching biases or provoking defensiveness. Games, by contrast, create a psychologically safe "third space" where learners can experiment with disinformation tactics without fear of judgment, and then critically reflect on what made those tactics effective.

Role-play also fosters metacognitive awareness. By adopting the perspective of "misinformation producers," learners experience first-hand how emotional targeting, algorithmic amplification, and identity-based appeals function. The *Fake It to Make It* game, for instance, positions players as disinformation entrepreneurs tasked with maximizing clicks and shares for profit, forcing them to deploy outrage, scapegoating, and bot amplification to succeed (Urban et al., 2018). Similarly, the *Bad News Game* invites learners to craft conspiracy theories and troll campaigns, after which participants consistently rated manipulative headlines as less credible than before playing. (Roozenbeek & van der Linden, 2019) These exercises cultivate a form of cognitive empathy: learners begin to see how their own biases, emotions, and cultural frames can be exploited, which is critical for building resilience.

Another strength lies in cultural adaptability. Unlike checklist models imported wholesale from Western pedagogical traditions, gamified and role-play methods can be tailored to Qatar's multilingual and diverse context. Scenarios can be designed to reflect region-specific disinformation narratives, migrant-worker misinformation, or health-related rumors, making them more resonant with local learners. Evidence from comparative studies of cross-cultural propaganda analysis shows that students who explored disinformation across national boundaries developed greater awareness of cultural specificity and manipulation techniques. (Hobbs et al., 2018) This suggests strong potential for adapting such approaches to Qatar's heterogeneous population.

Finally, gamification and role-play lend themselves naturally to collaborative learning, which research consistently identifies as effective for developing critical thinking and resilience (Boler et al., 2025) With this in mind, classroom simulations of "troll farms," group-based rumor verification exercises, or debates where teams adopt the roles of fact-checkers, bots, and influencers not only build individual skills but also foster collective responsibility for information health.

Taken together, these approaches mark a significant evolution beyond "spot the fake" models of media literacy. They embody a systemic pedagogy that addresses the psychological, emotional, and cultural dimensions of disinformation, while fostering skills that are both durable and adaptable. For Qatar, integrating gamification and role-play into teacher training and classroom practice will not only improve digital literacy outcomes but also strengthen long-term information resilience across diverse communities.

Gamified and role-play approaches-including classroom simulations, interactive activities, and narrative-based learning- add distinctive value:

- **Durability of Learning:** Interactive approaches produce longer-lasting effects, with gamified interventions showing measurable protection against manipulation weeks or months after exposure.
- **Engagement and Motivation:** Games and role-plays lower barriers to participation, especially in politically sensitive contexts. Learners are more willing to experiment and reflect critically when framed as simulation or play rather than confrontation.
- **Metacognitive Awareness:** By adopting the role of "misinformation producers" (e.g., through games like *Bad News, Chaos Corp,* or *Fake It to Make It*), learners experience first-hand how manipulation strategies exploit biases and emotions. This perspective-taking builds self-awareness about when *they themselves* might be vulnerable.
- **Cultural Adaptability:** Games can be tailored to Qatar's context, embedding region-specific disinformation narratives, migrant-worker misinformation, or health rumors, making them more relevant and credible than imported models.
- **Collaborative Learning:** Interactive classroom activities, such as group simulations of "troll farms" or rumor-verification role plays, foster collective discussion and shared responsibility for information health.

Together, these methods align with inoculation theory, which exposes learners to weakened versions of manipulation in controlled settings, helping them build resilience. They also mirror Qatar's cultural emphasis on participatory, community-based learning, making them especially well-suited to classrooms and community programs.

Classroom Simulation: "Burner Accounts, Bots & Trolls"

To demystify how disinformation networks and coordinated online influence campaigns operate, a controlled workshop simulation titled *"Burner Accounts, Bots & Trolls"* was formulated, and conducted with remarkable enthusiasm and learning impact. In line with the aforementioned effectiveness of becoming 'misinformation producers' in a gamified setting, the exercise immersed students in a wholly fictional information war between the invented nations of Thababia and Khonshan, enabling them to experience the logic of inauthentic amplification without any exposure to real-world harm. Every aspect of the activity was explicitly framed as a work of fiction and conducted under strict ethical guidelines designed to safeguard both participants and the integrity of online spaces.

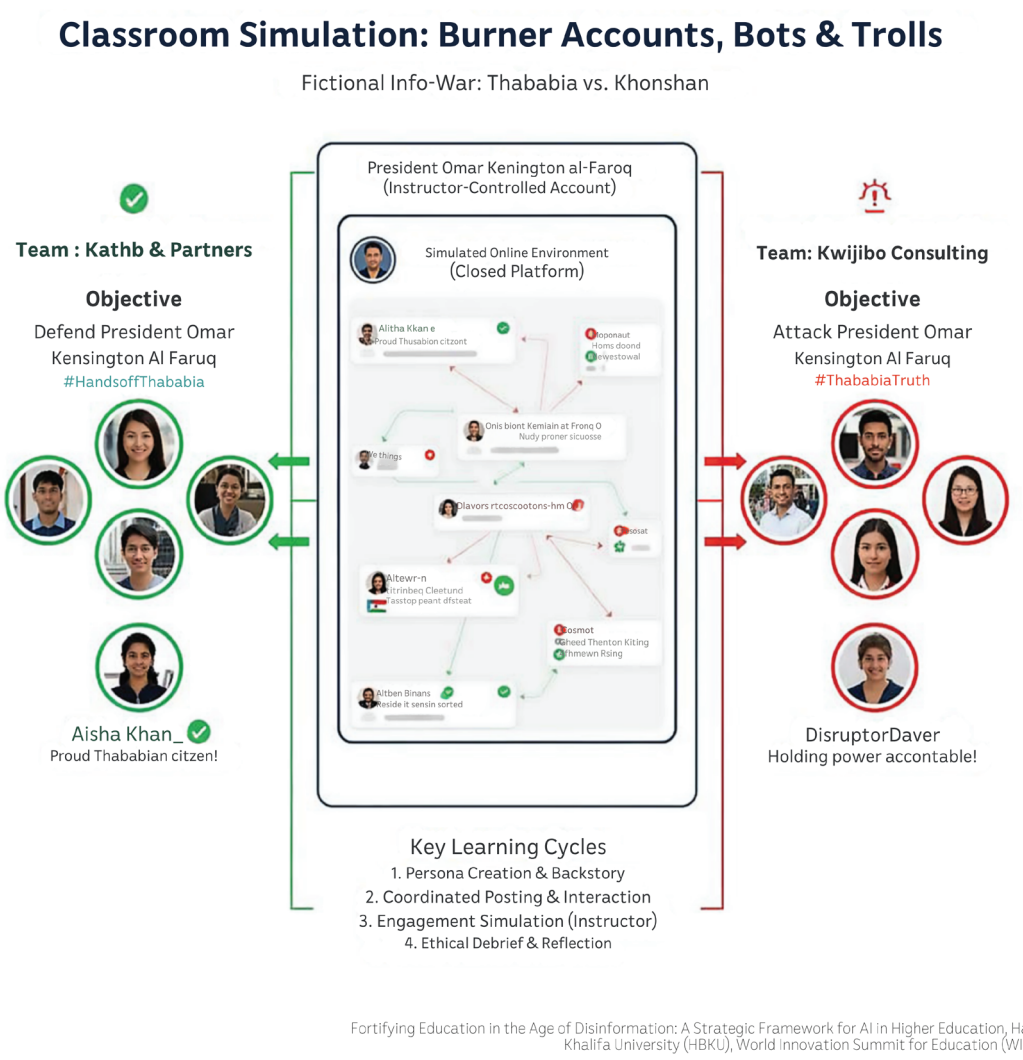


Image 17.1.Schematic showing overview of the game

The group was divided into two competing teams representing fictional consultancies working for opposing governments. One group, representing *Kwijibo Consulting* on behalf of the Kingdom of Khonshan, was tasked with destabilizing the presidency of *Omar Kensington al-Faroq* through coordinated criticism and parody. The opposing team, representing *Kathb & Kathb Partners* and contracted by the Republic of Thababia, was charged with defending the president and amplifying his legitimacy. It is worth noting that 'Thababia' was a play on the regional dialect for 'bots' (flies), while terms like 'Kathb and Kathb' translate as lies. All these touches are designed to tap into pre-learned pieces of knowledge, inject humor, and be adaptive to local context.

Within this imagined world, students were asked to simulate online coordination by building fictitious personas, creating posts, and strategically interacting with teammates to maximize engagement. Because safety and ethics were paramount, all activities took place within a closed or simulated online environment created specifically for the class. No real accounts, platforms, or individuals were used or referenced. The simulation of groups of bots and trolls working on behalf of political consultancies contracted by real nations reflected the structure of 'deception supply chains', i.e. that disinformation is often an industry. The format also resembled the known structure of campaigns used to target Qatar during the blockade years.

Students began by constructing their fictional online identities on X, choosing invented names, writing short biographical statements, and creating generic avatars using AI-generated images or stock visuals. Students were able to choose any identity they wished for and generate personas relevant to their own taste. This process allows for culturally adaptive role-playing, i.e. students can exercise autonomy over their avatars to meet their needs. Each persona carried a symbolic team identifier, 🟥 for anti-Omar participants and for 🟢 pro-Omar defenders, so that roles were clear. Working from a richly detailed backstory of Thababia and Khonshan, students familiarized themselves with their respective government's political narratives, scandals, and national myths. This imaginative context ensured the exercise remained self-contained while providing fertile ground for creativity and critical thinking.

Once the fictional world was established, the simulation unfolded in several cycles of activity. Students crafted posts, memes, and short commentaries, tagging their messages with fictional hashtags such as #ThababiaTruth or #HandsOffThababia. They interacted with one another to simulate coordinated behavior by replying, reposting, and reacting to their own side messages to create the appearance of networked support. In the classroom version of the exercise, metrics such as "likes" and "retweets" were simulated by the instructor to mirror common dynamics without using any live social media infrastructure. Students quickly grasp how coordination, repetition, and emotional framing could make content appear more authentic or more viral, even within a closed system.

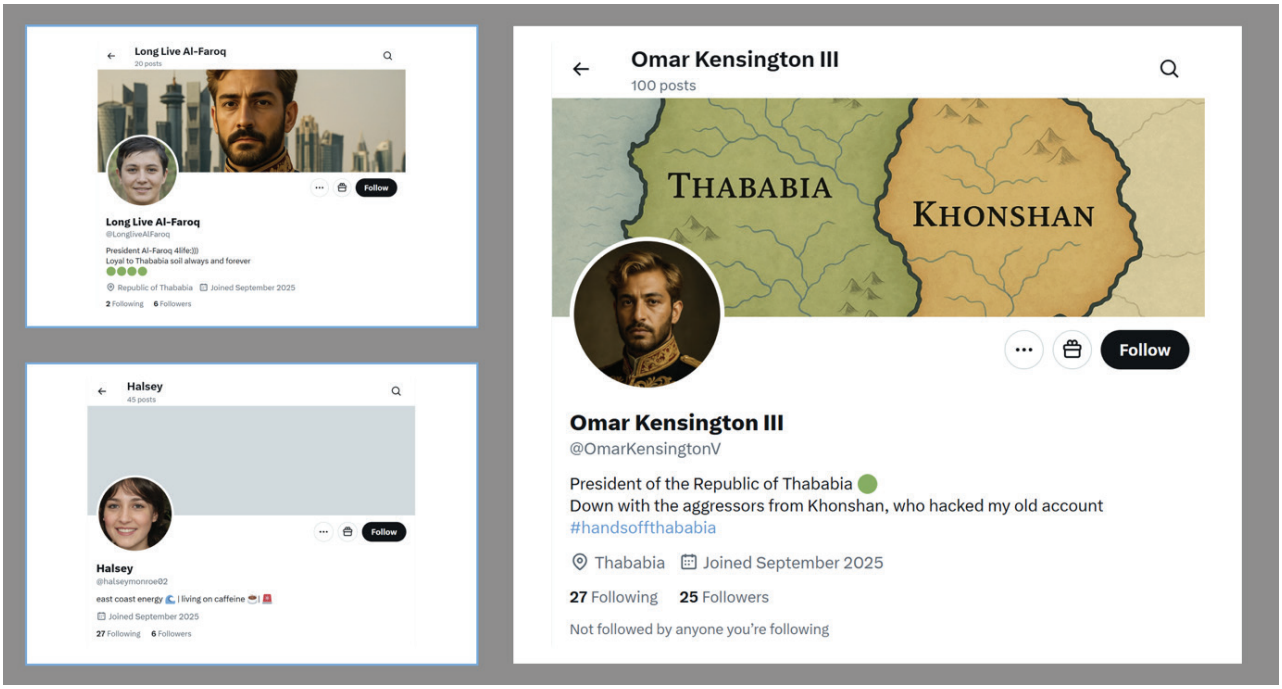


Image 17.2. Screenshots from three of the accounts involved in the activity

The instructor's role was twofold: to moderate and ensure that all content stayed within the fictional boundaries, and to facilitate discussion on the ethical dimensions of digital deception. Clear rules prohibited real-world targeting, harassment, or references to actual people or events. The exercise was carefully positioned as an analytical simulation rather than a creative writing game or a social-media experiment. Students were regularly reminded that their actions represented the reproduction of disinformation tactics for learning purposes only, and that any materials produced could be deleted or anonymized after the course. This framing underscored that ethical reflection was as important as technical understanding. The instructor also ran the account of the President of Thababia, reacting as appropriate to comments. In order to mimic real world structures in the disinformation ecosystem, students were encouraged to compete by being reminded that their total engagements would be tallied at the end.

During the debrief sessions, students reflected on how surprisingly easy it was to fabricate convincing identities and coordinated networks, even in a fictional context. They noted that simple consistency, shared imagery, unified slogans, and the illusion of interaction were enough to create perceived authenticity. Many recognized familiar patterns from real-world social media: repetitive phrasing, identical visuals across multiple accounts, and synchronized posting all contributed to the appearance of a movement. These moments of recognition proved powerful, transforming abstract discussions about "bot farms" and "troll networks" into tangible, embodied experiences. More importantly, they provided active exposure to forms to multiple key dimensions of critical literacy around disinformation; fake identities, AI-generated content, mobbing, astroturfing, politicized narratives and engagement boosting.

At the same time, the simulation provoked valuable ethical unease. Some students expressed discomfort at performing manipulative behavior, even in a fictional setting, which opened a meaningful dialogue about researcher responsibility, the moral limits of studying disinformation, and the potential emotional toll of engaging with deceit as pedagogy. This reflective component was integral to the exercise's success. By confronting their own reactions, students learned to distinguish between the technical ease of manipulation and the ethical complexity of doing so.

The outcomes of the *"Burner Accounts, Bots & Trolls"* simulation were profound. Students left with a clearer understanding of how quickly false narratives can spread, how algorithmic systems reward engagement regardless of accuracy, and how emotional or identity-based appeals drive virality. Equally important, they developed an appreciation for the fragility of truth in networked environments and for the necessity of ethical guardrails when studying or countering disinformation. The fictional design ensured that no real-world harm occurred, yet the experiential learning was vivid and authentic in its insight. Although this assessment was qualitative, widescale piloting could yield before and after surveying to assess impact.

From an instructional standpoint, the exercise demonstrates the pedagogical potential of immersive simulations in media literacy education. By situating students within an entirely fabricated but plausible digital ecosystem, educators can safely reproduce the logic of influence operations without breaching ethical or legal boundaries. The controlled environment allows students to learn detection of heuristics such as recognizing coordination patterns and linguistic repetition while cultivating critical empathy for those who encounter disinformation in real life. The simulation thus bridges theory and practice; it transforms abstract concepts about disinformation into lived experience while ensuring that ethical reflection and safety remain at the heart of the learning process.

In summary, *"Burner Accounts, Bots & Trolls"* shows that carefully constructed fictional simulations can powerfully illuminate the anatomy of digital manipulation. When conducted responsibly, with clear fictional framing, strong ethical safeguards, and structured debriefing, they not only enhance students' analytical skills but also nurture an ethical sensibility essential for future journalists, researchers, and digital citizens.

Such games can and should form a part of educators' toolkits in combatting disinformation. With an appropriate introduction and preamble on basic definitions (e.g. bots, trolls, disinformation, misinformation) and the structure of influence operations gamified methods allow teachers to address multiple facets of information manipulation, promoting skepticism in social media as a reliable source of information.

17.1.3 Policy Intervention

- Teachers and community educators to be trained to integrate digital literacy into both formal and informal learning environments. Gamified approaches as described above can be done relatively simply, inexpensively, and with minimal training. They should also function as important parts of digital literacy training. Early Learning Integration: Digital literacy competencies introduced at the primary level, embedded into existing curricula through age-appropriate games, storytelling, and activities.
- Secondary & University Levels: Digital/media literacy established as cross-curricular competencies across disciplines, not just communication studies. Universities will be encouraged to make digital literacy a core graduate skill, relevant for medicine, engineering, business, and law.

Training to focus on five core competencies:

- **Critical evaluation -** assessing credibility, verifying sources, and distinguishing between legitimate and manipulative content, e.g. manipulated or misleading visuals, e.g., out-of-context photos, altered metadata.
- **Algorithmic awareness** - understanding how recommendation systems, feeds, and search engines shape what people see online.
- **Astroturfing** - understanding the role of bots, trolls, synthetic persons and astroturfing in manipulating public opinion and generating narratives.
- **Emotional regulation** - recognizing and managing the affective techniques used in disinformation campaigns, such as outrage bait, fearmongering, or scapegoating.
- **Lateral reading** - cultivating habits of cross-checking information across multiple sources before accepting or sharing it.

Educators to also be trained to address *emerging threats*, including:

- **AI-generated text and images:** Fabricated or misleading content created by generative AI tools.
- **Deepfake video and synthetic audio:** Realistic falsifications impersonating people or events.
- **Algorithmic amplification:** Recommendation systems that boost divisive or false content.
- **Synthetic personas and bot networks:** Fake or automated accounts that simulate real users to spread disinformation.
- **Cross-platform disinformation flows:** False narratives moving between messaging apps, social media, and news outlets.

17.1.4 Recommended Implementation in Qatar

- A National Educator Micro-Credential in Digital Literacy (NEMCDL) should be established, accredited by the Ministry of Education and Higher Education (MOEHE), and could be delivered through blended online and in-person modules.
- A program to increase capacity by training the trainers. Disinformation experts in conjunction with the NEMCDL will train designated contact points at educational institutions.
- Certification will be mandatory for all K–12 teachers and offered voluntarily to university lecturers, librarians, community educators, and selected media and religious leaders who function as trusted sources of information.
- Modules would be developed in both Arabic and English (with potential optional tracks in Hindi, Nepali, Tagalog, and Urdu) to reflect Qatar's multilingual workforce.
- Practical toolkits - including lesson plans, discussion guides, games, and role-play scenarios - will allow educators to embed digital literacy into existing curricula and community programs without excessive workload.
- A crisis-response module will train educators to adapt lessons during periods of political or health-related disinformation spikes (e.g., rumor verification workshops linked to current events).
- Compliance will be monitored by MOEHE, with completion linked to career progression and professional development credits.

17.1.5 Expected Outcomes

If implemented as advised above, by 2035, every teacher and community educator in Qatar will be certified in digital literacy pedagogy. As a result:

- Students will graduate with the ability to critically evaluate digital content and recognize manipulation techniques.
- Adults engaged in community and workplace learning will develop resilience against both everyday misinformation and targeted disinformation campaigns.
- Qatar will be better prepared for future challenges posed by AI-enabled manipulation and cross-border influence operations.
- Early learners will acquire foundational habits of questioning, verification, and safe digital engagement.
- University graduates will leave with digital literacy embedded as a core competency, strengthening Qatar's workforce readiness.

17.1.6 Potential Key Performance Indicators (KPIs)

- By 2027: Micro-credential program designed and piloted with several educators certified.
- By 2029: 20% of K–12 teachers and librarians certified; crisis-response modules integrated.
- By 2030: At least 100 educators across schools, universities, and community institutions will be trained.
- By 2035: 20% of secondary students demonstrate proficiency in verification and evaluation tasks.
- Ongoing: Annual educator surveys report increased confidence in teaching both core competencies and emerging threats.

17.2 Adult & Vulnerable Populations

17.2.1 The Problem

While Qatar has invested heavily in school and university digital transformation, gaps remain in reaching both early learners and adults. Globally, most media literacy programs target school-age children, but in Qatar, these efforts are uneven, particularly at the pre-primary and early primary stages, where lifelong habits of inquiry and critical evaluation can be formed. At the other end of the spectrum, adults, especially older residents and migrant communities, remain underserved, even though research shows they are more likely to share or believe misinformation (Vivion et al., 2024).

This creates a dual vulnerability:

- Children and youth risk developing as fluent technology users without acquiring the critical and ethical foundations needed for safe participation.
- Adults outside formal education, including older Qataris and migrant workers, face challenges of language, literacy, and media unfamiliarity, leaving them exposed to manipulative campaigns.

In higher education, while universities in Qatar (such as QF institutions, Qatar University (QU), Northwestern University (NU-Q), and HBKU) are increasingly engaged with digital society topics, there is still no coordinated national approach to embedding digital literacy as a core competency across disciplines. This leaves graduates entering the workforce with variable preparedness.

17.2.2 Policy Intervention

Qatar will establish a lifelong digital literacy pathway that spans early childhood through adulthood, ensuring every learner, whether in school, university, or community education, builds resilience against disinformation.

Key elements include:

- Adult & Community Programs: Community-based delivery through libraries, community centers, and labor accommodations, supplemented by mobile outreach.
- Multilingual Access: Content developed in Arabic, English, Hindi, Nepali, Tagalog, and Urdu to reflect Qatar's diverse population.
- Trusted Intermediaries: Training community leaders, healthcare workers, religious figures, and youth mentors to serve as literacy advocates.
- Health & Crisis Modules: Special focus on combating health misinformation, rumors during emergencies, and politically motivated narratives.
- Implementation in Qatar

- MOEHE will coordinate curriculum reforms to ensure digital literacy starts in early primary education and is reinforced through secondary school and university programs.
- University-level integration will be developed in partnership with Education City universities, QU, and international accreditation bodies.
- Adult-focused programs will be jointly run by MOEHE, the Ministry of Labor, and MOPH (Ministry of Public Health), with partnerships through embassies, major employers, and expatriate associations.
- Pilot community hubs will begin in 2027, scaling nationwide by 2028.

17.2.3 Expected Outcomes

By 2035, Qatar will have created a national digital literacy pathway that equips citizens and residents at all life stages with the skills to navigate misinformation. Specifically:

- Older adults and migrant communities will gain accessible, practical tools to resist manipulation.
- Society-wide resilience will reduce the impact of both local rumors and cross-border disinformation campaigns.
- Suggested Key Performance Indicators (KPIs)
- By 2027: Digital literacy competencies introduced into the national primary curriculum.
- By 2028: At least 10% of universities in Qatar embed digital/media literacy as a cross-disciplinary graduate competency.
- By 2029: Multilingual community programs launched in six languages across five major population centers.
- By 2030: 1,000 adults reached through community and workplace programs; 25% of secondary students demonstrate proficiency in verification tasks.
- By 2035: 40% of university graduates assessed as competent in digital/media literacy; 30% of adult participants demonstrate improved evaluation skills in post-program assessments.

17.3 Platform Transparency & Data Access

17.3.1 The Problem

Qatar is one of the most digitally connected societies in the world, with social media penetration rates among the highest globally. Platforms such as WhatsApp, X, Instagram, and TikTok serve as primary sources of information for both citizens and residents. Yet these platforms remain opaque in their operations:

- Political advertising transparency tools are often unavailable or incomplete in Arabic.
- Access to platform data for independent research is restricted, particularly in the Global South.
- Coordinated inauthentic behavior, scams and synthetic accounts targeting Qatar since the 2017 blockade highlight the vulnerability of smaller states to information operations.
- Platform moderation policies are inconsistently enforced in non-English contexts, allowing harmful narratives to circulate unchecked.

Without greater transparency and data access, Qatar's ability to track, respond to, and understand disinformation campaigns is severely limited.

17.3.2 Policy Intervention

Qatar will adopt a National Platform Accountability Framework designed to ensure transparency, cooperation, and access to data for both regulators and researchers. Core measures include:

- Ad Libraries: Require platforms operating in Qatar to maintain searchable, publicly accessible advertising databases for political and issue-based content, including ads in Arabic and other widely used languages.
- Data Access Agreements: Establish partnerships between platforms and accredited Qatari research institutions (e.g., QU, QCRI, NU-Q, HBKU) to provide anonymized datasets for the study of disinformation trends.
- Content Moderation Transparency: Platforms must publish regular reports on takedowns, account suspensions, and content labeling in Qatar, disaggregated by language.
- Legal/Regulatory Levers: Qatar's Communications Regulatory Authority (CRA) will introduce guidelines requiring companies to comply with transparency standards as a condition of operating in the country.
- Regional Leadership: Position Qatar as a convener for GCC-level cooperation on platform accountability, recognizing that many disinformation campaigns targeting the country operate across regional borders.

17.3.3 Implementation in Qatar

- CRA, in partnership with the Ministry of Communications and Information Technology (MCIT), will lead negotiations with platforms to establish minimum transparency standards.
- A Platform Transparency Taskforce will be created under the proposed National Digital Literacy & Information Resilience Strategy (NDLIRS) Steering Committee.
- Pilot data-sharing agreements will be established with at least two major platforms (e.g., Meta, TikTok) by 2027.
- Universities and think tanks will be certified to receive and analyze platform data under strict ethical and privacy protocols.
- Qatar will coordinate with UNESCO and the International Telecommunication Union (ITU) to align national efforts with global norms on platform accountability.

17.3.4 Expected Outcomes

By 2035, Qatar is expected to have significantly improved oversight of the information environment through enhanced platform transparency. Outcomes include:

- Public access to advertising libraries that reveal the sources and funders of political messaging.
- Independent, evidence-based monitoring of disinformation campaigns affecting Qatar.
- Greater accountability from global technology companies operating in the country.
- Stronger protection for Qatar's information sovereignty in the face of cross-border influence operations.

17.3.5 Suggested Key Performance Indicators (KPIs)

- By 2027: Data-sharing agreements signed with at least two major platforms; public ad library established in Arabic.
- By 2028: Annual transparency reports from all major platforms operating in Qatar, disaggregated by language and content type.
- By 2029: At least three Qatari universities will be accredited to analyze platform data for disinformation research.
- By 2035: Evidence from platform datasets informs at least five published studies on disinformation in Qatar and the region.
- Ongoing: Annual audits confirm platform compliance with CRA (Ministry of Communications and Information Technology) transparency standards.

17.4 Public Service Media & Information Health

17.4.1 The Problem

Social media creates fertile ground for misinformation, rumors, and politically motivated health disinformation. Public trust in digital information sources is uneven, and there is limited coordination across sectors to provide timely, reliable, and culturally resonant counter-narratives.

At the same time, Qatar's media ecosystem has unique strengths, such as Al Jazeera, Qatar Media Corporation, and community media; however, these assets are not yet fully leveraged to support domestic information health. Without systematic engagement of public service media and trusted community voices, Qatar risks leaving its population vulnerable to rumor cycles during crises, health scares, and geopolitical tensions.

17.4.2 Policy Intervention

Qatar will launch a National Information Health Campaign, anchored in public service media, to build resilience against misinformation while strengthening civic trust. Core measures include:

- National Awareness Campaigns: Sustained campaigns promoting verification practices ("pause before you share"), fact-checking habits, and responsible online behavior.
- Trusted Messengers: Engage journalists, healthcare professionals, educators, religious figures, and influencers as credibility multipliers to reinforce accurate information.
- Multilingual Messaging: Produce campaigns in Arabic, English, Hindi, Nepali, Tagalog, and Urdu to ensure inclusivity across Qatar's diverse population.
- Crisis Communication Protocols: Develop rapid response communication strategies for health emergencies, security incidents, or disinformation spikes, coordinated through the MOI (Ministry of Interior) and MOPH.
- Youth Engagement: Partner with schools and universities to co-create campaigns, ensuring youth voices shape narratives that resonate with younger audiences.
- Integration with Digital Platforms: Work with platforms to promote reliable local information during emergencies and label false or misleading content.

17.4.3 Implementation in Qatar

- The Ministry of Information and Communications, Qatar Media Corporation, and Al Jazeera Media Institute could co-lead implementation.
- A National Information Health Unit will coordinate messaging across ministries, public health authorities, and community organizations.
- Campaigns will use a mix of television, radio, print, digital platforms, and outdoor advertising to maximize reach.
- Partnerships with embassies, expatriate associations, and employers will ensure multilingual campaigns reach migrant workers effectively.
- An annual "Media and Information Health Week" will be established, aligned with UNESCO's Global MIL Week, to highlight progress and renew public engagement.

17.4.4 Expected Outcomes

By 2035, Qatar will have established an ecosystem of trusted information partnerships, where accurate, reliable content reaches communities more quickly and credibly than misinformation. Specifically:

- Public trust in national and community media increases through transparency and responsiveness.
- Citizens and residents adopt healthier information-sharing behaviors.
- Rapid-response communication reduces the spread of rumors during crises.
- Qatar strengthens its regional reputation as a leader in responsible public communication and media literacy.

17.4.5 Suggested Key Performance Indicators (KPIs)

- By 2027: Launch of the National Information Health Campaign across all media channels.
- By 2028: Crisis communication protocols piloted during at least one national emergency simulation.
- By 2029: Multilingual campaigns covering several of Qatar's major language groups in circulation.
- By 2030, 20% of surveyed residents report exposure to information health campaigns; 20% report adopting fact-checking or "pause before sharing" practices.
- By 2035: Independent surveys show a measurable increase in trust in public service media and official communication channels.

17.5 Research & Threat Observatory

17.5.1 The Problem

Qatar has been subject to sustained disinformation campaigns, especially since the 2017 blockade, but lacks a permanent institutional mechanism for monitoring, analyzing, and testing responses to influence operations. Current research efforts are fragmented across universities, think tanks, and government bodies, with limited access to shared data or standardized methodologies. This fragmentation hampers the ability to:

- Track coordinated inauthentic behavior targeting Qatar in real time.
- Evaluate which interventions are effective in building resilience.
- Provide policymakers and the public with timely, evidence-based threat assessments.

17.5.2 Policy Intervention

To establish the Qatar Information Resilience Observatory (QIRO), a permanent, multidisciplinary unit combining expertise in communication, psychology, human-computer interaction, and Arabic natural language processing.

QIRO will:

- **Monitor**: Detect and analyze influence operations targeting Qatar across social, digital, and broadcast media.
- **Experiment**: Run randomized controlled trials (RCTs) and field experiments to test interventions (e.g., fact-checking prompts, gamified literacy, narrative inoculation).
- **Report**: Publish open-access dashboards and quarterly briefs to inform policymakers, educators, journalists, and the public.
- **Network**: Serve as a regional hub, linking Qatar's research institutions with global initiatives (e.g., UNESCO, ITU, EU, EDMO - European Digital Media Observatory).

17.5.3 Implementation in Qatar

- Hosted at a QF entity in partnership with QCRI, HBKU, and NU-Q, with CRA and MCIT oversight.
- Staffed by a core team of 20–30 researchers, with the capacity to expand during crises.
- Funded jointly by MOEHE, MCIT, and the Qatar National Research Fund (QNRF).
- Collaborative agreements with international partners to share best practices and methods.

17.5.4 Expected Outcomes

- Qatar becomes a regional leader in open-source intelligence (OSINT) and intervention research.
- Policymakers have access to timely threat assessments.
- Programs are evaluated against common metrics, ensuring resources are spent on interventions that work.

17.5.5 Suggested Key Performance Indicators (KPIs)

- By 2027: QIRO established, and the first public dashboard launched.
- By 2028: Four quarterly threat briefs published annually.
- By 2035: At least six RCTs or large-scale intervention studies completed.
- Ongoing: Open dashboards updated monthly; policy briefings provided to ministries on emerging threats.

17.6 Crisis & Health Misinformation Protocols

17.6.1 The Problem

During public health emergencies (e.g., COVID-19) and geopolitical crises (e.g., the 2017 blockade and Gaza conflict narratives), Qatar has faced waves of misinformation that spread faster than official communications. Current responses are often reactive, fragmented across ministries, and inconsistent in their multilingual approach. Delays in rebutting rumors or providing accurate information can reduce public trust and amplify harmful narratives.

17.6.2 Policy Intervention

Qatar will develop a Rapid Response Playbook to ensure coordinated, timely, and credible communication during crises. This will treat misinformation as a national security and public health issue, requiring pre-planned protocols.

Key elements include:

- Pre-approved templates: Rumor rebuttals and explainers that can be rapidly adapted to new crises.
- Multilingual communications: Arabic, English, Hindi, Nepali, Tagalog, and Urdu messaging prepared in advance.
- Coordinated spokesperson network: Trained representatives across MOPH, MOEHE, MCIT, and the Government Communications Office (GCO) to ensure unified messaging.
- Simulation drills: Twice-yearly exercises to test and refine the system, including response times and public trust measures.
- Digital platform integration: Agreements with social media companies to prioritize official health and crisis information in feeds during declared emergencies.

17.6.3 Implementation in Qatar

- The GCO will lead, with MOPH responsible for health-related misinformation, and MCIT/CRA ensuring technical delivery.
- Partnerships with religious leaders, healthcare professionals, and influencers will extend their reach into communities.
- Integration with Pillar 5.4 (Information Health Campaigns) ensures consistent messaging inside and outside crises.

17.6.4 Expected Outcomes

- Qatar establishes a reputation for fast, credible, and multilingual crisis communication.
- Rumors are rebutted within hours, not days.
- Citizens and residents increasingly rely on official resources during emergencies.
- Public trust in government communication is measurably strengthened.

17.6.5 Suggested Key Performance Indicators (KPIs)

- By 2028: Rapid Response Playbook developed; multilingual templates in place.
- By 2029: Two national-level drills conducted; <6 hours average time-to-response achieved.
- By 2030: Platform agreements operational for content prioritization.
- By 2035: Independent surveys show the number of residents who trust and use official crisis communication channels.

17.7 Standards & Quality Assurance (QA)

17.7.1 The Problem

Digital literacy and misinformation resilience programs in Qatar are proliferating across schools, universities, community centers, and workplaces. However, many initiatives remain fragmented, with inconsistent quality, cultural fit, and evaluation standards. Without a national framework, programs may lead to risk duplication, uneven effectiveness, or even counterproductive outcomes (e.g., reinforcing cynicism or mistrust).

At present:

- No unified quality rubric exists for assessing digital/media literacy interventions.
- Procurement of digital tools and AI systems often prioritizes technical functionality without a systematic review of ethical and privacy safeguards.
- Publicly funded programs vary in whether they include evaluation plans or culturally responsive design.

This fragmentation undermines Qatar's goal of building a cohesive national digital literacy ecosystem.

17.7.2 Policy Intervention

Qatar will introduce a National Standards & Quality Assurance Framework to ensure all digital literacy and misinformation resilience programs meet clear benchmarks before scaling.

Key measures include:

- Quality Mark for MIL Programs: Independent certification process assessing cultural responsiveness, accessibility (including multilingual delivery), inclusivity, and evidence-based evaluation plans.
- Evaluation Requirements: All government-funded programs must incorporate standardized metrics (aligned with Pillar 5.5's "Living Evidence Standards").
- Ethical Tech Procurement: Schools, ministries, and public institutions must apply minimum safeguards when procuring verification and AI tools, including:
    - Watermark and provenance detection capability.
    - Model-card transparency (documenting training data, limitations, and bias).
    - Privacy and data protection compliance.
- Continuous Review: Standards updated annually to reflect new threats (e.g., synthetic media, generative AI tools).

17.7.3 Implementation in Qatar

- Led by MOEHE and MCIT in coordination with CRA and QF's Education Development Institute.
- Independent Quality Assurance Board established, with representation from educators, technologists, legal experts, and community stakeholders.
- The certification process is streamlined through an online portal to reduce administrative burden.
- Procurement guidelines are embedded in government tendering rules and Ministry of Finance processes.

17.7.4 Expected Outcomes

By 2035, Qatar will have a robust and transparent system to guarantee the quality, safety, and cultural appropriateness of digital literacy initiatives. Outcomes include:

- All government-funded programs are evaluated against consistent national standards.
- Increased public trust in interventions, knowing they are vetted and evidence-based.
- Reduced risk of adopting ineffective or harmful technologies.
- Positioning Qatar as a regional model for standards in digital literacy and responsible tech adoption.

17.7.5 Suggested Key Performance Indicators (KPIs)

- By 2027: National Quality Assurance Framework launched; QA Board established.
- By 2028: 50% of new government-funded digital literacy programs certified under the framework.
- By 2029: Ethical procurement guidelines adopted across all ministries.
- By 2030: Annual QA audits published with compliance rates ⩾90%.
- By 2035: Evidence shows consistent improvement in program effectiveness and participant outcomes across certified initiatives.

18 Implementation Plan & Governance

Effective delivery of the National Digital Literacy & Information Resilience Strategy (NDLIRS 2026–2035) requires clear governance structures, defined responsibilities, and mechanisms to anticipate and mitigate risks. This section outlines a proposed framework for national implementation, emphasizing coordination, accountability, and resilience.

18.1 National Steering Committee and Delivery Unit

- A National Steering Committee (NSC) should be established, chaired by the Ministry of Education and Higher Education (MOEHE), with representation from the Ministry of Communications and Information Technology (MCIT), the Communications Regulatory Authority (CRA), the Government Communications Office (GCO), the Ministry of Public Health (MOPH), and Qatar University/Qatar Foundation institutions.
- The NSC will set strategic direction, approve annual plans, and ensure alignment with Qatar's National Vision 2030.
- A Delivery Unit housed within MOEHE will coordinate day-to-day execution, working closely with partner ministries, universities, and civil society organizations. The Delivery Unit will oversee progress monitoring, KPI reporting, and escalation of implementation challenges to the NSC.

18.2 Timeline (2026–2035)

Implementation will follow a phased approach:

Years 1-3: Pilot programs and infrastructure building

Years 4-6: Evaluation, refinement, and selective scaling

Years 7-10: Full implementation based on proven models

- 2026: Launch of NDLIRS, establishment of NSC and Delivery Unit, initiation of educator upskilling and pilot adult literacy programs.
- 2027–2028: Expansion of interventions across schools, universities, and community hubs; roll-out of platform transparency agreements and rapid-response protocols.
- 2030: Consolidation and scaling of successful programs; integration of QA standards and national evaluation instruments.
- 2035: Full coverage across education, community, and media sectors; publication of final progress report with recommendations for post-2035 strategy renewal.

18.3 Stakeholder Responsibilities

- MOEHE: Lead agency for curriculum integration, teacher training, and educator certification.
- MCIT/CRA: Regulatory oversight for platform transparency, tech procurement standards, and data-sharing agreements.
- GCO: Coordination of crisis communication and public information health campaigns.
- MOPH: Lead in health misinformation response and multilingual public health messaging.

- Universities (QU, NU-Q, HBKU, QF institutions): Research, intervention evaluation, and training delivery; hosting the QIRO.
- Qatar Media Corporation / Al Jazeera Media Institute: Production and dissemination of national awareness campaigns.
- Community and Religious Leaders: Act as trusted intermediaries for migrant and vulnerable populations.
- Private Sector (telecoms, tech platforms, employers): Support digital access, workplace programming, and co-financing community initiatives.

18.4 Risks and Mitigation Strategies

- Platform Cooperation: Risk that global tech companies resist transparency measures. *Mitigation*: Leverage Qatar's regulatory authority (CRA) and regional alliances (GCC) to negotiate compliance.
- Educator Overload: The risk that teachers perceive digital literacy as an additional burden. *Mitigation*: Provide practical toolkits integrated into existing curricula, with recognition and incentives for certification.
- Multilingual Outreach: Risk that migrant populations are excluded due to language barriers. *Mitigation*: Develop content in six core languages and partner with embassies, employers, and community groups for delivery.
- Public Trust Deficits: Risk that official campaigns are dismissed as top-down. *Mitigation*: Co-create messages with youth, influencers, clinicians, and imams to enhance credibility.
- Geopolitical Disinformation Surges: Risk of sudden spikes in cross-border campaigns during regional crises. *Mitigation*: Activate Rapid Response Playbook (Pillar 5.6) and draw on QIRO's monitoring for early warning.

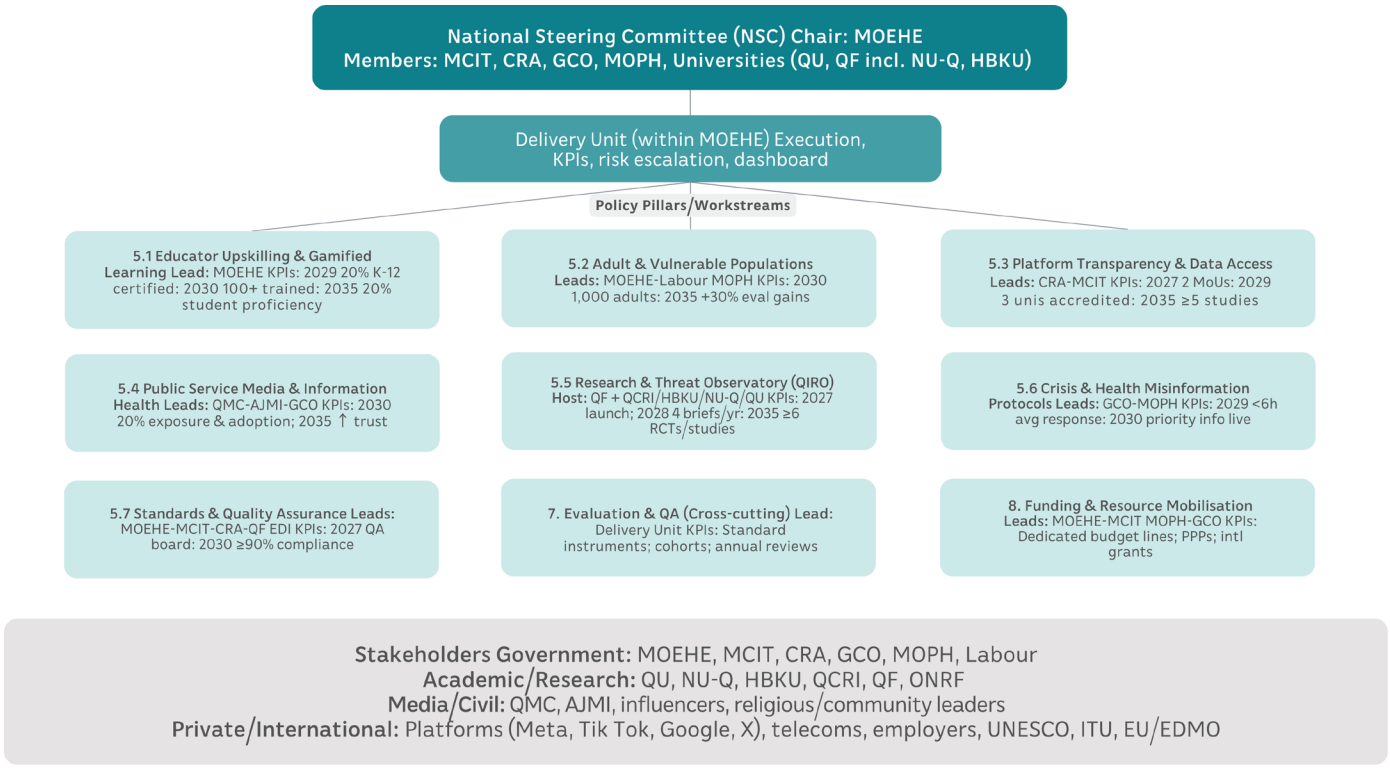


Image 18.1 Suggested flow chart detailing governance, coordination, pillars, and stakeholders

19 Evaluation and Quality Assurance

Robust evaluation and quality assurance are essential to ensure that the National Digital Literacy & Information Resilience Strategy (NDLIRS 2026–2035) delivers meaningful and measurable outcomes. Current global practice suffers from inconsistent metrics, limited longitudinal evidence, and fragmented evaluation protocols. Qatar has the opportunity to lead by adopting standardized, transparent, and culturally responsive evaluation mechanisms that provide both national accountability and international credibility.

19.1 Standardized Evaluation Protocols

- A national evaluation framework will be introduced, requiring all government-funded digital literacy programs to adopt common instruments for measuring knowledge, attitudes, and behaviors.
- Protocols will draw on validated international tools (e.g., lateral reading tests, deepfake detection tasks, susceptibility indices) and adapt them to Qatar's multilingual and cultural context.
- Independent evaluators (universities and research institutes) will be responsible for assessing program outcomes to avoid conflicts of interest.
- Alignment will be ensured with Pillar 5.5's "Living Evidence Standards," so all interventions contribute to a comparable evidence base.

19.2 Suggested Key Performance Indicators (KPIs)

- Each pillar of the strategy will be assessed through specific, measurable, and time-bound KPIs (e.g., proportion of certified teachers, speed of crisis response, number of adults trained).
- Cross-cutting KPIs will track:
    - Reach (number of educators, adults, students trained).
    - Quality (percentage of certified programs meeting QA standards).
    - Impact (measured shifts in verification skills, trust in information sources, and reduced susceptibility to misinformation).
- Annual KPI reviews will inform adjustments to programming and ensure continuous improvement.

19.3 Longitudinal Measurement Strategies

- Recognizing that the effects of digital literacy may fade or evolve, long-term cohort studies will track participants over multiple years.
- Longitudinal designs will measure not only short-term knowledge gains but also sustained behavior change, emotional resilience, and information-sharing practices.
- These studies will focus on diverse groups, including schoolchildren, university students, migrant workers, and older adults, to capture population-specific trajectories.
- Findings will feed directly into policy revisions and future iterations of NDLIRS.

19.4 Public Dashboard Reporting

- A national digital literacy and information resilience dashboard will be launched to provide transparent, real-time updates on progress.
- The dashboard will display KPIs, program reach, and selected research findings, with data disaggregated by sector, population group, and language.
- Managed by the Delivery Unit (Section 18.1) and updated quarterly, the dashboard will ensure accountability to policymakers, educators, and the public.
- This visibility will also support Qatar's role as a regional leader by showcasing progress to international partners such as UNESCO and ITU.

19.4.1 Expected Outcomes

By embedding rigorous evaluation and QA, Qatar will:

- Ensure programs are evidence-based, not ad hoc or duplicative.
- Build a cumulative knowledge base that strengthens resilience over time.
- Increase public trust through transparent reporting and demonstrated progress.
- Position itself as a global leader in measuring and evaluating digital literacy interventions.

20 Funding and Resource Mobilization

The success of the National Digital Literacy & Information Resilience Strategy (NDLIRS 2026–2035) depends on sustainable funding and effective mobilization of resources across sectors. While this document does not propose specific budget lines, it emphasizes the mechanisms and principles through which adequate support should be secured.

20.1 National Budget Allocations

- Government commitment is essential. Each participating ministry (MOEHE, MCIT, MOPH, GCO) should establish dedicated budget lines to support their respective components of NDLIRS.
- Allocations should be coordinated through the National Steering Committee (Section 18.1) to avoid duplication and ensure balanced coverage across education, research, crisis response, and public communication.
- Funding stability over the 2026–2035 cycle will be critical to avoid fragmented or short-term programming.

20.2 Public–Private Partnerships

- Given the central role of technology platforms, telecommunications firms, and media companies in shaping information flows, co-financing and in-kind contributions will be vital.
- Platforms operating in Qatar (e.g., Meta, TikTok, Google, X) should be encouraged to contribute to capacity-building, awareness campaigns, and transparency initiatives as part of their corporate responsibility.
- Employers and industry associations can support digital literacy for migrant workers and workplace communities, helping extend reach beyond the classroom.
- Partnerships with Qatar's financial and energy sectors can ensure alignment with the broader national innovation and resilience agenda.

20.3 International Collaboration (UNESCO, UN, EU)

- Qatar should leverage international partnerships to secure expertise, training materials, and co-funding opportunities.
- Collaboration with UNESCO's Media and Information Literacy (MIL) program and ITU digital skills initiatives can provide standardized frameworks and global recognition.
- Engagement with UN agencies and the EU's European Digital Media Observatory (EDMO) can bring in comparative lessons, technical expertise, and potential grant funding.
- By positioning itself as a regional leader, Qatar can attract external investment and technical support, while simultaneously contributing its own expertise on information resilience in the Gulf and wider MENA region.

20.3.1 Expected Outcomes

By diversifying funding sources and embedding resilience into existing national and international initiatives, Qatar will:

- Ensure long-term sustainability of digital literacy programming.
- Avoid dependency on any single sector or external partner.
- Strengthen international legitimacy by aligning with global best practices.
- Enhance regional leadership by demonstrating commitment to both domestic resilience and global cooperation.

21 Conclusion: The Stakes for Social Cohesion, Stability, and International Credibility

The digital information environment is now a critical domain of national resilience. For Qatar, where society is both highly connected and highly diverse, the stakes are especially high. Coordinated disinformation campaigns since 2017 have shown how external actors can exploit digital vulnerabilities to destabilize societies, amplify tensions, and undermine international reputation.

Investing in digital literacy is therefore not simply an educational or technological priority: it is a matter of social cohesion, national stability, and information sovereignty. By equipping students, adults, educators, and communities with the skills to navigate misinformation, Qatar strengthens its capacity to withstand external influence operations while promoting a healthier information ecosystem at home.

Globally, the ability of states to project credibility increasingly depends on their information resilience. By demonstrating leadership in this space, Qatar can enhance its international standing, positioning itself as a regional model for evidence-based, culturally responsive approaches to disinformation.

The challenges posed by misinformation, disinformation, and synthetic media are urgent and evolving. Incremental steps are no longer sufficient. What is required is a coordinated, systematic, and sustained response that brings together government, educators, researchers, media, communities, and international partners.

The National Digital Literacy & Information Resilience Strategy (NDLIRS 2026–2035) provides a roadmap for achieving this. Its success depends on strong political commitment, institutional collaboration, and societal engagement. By adopting and implementing the policies outlined in this strategy, Qatar can:

- Equip its population to recognize and resist manipulation.
- Strengthen social trust and cohesion in a diverse society.
- Build resilience against future crises and influence operations.
- Contribute to global debates on safeguarding the integrity of information.

The task ahead is clear: Qatar must act decisively to embed digital literacy across all levels of society, from early learners to adults, from classrooms to workplaces, and from media systems to regulatory frameworks. The cost of inaction is high, measured in vulnerability, mistrust, and diminished international credibility. The opportunity, however, is greater: to secure a future where Qatar's information environment is open, resilient, and trusted.

UNIFIED CONCLUSION

A Comprehensive Roadmap for Education City and Qatar

This integrated report has presented a three-part strategy to fortify education in the age of AI and disinformation. Each component addresses a distinct but deeply interconnected dimension:

- **Part 1: Institutional Foundations** provided empirical evidence from HBKU demonstrating that the Education City community is actively engaging with AI while facing challenges in policy clarity, pedagogical guidance, and digital literacy. This analysis revealed the urgent need for coherent governance frameworks and professional development, insights that inform the entire strategy.
- **Part 2: Healthy Digital Discourse** addressed the quality of online communication, a challenge amplified by AI-generated content. The curated Arabic toxicity dataset and bilingual digital literacy platform provide both evidence and practical tools for cultivating constructive engagement in educational settings.
- **Part 3: National Information** Resilience scaled these institutional insights to propose Haqiqatar: a comprehensive ten-year national strategy. This framework ensures that lessons learned at Education City can inform policy and practice throughout Qatar, building societal resilience against disinformation.

Unified Implementation Framework

Success requires coordinated action across institutional, technological, and national levels:

- **Immediate Actions (0-6 months)** - Establish Education City AI Governance Committee with representatives from all institutions - Launch faculty development programs using digital literacy resources from Part 2 - Begin pilot implementation of AI literacy curricula across Education City - Convene national stakeholders to discuss Haqiqatar implementation
- **Medium-Term Initiatives (6-18 months)** - Formalize unified AI policies across Education City institutions - Deploy digital literacy platform at scale within Education City and pilot nationally - Establish QIRO as proposed in Part 3 - Launch educator training programs for national rollout of media literacy curricula
- **Long-Term Transformation (18+ months)** - Full implementation of national digital literacy strategy across all educational levels - Position Qatar as a regional leader in ethical AI integration and information resilience - Establish Education City as an international research hub for AI ethics and digital literacy - Scale successful models globally through the WISE network and international partnerships.

The Stakes and the Opportunity

At stake is nothing less than the preservation of informed citizenship in an age of sophisticated manipulation. The challenges of AI and disinformation will not resolve themselves. They require vision, coordination, resources, and sustained commitment.

Education City, with its unique constellation of world-class institutions, research capacity, and alignment with Qatar's national vision, is optimally positioned to pioneer this transformation. By implementing this unified strategy, Qatar can:

- **Lead regionally** in responsible AI integration and information resilience
- **Protect nationally** against disinformation threats while fostering innovation
- **Contribute globally** by developing models that others can adapt and implement

By embracing this unified roadmap, Qatar has the unparalleled opportunity to transform the profound challenges of the digital age into a defining moment of national resilience and global leadership.

References

Aboulnaga, A., Chawla, S., & Elmagarmid, A. (2021). An AI-enabled future for Qatar and the region. *Communications of the ACM*, *64*(4), 34-36. https://doi.org/10.1145/3447725

Abu-Tineh, A. M., Alazaizeh, M., Shal, T., Alshaboul, Y., & Ghamrawi, N. School leaders' digital leadership and teachers' digital competency in government schools in Qatar: Predicting teachers' perceptions. *Leadership and Policy in Schools*, 1–15. https://doi.org/10.1080/15700763.2025.2531061

Almaraz-Menéndez, F., & Maz-Machado, A. (2022). Strategy, policy, practice, and governance for AI in higher education institutions. *IGI Global.*

Alvarez, B. (2021). Public libraries and the dismantling of disinformation. *Public Libraries*, *60*(2), 34–43.

Al-Zyoud, H. M. M. (2020). The role of artificial intelligence in teacher professional development. *Universal Journal of Educational Research, 8*(11B), 6263 - 6272. DOI: 10.13189/ujer.2020.082265

Appel, R. E., et al. (2025). Psychological inoculation improves resilience to and reduces willingness to share vaccine misinformation. *Scientific Reports, 15*(1), 29830. https://doi.org/10.1038/s41598-025-09462-5

Bao, W. (2024). A systematic review of AI-driven intelligent tutoring systems (ITSs) in K-12 education. *Education and Information Technologies*, *29*(1), 1-25. https://doi.org/10.1007/s10639-023-11968-y

Barus, O. P. (2025). Shaping generative AI governance in higher education. *Computers and Education: Artificial* Intelligence, 8, 100184. https://doi.org/10.1016/j.caeai.2025.100184

Billiot, T. (2023). Continuous learning and advancing technologies: a framework for professional development and training in artificial intelligence. Dev. Learn. Organ. 37, 28–31. DOI: 10.1108/DLO-04-2022-0064

Boler, M., Gharib, H., Kweon, Y.-J., Trigiani, A., & Perry, B. (2025). Promoting mis/disinformation literacy among adults: A scoping review of interventions and recommendations. *Communication Research*. Advance online publication. https://doi.org/10.1177/00936502251318630

Bozkurt, A. (2024). The rise of generative AI and the future of education in the attention economy. *Asian Journal of Distance Education*, *19*(1), 1-12.

Brashier, N. M., & Schacter, D. L. (2020). Aging in an era of fake news. *Current Directions in Psychological Science, 29*(3), 316–323. https://doi.org/10.1177/0963721420915872

Cernicova-Buca, M., & Ciurel, D. (2022). Developing resilience to disinformation: A game-based method for future communicators. *Sustainability, 14*(9), 5438. https://doi.org/10.3390/su14095438

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*(1), 55. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y. (2024). A comprehensive AI policy for a university: a 10-point framework. *Computers and Education:* X, 6, 100059. https://doi.org/10.1016/j.cex.2024.100059

Clark, N., & Scherpereel, J. A. (2024). Do political science simulations promote knowledge, engagement, skills, and empathy? *Journal of Political Science Education, 20*(1), 133–152. https://doi.org/10.1080/15512169.2023.2204236

Compton, J., Jackson, B., & Dimmock, J. A. (2016). Persuading others to avoid persuasion: Inoculation theory and resistant health attitudes. *Frontiers in Psychology, 7*(122). https://doi.org/10.3389/fpsyg.2016.00122

Dai, Y., & Li, Y. (2024). Ethical challenges of generative artificial intelligence in education. *Journal of Educational Technology Development and Exchange*, *17*(1), 1-16.

Eke, D. O. (2023). The ChatGPT conundrum: A review of literature on the ethical and academic integrity implications of generative artificial intelligence. Journal of Academic Ethics, 1-18. https://doi.org/10.1007/s10805-023-09503-6

Fanar Team. (2025). Fanar: An Arabic-centric multimodal generative AI platform. *arXiv.* https://arxiv.org/abs/2501.13944

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT in higher education. European Journal of Teacher Education, 1-20. https://doi.org/10.1080/02619768.2024.2324076

Grand View Research. (2024). *AI in education market size & share | Industry report, 2030*. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-education-market-report

Gruenhagen, J. H., Sinclair, P. M., Carroll, J. A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence*, *7*, 100273. https://doi.org/10.1016/j.caeai.2024.100273

Head, A. J., Wihbey, J., Metaxas, P. T., MacMillan, M., & Cohen, D. (2018, October 16). *How students engage with news: Five takeaways for educators, journalists, and librarians*. Project Information Literacy Research Institute. https://projectinfolit.org/publications/news-study/

Henley, J. (2023, November 7). 85% of people worry about online disinformation, global survey finds. *The Guardian*. https://www.theguardian.com/technology/2023/nov/07/85-of-people-worry-about-online-disinformation-global-survey-finds

Hobbs, R., Seyferth-Zapf, C., & Grafe, S. (2018). Using virtual exchange to advance media literacy competencies through analysis of contemporary propaganda. *Journal of Media Literacy Education, 10*(2), 152–168. https://doi.org/10.23860/JMLE-2018-10-2-9

Jones, M.O. (2023) The "Bot-iful" Game: Football, Social Media Manipulation, and Transnational Gulf Politics, https://pomeps.org/the-bot-iful-game-football-social-media-manipulation-and-transnational-gulf-politics

Jones, M. O. (2022). *Digital authoritarianism in the Middle East*. Hurst.

Jones, M.O. (2020). *Political Repression in Bahrain*. Cambridge University Press.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. Journal of Applied Research in Memory and Cognition, 6(4), 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lodge, J. M., & Thompson, K. (2024). What is the problem with academic integrity in the age of AI? *Australian Universities' Review*, *66*(1), 4-9.

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of* Experimental Psychology: Applied, 27(1), 1–16. https://doi.org/10.1037/xap0000315

Mahrishi, M., Abbas, A., & Siddiqui, M. K. (2025). Global initiatives towards regulatory frameworks for artificial intelligence (AI) in higher education. *Digital Government: Research and Practice*, *3*(2), 1-15. https://doi.org/10.1145/3672462

Ministry of Communications and Information Technology, Qatar. (n.d.). *Artificial Intelligence Committee & AI Strategy*. Retrieved September 29, 2025, from https://www.mcit.gov.qa/en/artificial-intelligence-committee/

Mutsvairo, B., & Bebawi, S. (2019). Journalism educators, regulatory realities, and pedagogical predicaments of the "fake news" era: A comparative perspective on the Middle East and Africa. *Journalism & Mass Communication Educator, 74*(2), 143–157. https://doi.org/10.1177/1077695819833552

PwC Middle East. (n.d.). *The potential impact of artificial intelligence in the Middle East*. PwC. Retrieved October, 17, 2025, from https://www.pwc.com/m1/en/publications/potential-impact-artificial-intelligence-middle-east.html

Qatar Computing Research Institute. (2019). *National artificial intelligence strategy for Qatar*. https://qcai-blog.qcri.org/wp-content/uploads/2020/04/QCRI-Artificial-Intelligence-Strategy-2019-ENG.pdf

Qatar Foundation. (2025, April 20). *AI tool 'Fanar' aims to protect the Arabic language in the digital world*. https://www.qf.org.qa/stories/ai-tool-fanar-aims-to-protect-the-arabic-language-in-the-digital-world

Randhawa, G. K., & Jackson, M. (2020). The role of artificial intelligence in learning and professional development for healthcare professionals. *Healthcare Management Forum*, *33*(1), 19-23. https://doi.org/10.1177/0840470419869032

Rizwan, A., & Hillman, S. (2021). Laboring to communicate: Use of migrant languages in COVID-19 awareness campaign in Qatar. *Multilingua, 40*(3), 303–337. https://doi.org/10.1515/multi-2020-0119

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5*, 65. https://doi.org/10.1057/s41599-019-0279-9

Ruffalo Noel Levitz. (2024, August 22). *Why universities need AI governance.* https://www.ruffalonl.com/blog/artificial-intelligence/why-universities-need-ai-governance/

Semmar, Y., & Fakhro, A. (2009). Investigating critical thinking skills practices in Qatari elementary schools. *The International Journal of Learning, 16*(1), 13–24. https://doi.org/10.18848/1447-9494/CGP/v16i01/46057

Southworth, J. (2023). Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence, 4, 100061. https://doi.org/10.1016/j.caeai.2023.100061

Surjatmodjo, D., Unde, A. A., Cangara, H., & Sonni, A. F. (2024). Information Pandemic: A Critical Review of Disinformation Spread on Social Media and Its Implications for State Resilience. *Social Sciences*, *13*(8), 418. https://doi.org/10.3390/socsci13080418

Szakács, J., & Bognár, É. (2021, June 24). *The impact of disinformation campaigns about migrants and minority groups in the EU* (In-Depth Analysis No. EXPO_IDA(2021)653641). European Parliament, Think Tank. Retrieved September 21, 2025, from https://www.europarl.europa.eu/thinktank/en/document/EXPO_IDA(2021)653641

The Peninsula Newspaper. (2023, November 18). Qatar Foundation professor highlights media literacy in the face of misinformation. *The Peninsula*. https://thepeninsulaqatar.com/article/18/11/2023/qatar-foundation-professor-highlights-media-literacy-in-face-of-misinformation

UNESCO. (2023, June 1). *UNESCO survey: Less than 10% of schools and universities have formal guidance on AI*. UNESCO. https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai

UNESCO. (2023, November 6). *Online disinformation: UNESCO unveils action plan to regulate social media platforms* [Press release]. Retrieved August 25, 2025, from https://www.unesco.org/en/articles/online-disinformation-unesco-unveils-action-plan-regulate-social-media-platforms

UNESCO. (2024, February 6). *Global media and information literacy assessment framework: Country readiness and competencies*. UNESCO. Retrieved September 22, 2025, from https://www.unesco.org/en/articles/global-media-and-information-literacy-assessment-framework

UNESCO. (2025). *UNESCO strengthens global push to media and information literacy*. UNESCO. https://www.unesco.org/en/articles/unesco-strengthens-global-push-media-and-information-literacy-media

UNESCO. (2025). *UNESCO survey: Two-thirds of higher education institutions have or are developing guidance on AI use*. UNESCO. Retrieved October 22, 2025, from https://www.unesco.org/en/articles/unesco-survey-two-thirds-higher-education-institutions-have-or-are-developing-guidance-ai-use

UNESCO. (n.d.). *Media and information literacy*. UNESCO. Retrieved September 21, 2025, from https://www.unesco.org/en/media-information-literacy

United Nations. (n.d.). *Global media and information literacy week*. United Nations. Retrieved September 21, 2025, from https://www.un.org/en/observances/media-information-literacy-week

Urban, A., Hewitt, C., & Moore, J. (2018). Fake it to make it, media literacy, and persuasive design: Using the functional triad as a tool for investigating persuasive elements in a fake news simulator. *Proceedings of the Association for Information Science and Technology*, *55*(1), 915–916. https://doi.org/https://doi.org/10.1002/pra2.2018.14505501174

Vivion, M., Reid, V., Dubé, E., Coutant, A., Benoit, A., & Tourigny, A. (2024). How older adults manage misinformation and information overload: A qualitative study. *BMC Public Health, 24*, 871. https://doi.org/10.1186/s12889-024-18335-x

World Bank. (n.d.). *World Bank Open Data*. Retrieved July 2, 2025, from https://data.worldbank.org

World Economic Forum. (2025, January). *Global Risks Report 2025: Conflict, Environment and Disinformation Top Threats*. Retrieved July 16, 2025, from https://www.weforum.org/press/2025/01/global-risks-report-2025-conflict-environment-and-disinformation-top-threats/

World Economic Forum. (n.d.). *Education 4.0*. Reskilling Revolution Initiative. Retrieved October, 17, 2025, from https://initiatives.weforum.org/reskilling-revolution/education-4-0

Young, J. C., Boyd, B., Yefimova, K., Wedlake, S., Coward, C., & Hapel, R. (2021). The role of libraries in misinformation programming: A research agenda. *Journal of Librarianship and Information Science, 53*(4), 539–550. https://doi.org/10.1177/0961000620966650

Zawacki-Richter, O., & Latchem, C. (2018). Exploring four decades of research in Computers & Education. *Computers & Education*, *122*, 136-152. https://doi.org/10.1016/j.compedu.2018.04.001

About WISE

WISE is a global education platform and think-and-do tank convening leaders to shape the future of learning. Established in 2009 by Qatar Foundation under the leadership of its Chairperson, Her Highness Sheikha Moza bint Nasser. WISE drives educational innovation through policy engagement, research, leadership development, and practitioner programs. Through our year-round activities and flagship bi-annual Summit, WISE is building the future of education through strategic local, regional, and international collaborations.

Acknowledgment

We would like to extend our sincere gratitude to Chaïmaa Benkermi, Advocacy and Community Development Associate at WISE, and Romane Monnier, Consultant at WISE, for their creative and design support throughout this project. We also thank Eva Keiffenheim for her invaluable editorial contribution during the writing process.

Special thanks to Sabina Huseynova, Senior Project Analyst in the College of Humanities and Social Sciences at HBKU, for her assistance with the literature review and report formatting.

Report Citation

World Innovation Summit for Education (WISE), Hamad Bin Khalifa University (HBKU), & Northwestern University in Qatar (NU-Q). (2025). Fortifying education in the age of disinformation: A strategic framework for AI in higher education in Qatar. Qatar Foundation.

AI Disclaimer

The authors acknowledge the use of AI tool in the writing process of this study, primarily to enhance the readability of the findings. AI-generated content was not used verbatim; instead, it was thoroughly reviewed, edited, and curated by the authors to ensure accuracy, authenticity, and integrity. Human oversight and judgment were essential in interpreting and validating the AI's contributions.

وايز WISE

مؤسسة قطر Qatar Foundation

Unlocking human potential لإطلاق قدرات الإنسان

وايز
WISE

مؤسسة قطر
Qatar Foundation
Unlocking human potential لإطلاق قدرات الإنسان

ThinkBay
Qatar Foundation – Education City
P.O. Box 5825 Doha, Qatar

© WISE, Qatar Foundation. All Rights Reserved.