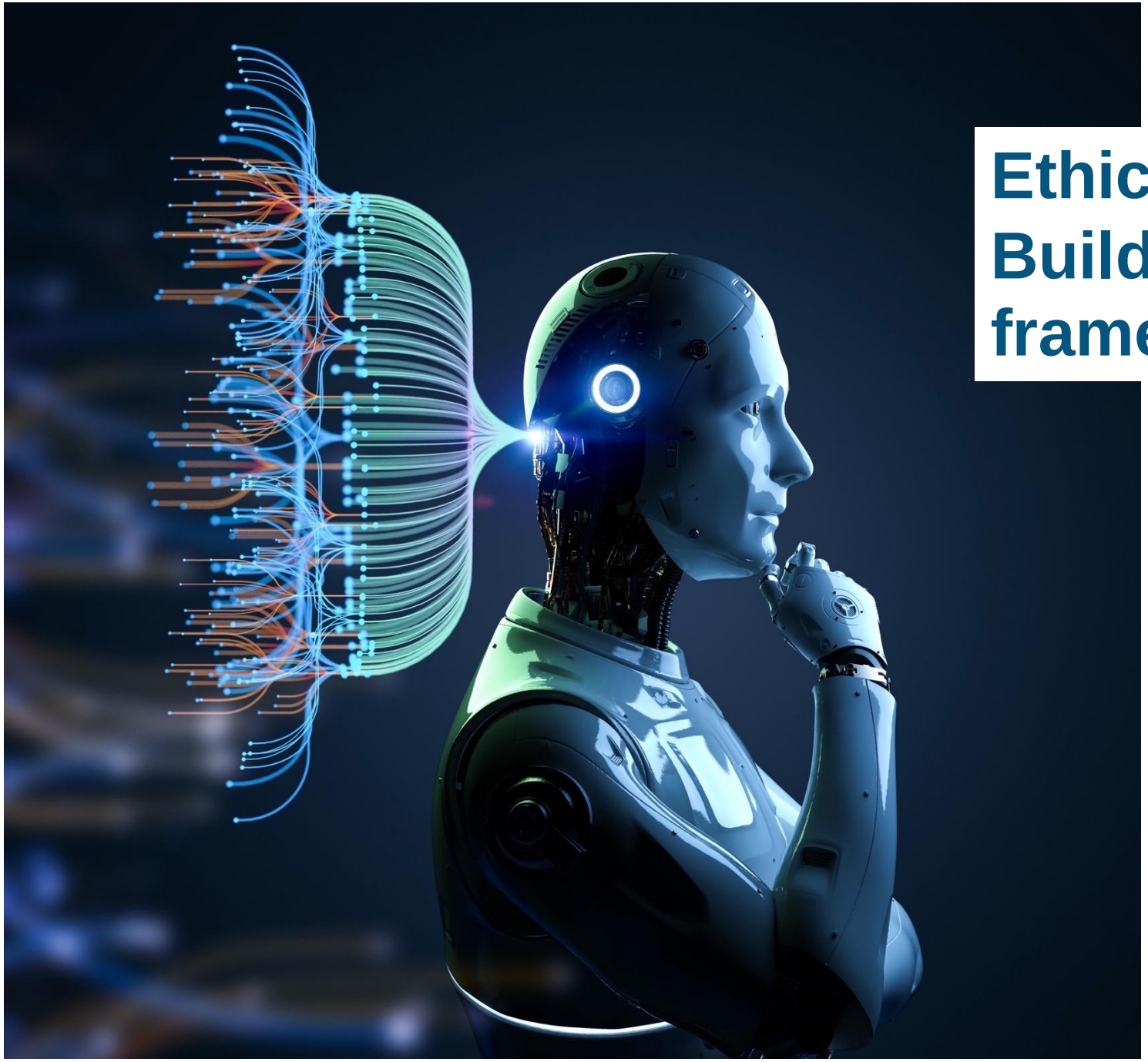


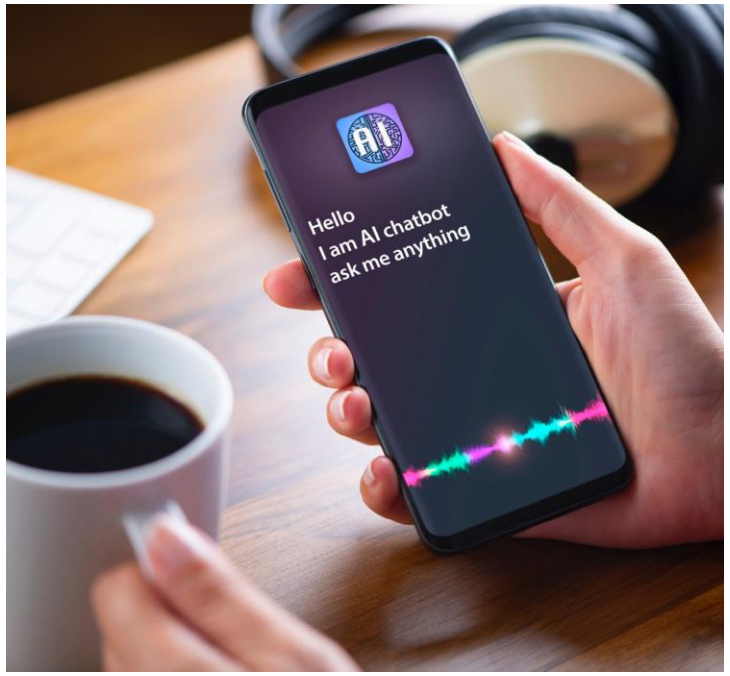
Ethical AI in Education: Building transparent frameworks for the future

Pierrette Renée Dagg, Ph. D

Director of Research, Merit Network

Research Associate, Quello Center for Media & Information Policy, Michigan State University

Adjunct Instructor, Oakland Community College



GOALS

- Establishing and structuring an AI Governance Committee
- Developing transparent evaluation metrics for new educational technologies
- Conducting environmental scans and creating organizational AI roadmaps
- Identifying actionable components of ethical AI tailored to your district or county office of education

Housekeeping

•Please use the Q&A feature to submit your questions. This is located on the ribbon below the speaker. The chat feature has been turned off.

•Your questions will be answered either in the Q&A feature in writing or live by our presenters

•All attendees will receive a link to this video recording as well as resources in a follow up email after the webinar.

Statement of the Problem

> No framework has been devised, no guidelines have been agreed, no policies have been developed, and no regulations have been enacted to address the specific ethical issues raised by the use of AI in education.

(Holmes, 2019, p. 505)

AI vs. Generative AI

- New type of AI that creates new data based on training data. In the instance of large language models (ie: Bard or GPT), the AI performs a variety of natural language processing tasks, like answering questions

- Generative AI enables users to quickly generate new content based on a variety of inputs. Inputs and outputs to these models can include text, images, sounds, animation, 3D models, or other types of data.

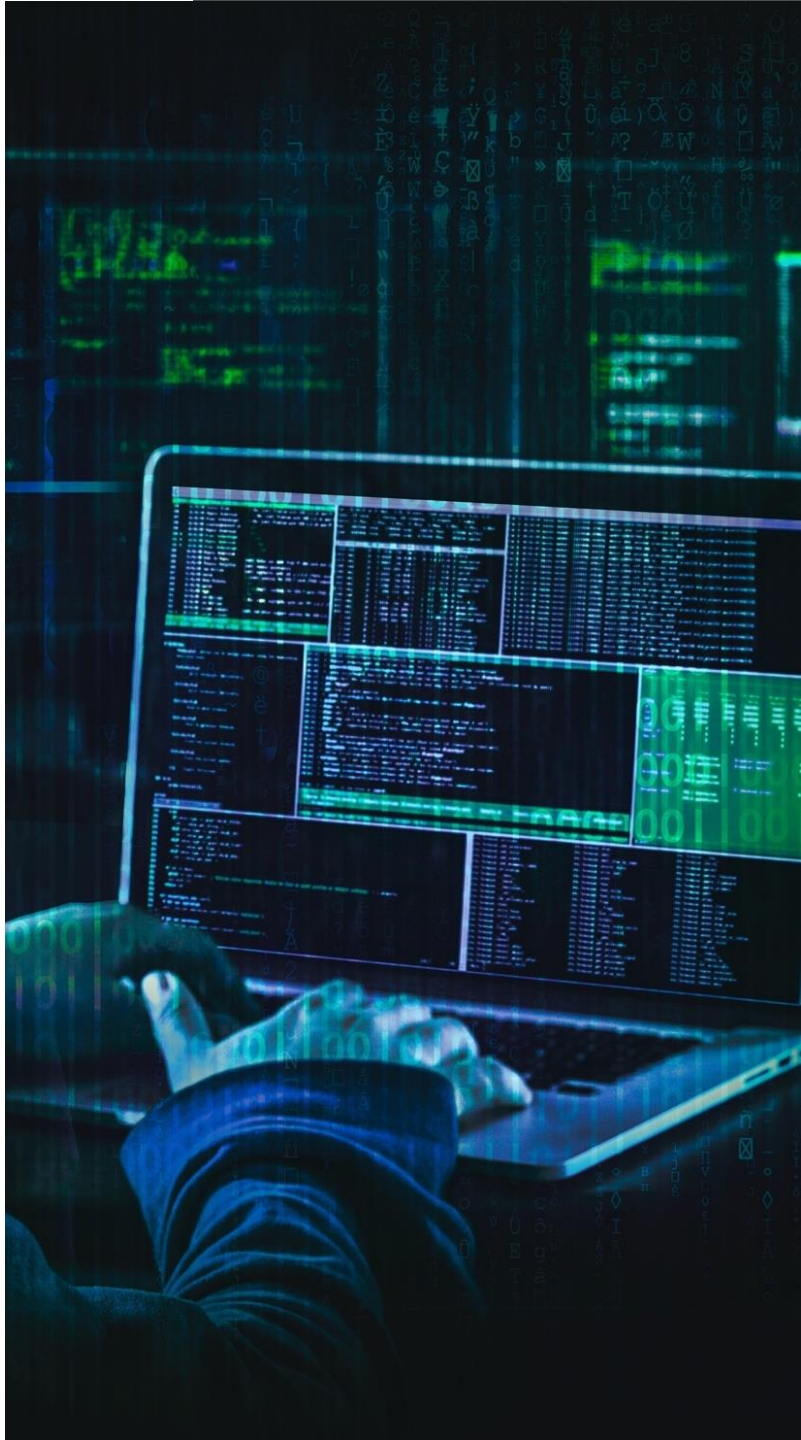
- Traditional AI systems are primarily used to analyze data and make predictions, while generative AI goes a step further by creating new data similar to its training data.

- In other words, traditional AI excels at pattern recognition, while generative AI excels at pattern creation.

- Generative Adversarial Networks (GANs) is a popular technique in GAI (though there are others). This is made of two neural networks (Generator and Discriminator). Generator produces data and Discriminator tries to distinguish between real data and data produced by the Generator. Over time, Generator improves.
    - No shortage of applications in society - new pieces of art and music, data augmentation (creating additional training data for machine learning models), drug discovery (molecular structures for new potential drugs), gaming, etc.

AI in the Educational Sector

- Forms of AI are increasingly utilized to enhance learning and operational efficiency
- Facilitate personalized learning experiences by adapting educational content to meet individual student needs, automating routine tasks like grading, and providing real-time feedback
- Assist in administrative tasks, streamlining processes like scheduling, student admissions, and resource allocation, thereby allowing educators to focus more on teaching and less on bureaucratic tasks



Delphi Round 1

- ChatGPT was a major concern/interest
- Most schools slower to adopt AI than anticipated, citing budget and personnel constraints
- Participant's definitions of AI varied greatly
- Many did not demonstrate an accurate understanding of when they were interacting with AI

Administrators:
- Expect increase of AI technologies within the next 2 years (student management, recruiting, student success)
- Lack of policies in place, reluctance to become involved in classroom technology decisions
- Concerns included reevaluation of governance, security, data compliance, business continuity, ethics, staff training, digital inequalities, competition for recruiting, staff displacement

Educational Technologists:
- Use AI in ESL, social studies, language arts, surveillance camera systems
- "Believers" in AIEd: coaching, student support, disability accommodations, administration efficiencies
- Concerns included biases, teacher training, low student AI literacy/fact checking skills, protecting students from manipulation (coercion and deepfakes)

Educators:
- Applications and understanding varied
- Lack of policies or guidelines for AI use and evaluation
- Concerns included digital inequalities, lack of transparency, changes in student/teacher relationships, pedagogical implications, ethical issues with students using GPT, social isolation
- Several equated GPT to early concerns with Wikipedia

Students:
- All acknowledged use of GPT for education – primarily as a "time saving device" (unsupervised machine learning, Grammarly, debugging code)
- Concerns included inaccuracies in the technologies, lack of knowledge about how schools were using AI in decision-making, data security

Delphi Round 2

Table 1

Delphi Round 2 Questionnaire - AI Systems

AIEd Consideration	Agree (n)	Agree (%)	Disagree (n)	Disagree (%)	Unsure (n)	Unsure (%)	Total (n)
Users should understand when they are interacting with AI	21	87.5	1	4.2	2	8.3	24
Stakeholders should be aware of when AI is used to make student-related decisions	23	95.8	0	0	1	4.2	24
Educational institutions should increase future use of AI	17	70.8	3	12.5	4	16.7	24
Educational institutions need to adopt clear guidelines for selection of AI tools	23	95.8	0	0	1	4.2	24
Privacy and information protections should be clear to all stakeholders	23	95.8	0	0	1	4.2	24
Clear accountability metrics for AI are lacking in education	20	83.3	0	0	4	16.7	24
Access to AI tools in education are generally equitable	6	25	9	37.5	9	37.5	24

INCEPTION OF ED TECH

1920s

Radio as the first MOOC

1950s

Teaching machines proliferate

1963

BASIC is developed

Late 1960s

Vocational training programs for mainframe and computer maintenance

1972

Scantrons in schools

1975

Computers move into schools (Apple dominating edu market by 1983

1985

Mass marketed PCs (Toshiba and Apple)

1990

HTML is developed

Late 1990s – Early 2000s

Most classrooms equipped with a computer. LMS systems like Blackboard debut

2010

1MM+ students attend school virtually

Teaching Machines

Early Innovations (1910-1940)

- Film's limited quality materials hindered adoption
- Edison predicted motion pictures would take over education
- Critics were concerned about "emotional engineering"
- Teachers viewed distance learning radio as an existential threat

Mid-century Challenges (1950s – 1970s)

- Television had promise, but curriculum designers had little experience in teaching and teachers had little involvement in planned implementation
- Teaching Machines introduced mechanized instruction
  - Proponents – teachers would become more important as time was freed
  - Opponents – issues with "shaping behavior," job security issues, dehumanization of the profession, poor curriculum, too much power given to a small number of developers

Rise of Computing (1960s – 1990s)

- Concerns with achievement gap
- Teachers unequipped with skills and pedagogical approaches
- Clash of values with capitalism in education
- Suspicions of machines teaching human values
- Questions of local versus national control and deskilling of the profession

Internet Era (1990s – 2000s)

- Teachers unprepared to use computers and the internet
- Digital divide grows
- Predatory online institutions with poor curriculum

FOUR EMERGING THEMES: 1970-2000

EFFECTIVE DIGITALLY SUPPORTED PEDAGOGY

INSUFFICIENT TECHNICAL TEACHER TRAINING

PHILOSOPHICAL AND NORMATIVE CONCERNS OF ED TECH

TECHNOLOGY'S PROMISE VERSUS PERFORMANCE

“WHILE IT IS VERY DESIRABLE FOR PEOPLE TO KNOW HOW TO USE COMPUTERS AS TOOLS, IT IS EVEN MORE IMPORTANT THAT THEY LEARN HOW TO THINK, SOLVE PROBLEMS, MAKE DECISIONS, AND INTERACT USING THOSE TOOLS” (KEARSLEY, 1998).

EFFECTIVE DIGITALLY SUPPORTED PEDAGOGY

The way we conceive of technology in the classroom and the importance given to tech over pedagogy (Kearsley, 1998)

Shift in strategic goals of ed away from short-term outcomes to long term changes in cognitive style and well being was needed (Kerr, 1989)

Early ed tech didn’t facilitate teaching of different intellectual skills to students based on need (Johnson, 1981)

Failure to capitalized on personalized learning in a way that engages students (Kearsley, 1998)

Loss of hands-on opportunities and continued failure of progressive education (Lipson and Fisher, 1983)

Educators were unable or reluctant to revise curriculum to capitalize on technology’s potential contributions

INSUFFICIENT TECHNICAL TEACHER TRAINING

Demand for enhanced technical teacher training - core grievance

Perspectives on this vary depending upon job role
Problems exist because teachers don't follow ed tech plans to the letter (Kerr, 1989)

Educators as operants (Kerr, 1989)

Training should be ongoing and sustained. Training is unsuccessful because it focuses on how to use the technology, rather than how the technology might be used to teach (Kearsley, 1998)

PHILOSOPHICAL AND NORMATIVE CONCERNS OF ED TECH

HOW WILL ED TECH IMPACT THE LEARNING PROCESS LONG TERM? (JOHNSON, 1981)

If technology's application presented a roadblock for interpersonal skills, the ethics of IT within an educational context should be reconsidered (Johnson, 1981)

Changing roles of teachers in the classroom
Lack of commercial development of materials (Lipson and Fisher, 1983)

Intrusion of commercial interests in education (Spitzer, 1987)

The digital divide and the homework gap

TECHNOLOGY'S PROMISE VERSUS PERFORMANCE

"EDUCATIONAL TECHNOLOGY HAS FAILED TO HAVE THE KIND OF IMPACT ON EDUCATION AND SOCIETY THAT THE FOUNDING FATHERS OF THE FIELD HAD ENVISIONED FOR IT" (SPITZER, 1987)

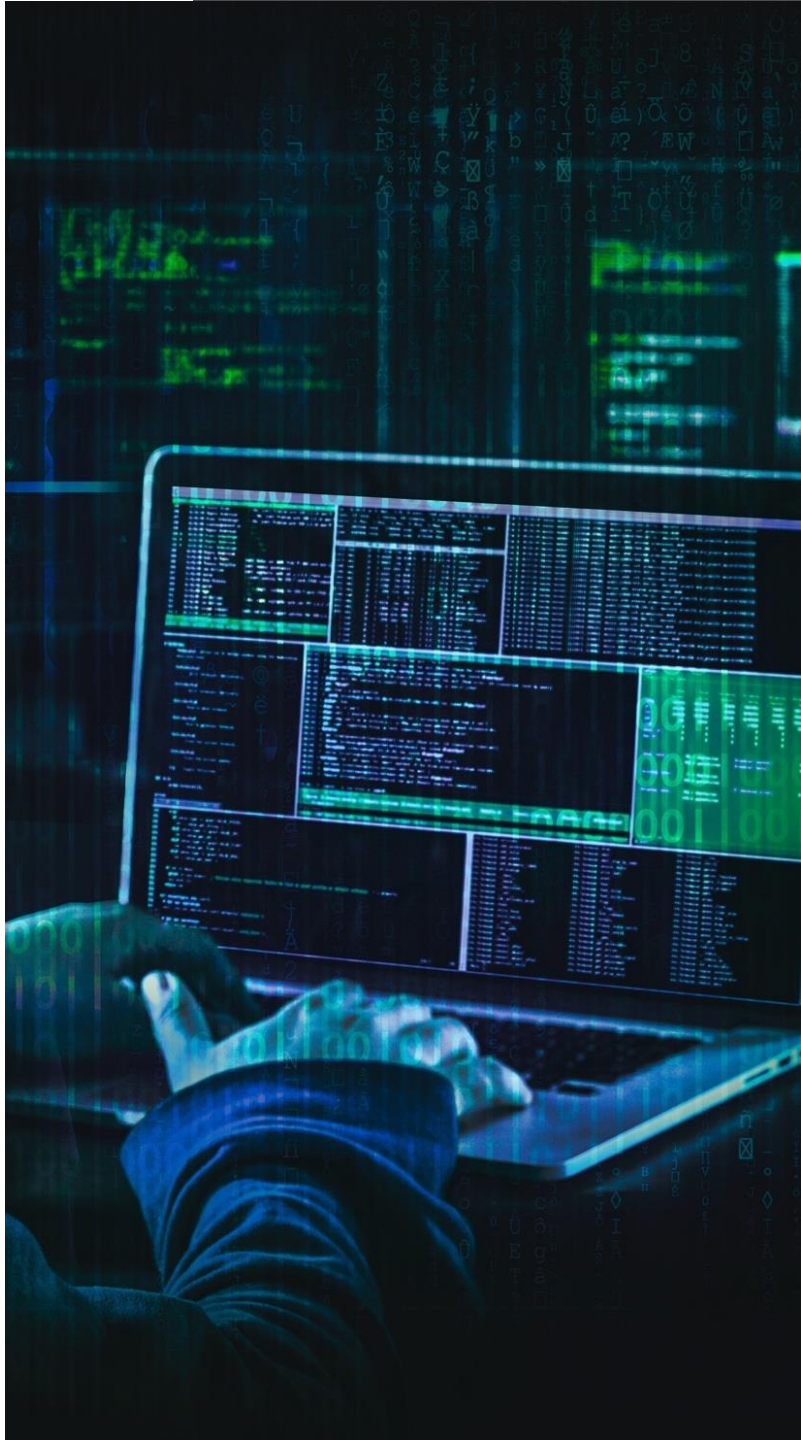
Need for pedagogical and epistemological reform (Kearsley, 1998)

Bifurcation of local vs national control

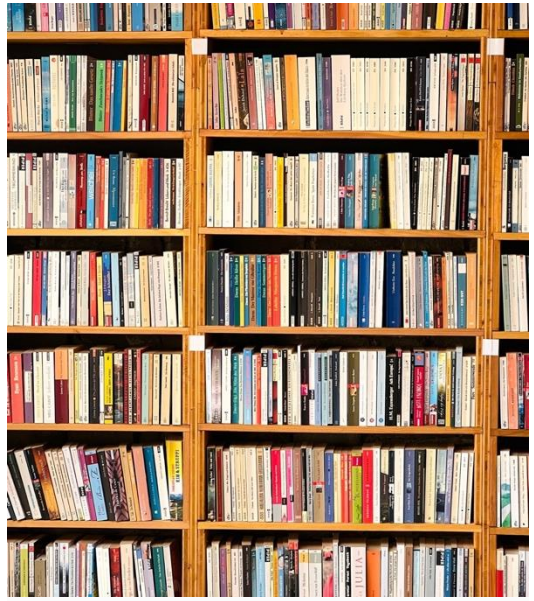
PROBLEMATIC ELEMENTS of AIEd

Normative + Ideological AIEd Dilemmas

- Algorithmic bias
- Governance and productization of data
- Digital surveillance
- Behavioral manipulation
- Harm to health and well-being
- Threats to democracy
- Pedagogical obstacles
- Cheating
- Automation/deskilling of teaching profession/job threats
- Spread of misinformation
- AIEd divide amongst institutions



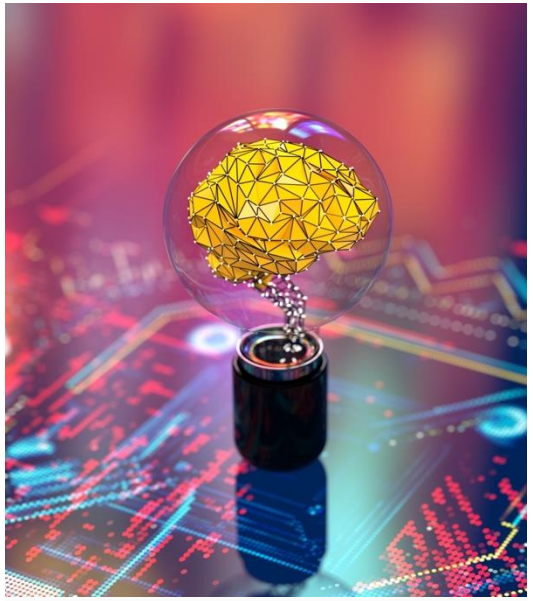
Mixed-Methods Approach



Literature review: Identify normative, ideological, and implementation challenges

Delphi method: Confirm findings from the literature, identify additional areas of concern

Document analysis: Identify ethical components from existing frameworks that are congruent with Stoic virtue + findings from Delphi

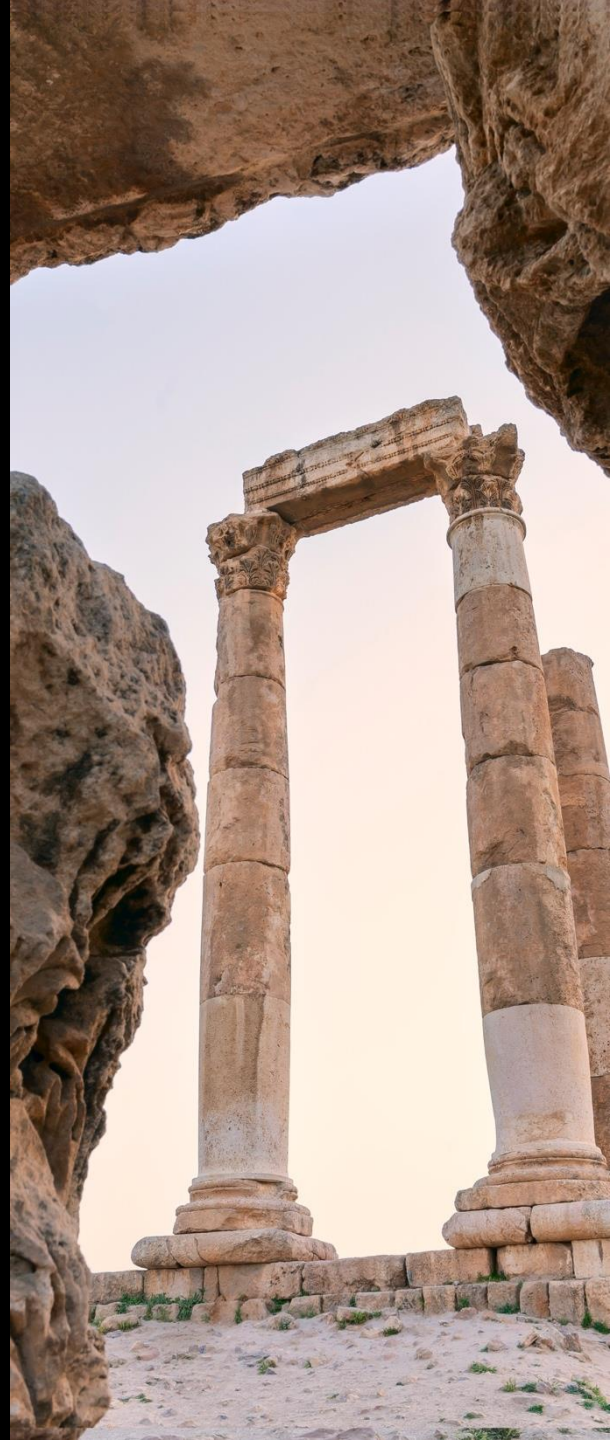


Propose supplemental components: Support fallow principles from document analysis

Stoicism as A Theoretical Framework

Eudaimonia is cultivated through:

- Cultivation of virtues
- Living in accordance with reason and nature
- Focusing on what one can control
- Adopting indifference to everything else



Prudence - the science of what ought and ought not to be done, and encompasses good judgment, shrewdness, and resourcefulness in difficult situations

Courage - knowing what ought and ought not to be tolerated, embodied through unflinching movement toward that which preserves our existence

Temperance - what ought and ought not to be chosen; knowing when action should be taken, and fortifying one's undertakings with control, honor, and blame avoidance

Justice - what ought and ought not to be distributed; encompasses concepts such as piety, kindness, compassion, doing good to others, and equality

Comprehensive Components

- Combination of document analysis, Delphi method and reverse analysis

- Resulted in comprehensive set of 90 ethical components
  - Most robust frameworks are similar, or larger in size

- Component areas included:
  - Ethics
  - Accountability
  - Transparency
  - Well-being
  - Autonomy
  - Equity and Inclusion
  - Pedagogy and Teacher Training
  - IT Risk Management
  - Administration Considerations
  - Supplier Scrutiny

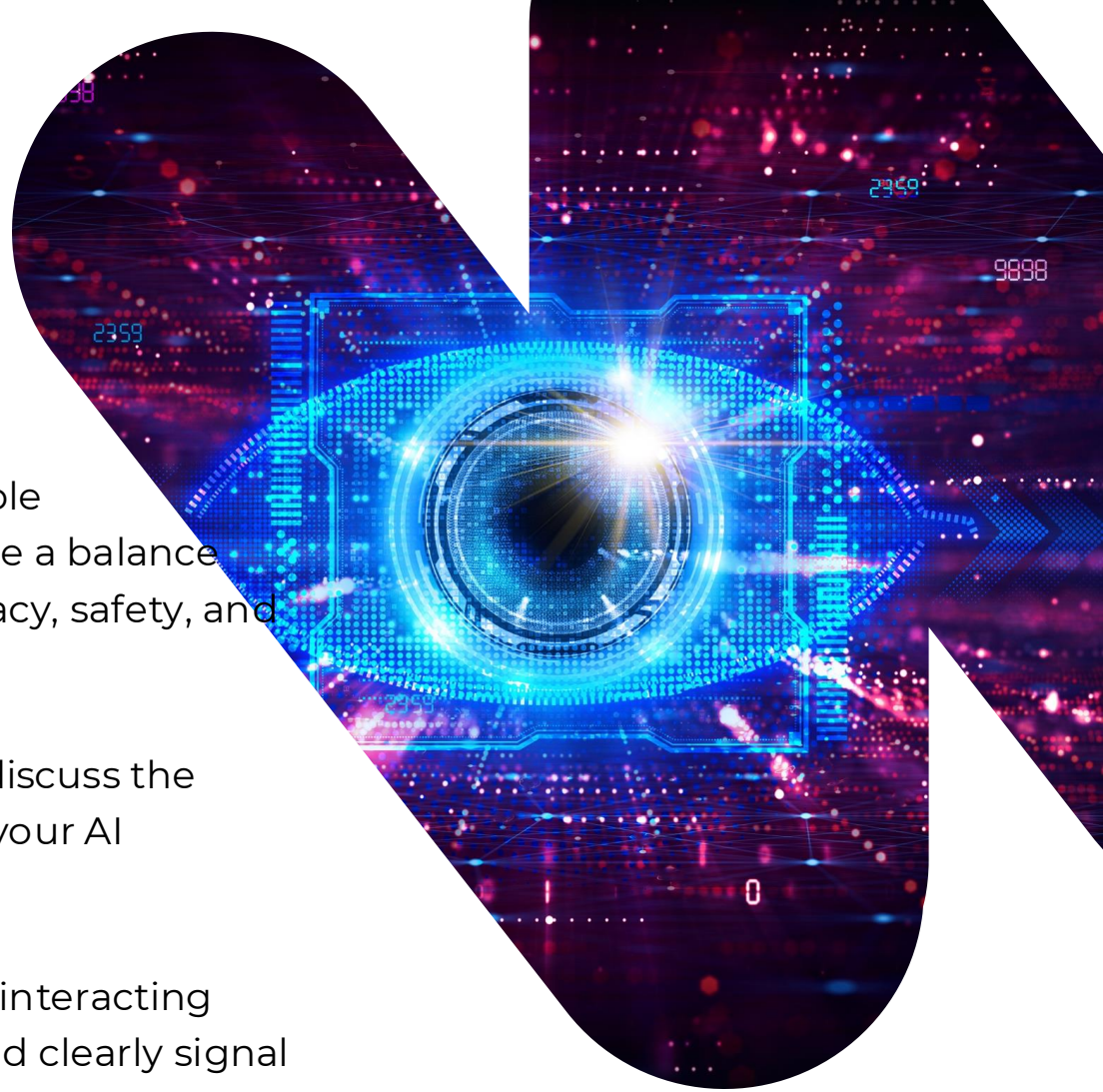
ETHICS COMPONENTS (9)

- Peace, inclusivity, justice, interconnectedness, democracy, equity, and freedom of expression should be promoted throughout the AI life cycle. AI systems should not undermine freedom, safety, or turn individuals against each other.
- Recognize that AI technologies do not ensure human flourishing.
- Develop mechanisms to instill the understanding of students' ethical use of AI tools in the learning process (e.g. ChatGPT).
- AI should always protect and empower users

ACCOUNTABILITY COMPONENTS (9)

- Remember that humans are ultimately responsible for educational outcomes, an AI system is not a replacement. Oversight of AI systems should come from humans, a governance group, and the public, as appropriate.
- Develop accountability assessment measures and requirements of AI systems and tools before implementation.
- Continually monitor the achievement of educational objectives and intended impacts for AI systems. When goals are not met, or the use of AI systems result in harm, investigate potential causes, such as its implementation or design, and create a remediation plan.
- Create enabling environments for stakeholders and the public to understand any benefits and negative consequences of AI's use in your institution.
- Assess and monitor the reproducibility or variability of decisions made by AI systems.

TRANSPARENCY COMPONENTS (7)



- The principle of transparency should be followed as much as possible within institutional data applications, however there may need to be a balance between transparency, explainability, and other principles like privacy, safety, and security.

- When the use of AI could be considered a surveillance of learners, discuss the justification of the direct or indirect benefits of this practice within your AI governance group.

- Users and their caregivers should easily understand when they are interacting with AI. The AI system, even when imitating human behavior, should clearly signal that its behavior is simulated and that the system has no emotional capacities or feelings. Ensure this explanation is written in age-appropriate language.

- Consider within your AI governance group whether the AI system should communicate to users that a decision, content delivery choice, advice, or outcome is the result of an algorithmic decision.

WELL-BEING COMPONENTS (12)

- If AI assessment models are continuous, designate safe spaces where students are not assessed.
- Emphasize students' agency and social well-being in the process of integrating AI-based tools: Protect students' agency and motivation to grow as individuals; protect play and leisure time, social interaction, and school breaks.
- Test and scale up evidence-based avenues of applying AI in learning; this includes fostering broad, transferable abilities including social-emotional skills, meta-cognition, collaboration, problem-solving, and creativity.
- When the risk of job loss or de-skilling is present, take action to counteract potential consequences.

AUTONOMY COMPONENTS (7)

- Ensure the AI system does not affect human autonomy by interfering with the user's decision-making process in an unintended way.
- Establish whether AI's use could undermine or marginalize the authority of practitioners or educators, or disrupt accountability structures and take action based on the assessment.
- AI systems should be used to increase the level of control that learners have over their learning and development, including the decision to opt out of their use.
- Promote data agency whenever possible. Users should retain control of their own data and digital identities.

EQUITY & INCLUSION COMPONENTS (6)

- Develop definitions for fairness, inclusion, and equity within your governance group.
- Build pathways for diverse students to enter the AIEd field.
- Establish and monitor measurable targets to ensure inclusion, diversity, and equality in AI decision-making, teaching practices, systems, and data sets.
- Encourage and facilitate the participation of individuals with diverse backgrounds and beliefs, marginalized people, girls and women, and students of all ages in AI leadership, governance, and decision-making roles within your organization.

PEDAGOGY & TEACHER TRAINING COMPONENTS (11)

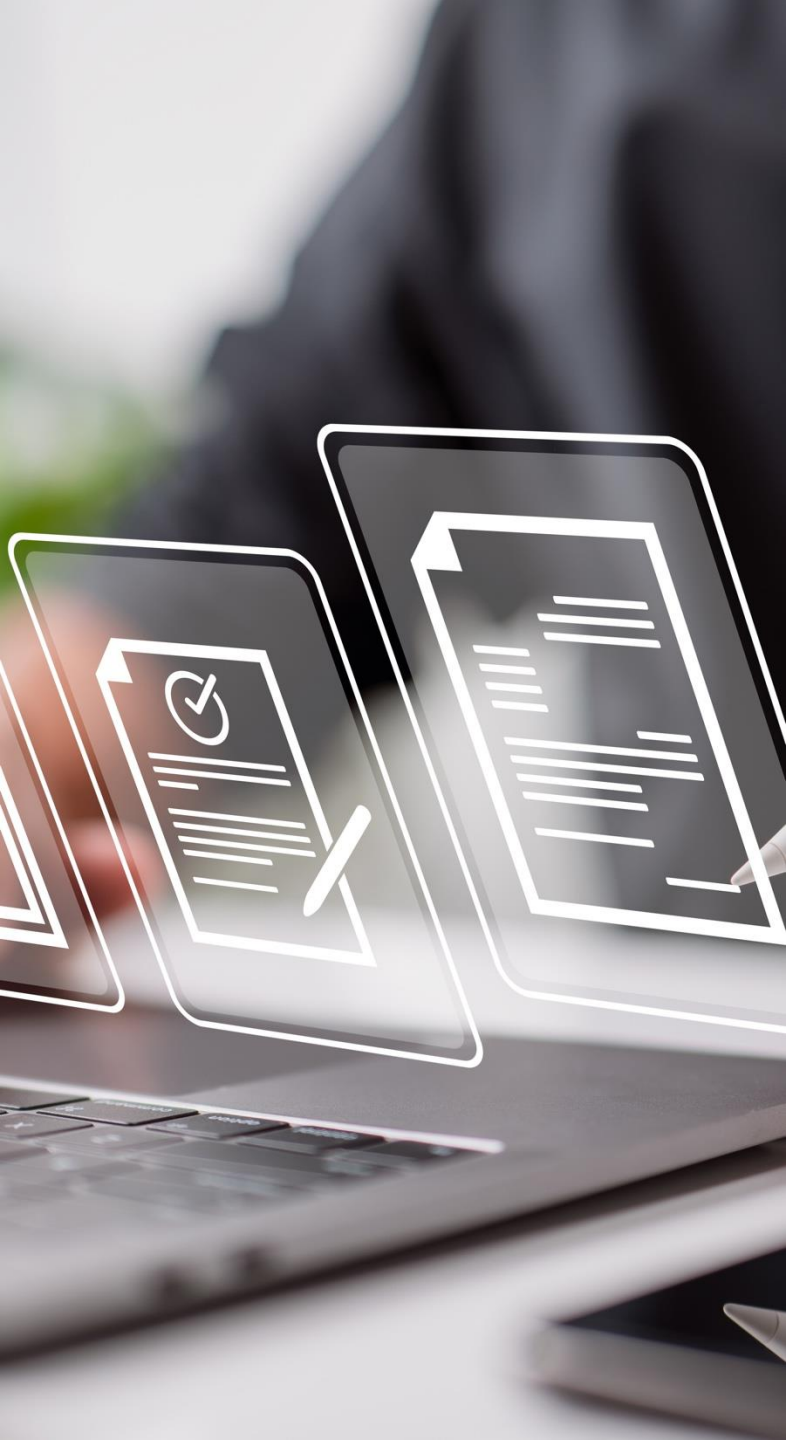
- Independently verify vendors' claims about AI's potentials based on pedagogical research, and ensure that student performance measures align with recognized evidence-based measurements.
- Have clear educational goals when AI is used in curriculum or assessment.
- Identify and develop AI literacy curriculum for students.
- Analyze potential changes in teachers' roles in facilitating knowledge transfer, higher order thinking, and human values.
- Analyze potential changes and best practices concerning the teaching and learning of human values, such as love, tolerance, and respect.

IT and RISK MANAGEMENT COMPONENTS (9)

- Consider necessary balances between privacy and the use of data to achieve organizational and educational goals.
- Require ongoing testing of AI systems for safety and security, including the consideration of vulnerabilities such as data pollution, physical infrastructure, and cyber-attacks. Develop organizational policies for risk prevention and mitigation.
- Implement a strong AI information technology training program to build stakeholders' understanding of safety and security protocols.
- Ensure AI systems are properly constrained to the places they can write data to; dynamic decisions made by AI can impact systems related or connected to AI tools.

ADMINISTRATION CONSIDERATIONS COMPONENTS (10)

- Set up a structure for policy governance and coordination of AIEd.
- Support knowledge transfer and capacity building on AI and its use in your educational environment for top management, administration, and decisionmakers.
- Identify, monitor, and evaluate the ways AI can increase organizational capacity and improve processes and educational information management while respecting human relationships.
- Consider the values of current practices before automating them with AI.



SUPPLIER SCRUTINY COMPONENTS (11)

- Insist that suppliers explain how their AI resource achieves the desired objectives and impacts, including information relating to algorithmic decision-making.
- Require confirmation that AI resources were not designed to coerce learners. Insist that suppliers demonstrate that when AI is used for behavioral influence, this influence is positive and supported by scientific evidence.
- Evaluate with suppliers that solutions are relevant to local needs in terms of educational best practices, diversity, context, cultures, and objectives.
- Data ownership should be actively examined. It is critical to understand which entities own the data that is generated by students as a result of interacting with AI systems, and the rights retained by the data owners.

AIEd Stakeholders

- Administrators (school, ISD, district, school board, etc.)
- Educational Technologists
- Educators
- Students
- Parents
- Community members
- Related Institutions

Ethical AI in Education Project Logic Model

How might the ethical components and governance recommendations for AIEd as identified by Dagg (2023) be further refined for understanding, achievability, and applicability across U.S. institutions?

Problem Statement

Educational institutions face increased pressure to increasingly adopt and use AI tools, but no holistic manual to guide the ethical application of AI in education (AIEd) exists.

Guiding Questions

SQ1: What considerations should exist for developing culturally appropriate intake models for AIEd programs, implementation, processes, sentiment, knowledge bases, concerns, needs, and policies?

SQ2: How should the principles identified in this study be leveraged as an ongoing decision-making model for technology governance beyond AI applications?

Inputs

- AIEd Framework for ethical decisionmaking and governance
- Research personnel
- Support personnel
- Funding
- Stakeholder governance group
- Participating institutions

	Activities	Outputs
Phase 1	Institutional onboarding Form stakeholder governance groups Review existing AIEd framework	Sustainable, inter-institutional, collaborative technology governance groups are formed
Phase 2	Develop culturally appropriate intake models Present results of intake research to governance groups	Intake models are developed
Phase 3	Framework component reviews Develop/identify evaluation metrics (if appropriate) for AI tools and applications Monthly virtual institute community of practice meetings	Evaluation metrics are developed
Phase 4	Develop Strategic AI adoption plan for each stakeholder Identify professional development needs by role Conduct summative AIEd discussion Refine AIEd framework and guidelines Conduct project evaluation	Strategic AI adoption plans are developed Professional development recommendations are created Revised guidelines for ethical AIEd and related governance models are written
Phase 5	Disseminate project findings and resources	Case studies, implementation guides, and webinars are created

Outcomes

Stakeholder institutions are empowered to evaluate and select AI tools for adoption that do not compromise cultural priorities or introduce additional inequity into the learning system

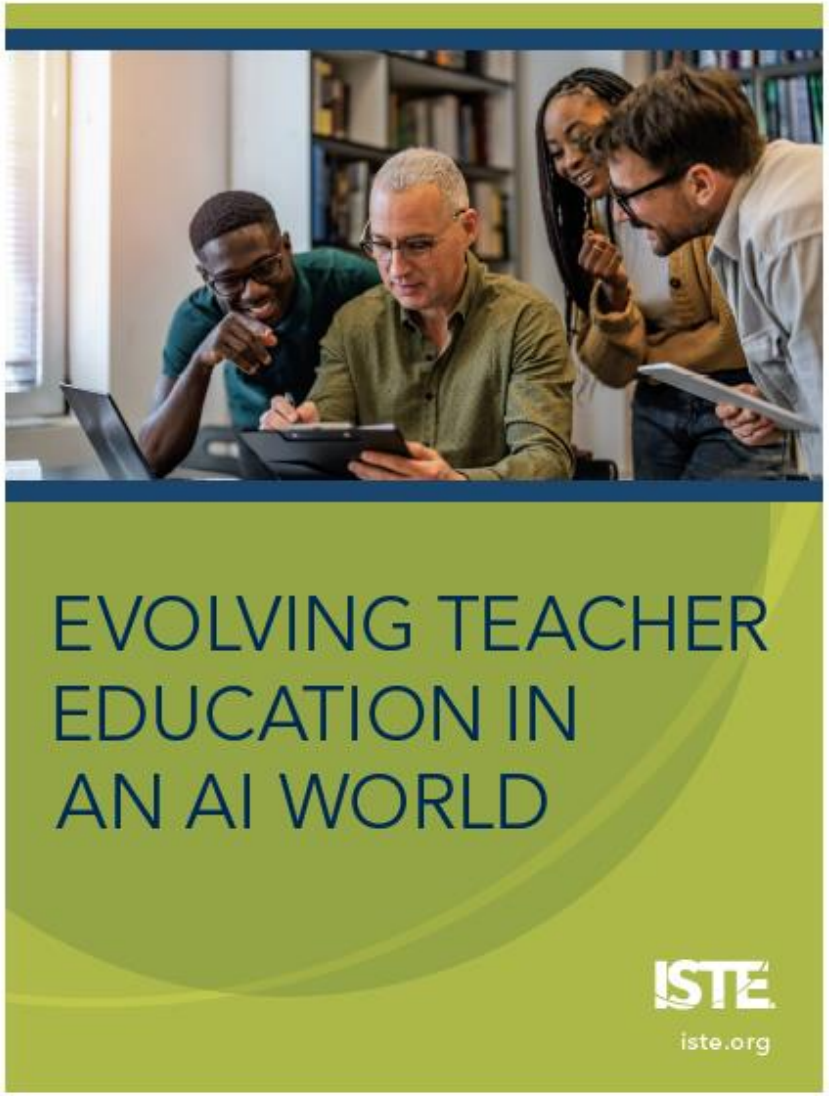
Stakeholders adjust policies and decision-making based on AIEd

Faculty, staff, and administrators are confident in AI tool adoption and use

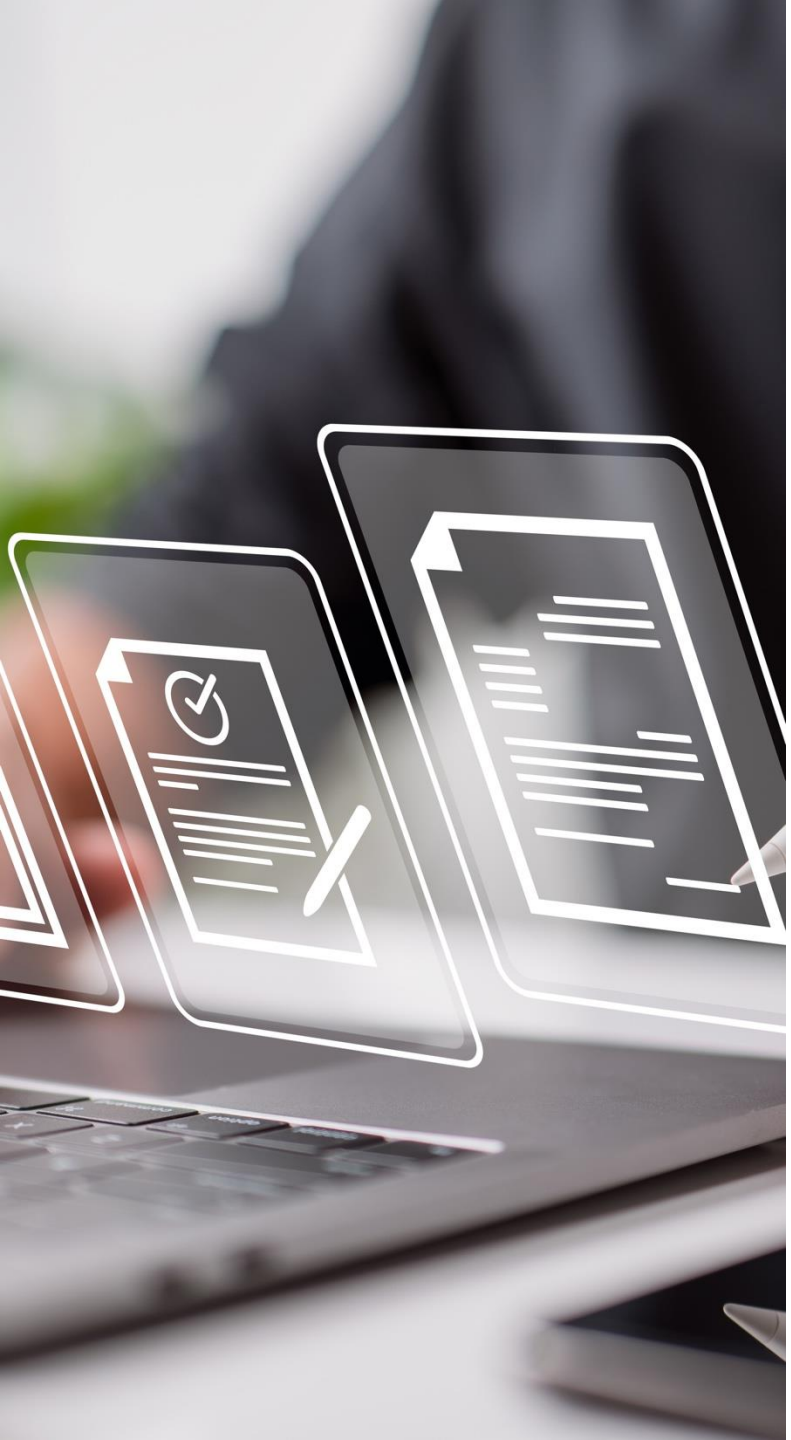
Increased awareness of cultural and community impacts and ethical decision-making among higher education institutions regarding AI adoption

Governance Model

- Begin with a review of existing frameworks and decision-making models in your institution
- Deploy intake models to identify current applications, sentiment, processes, knowledge base, and policies
- Organize regular cadence to review each component area of ethical framework against current and future AI use
- Develop strategic AI adoption plan, including remediation to current policy and use, professional development needs, and oversight mechanisms



https://info.iste.org/openai-white-paper?hs_preview=WWHGnwJv-168688965032&utm_campaign=ISTESproutSocial&utm_medium=social&utm_source=linkedin.com



VISION AND EXPLORATION: Laying the groundwork for AI adoption

- Understanding the Landscape: Engage with stakeholders to assess current AI use and identify needs, gaps, and risks.
- Establishing Vision: Set clear, balanced objectives for both immediate and long-term AI integration.
- Key Actions:
  - Gather insights on perceptions and usage of GenAI.
  - Develop a phased AI roadmap focusing on integration, training, and curriculum development.
- Key Questions:
  - How is AI currently enhancing teaching and learning?
  - What long-term changes are envisioned through AI integration?

STRATEGY and SUPPORT: Deepening understanding and building governance

- Faculty Development: Enhance understanding of GenAI's potential and limitations. Model effective use and ethical practices.

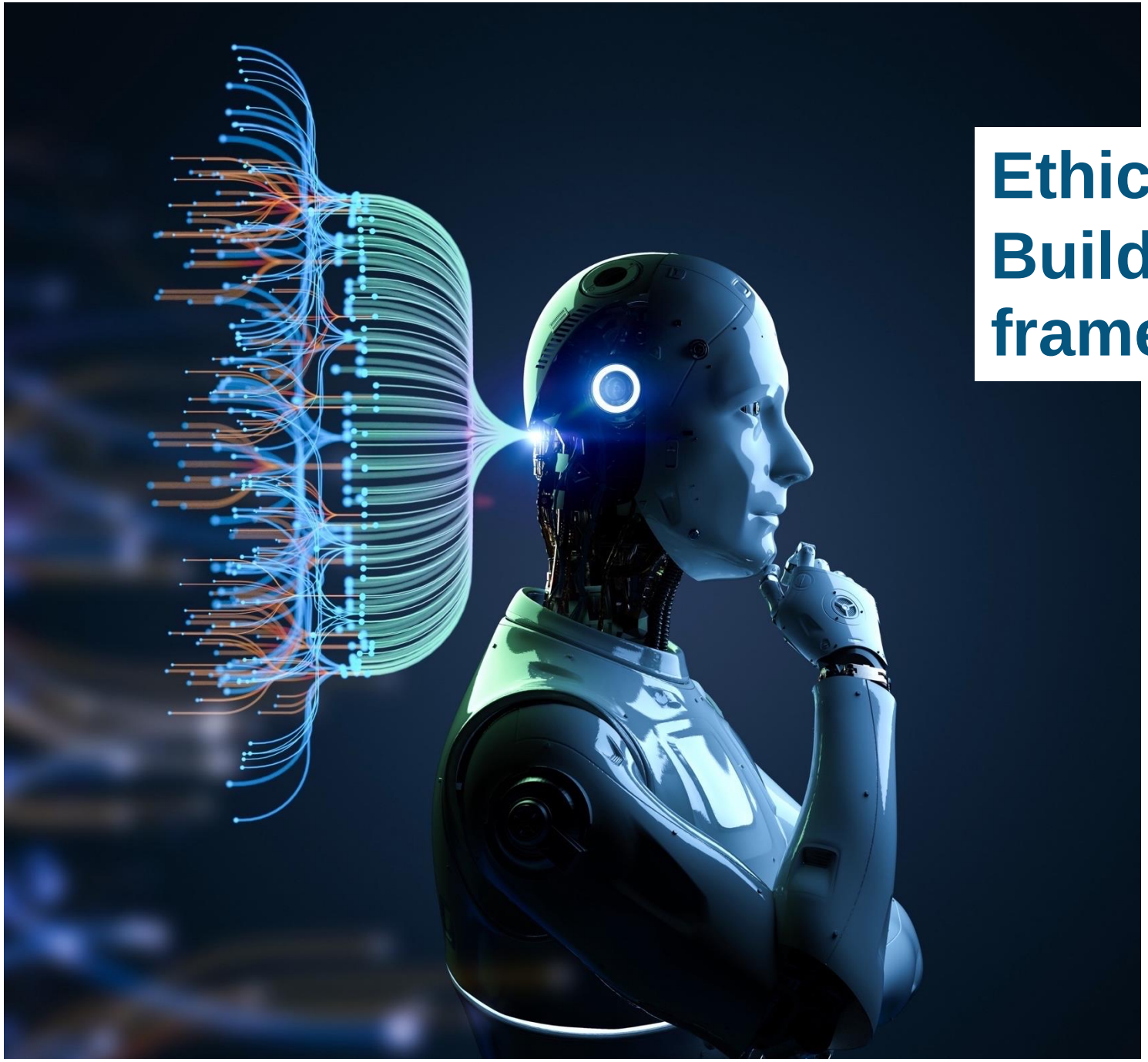
- Governance and Partnerships: Establish governance structures that include diverse voices. Build partnerships within and beyond the institution.

- Key Actions:
  - Provide professional learning resources and ongoing training for faculty.
  - Create interdisciplinary governance groups to evaluate and refine AI tools and strategies.

- Key Questions:
  - What are the most promising AI tools for educators?
  - How can partnerships enhance AI integration and effectiveness?

IMPLEMENTATION and ETHICAL CONSIDERATIONS: Executing AI integration while addressing risks

- Skill Modeling for Teacher Candidates: Focus on operational efficiency and ethical AI use.
- Mitigate Risks: Develop shared visions for responsible AI use, update policies, and ensure ethical application.
- Key Actions:
  - Identify tasks that AI can accelerate, providing more time for critical human tasks.
  - Review and adjust academic integrity policies and grading practices to incorporate ethical AI use.
- Key Questions:
  - What educator tasks can GenAI streamline?
  - How will we safeguard data privacy and maintain academic integrity with AI tools?



Ethical AI in Education: Building transparent frameworks for the future

Pierrette Renée Dagg, Ph. D

Director of Research, Merit Network

Research Associate, Quello Center for Media & Information Policy, Michigan State University

Adjunct Instructor, Oakland Community College