

AI and Assessment: Navigating Ethical Implementation and Future Possibilities

Workshop Report

November 2025

This workshop was hosted in June 2025 by the AI & Education Community at the University of Cambridge (Faculty of Education) and organised in partnership with ai@cam, Cambridge University Press & Assessment, Google Research, and the Accelerate Programme for Scientific Discovery

Authors:

Megan Ennion, Faculty of Education, University of Cambridge

Inbar Bobrovsky, University of Cambridge

The authors would like to thank Dr Martina Kuvalja (Education Futures, Cambridge University Press & Assessment) for reviewing the earlier drafts of this report.

Executive Summary

On 23 June 2025, the AI & Education Community at the University of Cambridge convened a multidisciplinary event at the Faculty of Education to address one of the most urgent issues in contemporary education: how artificial intelligence is reshaping assessment, and how to ensure its implementation is both ethical and educationally valuable. Organised in partnership with ai@cam, Cambridge University Press & Assessment, Google Research, and the Accelerate Programme for Scientific Discovery, the gathering brought together leading researchers, educators, and industry representatives to explore the rapidly evolving relationship between AI and educational assessment, and to consider its implications for the future of education.

The discussions revealed a field in transition, where technological capabilities are advancing more rapidly than institutional readiness and regulatory frameworks.

Participants identified significant opportunities for AI to enhance educational assessment while highlighting serious concerns about equity, authenticity, and the preservation of human judgment in educational processes.

Many participants cautioned that assessment goals should not be dictated by technological capabilities or limitations. Instead, the design and purpose of assessment must remain anchored in educational priorities, with technology serving those priorities rather than defining them.

Four key tensions emerged:

- **Tools versus Systems** - The workshop highlighted a tension between focusing on individual AI tools and understanding AI integration as part of broader systemic transformation. While immediate practical needs drive attention to specific applications, participants emphasised the importance of systematic, higher-level thinking about how AI changes entire educational ecosystems.
- **Binary Decision-Making versus Iterative Learning** - Participants identified a tension between institutional pressure for binary yes/no decisions about AI adoption and the recognition that AI integration represents a wicked problem requiring iterative learning, experimentation, and correction over time.
- **Efficiency versus Educational Authenticity** - While AI offers unprecedented speed and consistency in assessment, participants questioned whether this efficiency comes at the cost of meaningful evaluation of genuine learning and the nuanced understanding that human judgment provides.
- **Innovation versus Equity** - Rapid AI adoption offers potential benefits for educational enhancement, but risks creating new forms of inequality between students with access to advanced AI tools and those without, potentially reinforcing existing educational disparities rather than addressing them.

The workshop concluded with recognition that this represents a unique moment in time for the educational community to exercise agency within the evolving ecosystem, using current challenges as an opportunity to redefine, question, and reopen fundamental questions around epistemology and the goals of education itself. Rather than simply adapting to technological change, participants were invited to view this as an opportunity to actively shape how technological and social systems develop in relation to educational values and purposes.

The sections that follow explore these themes in detail, drawing on speaker contributions, workshop discussions, and emerging research to outline both the opportunities and the risks of AI in assessment.

Workshop Overview

The workshop featured presentations from distinguished speakers across multiple disciplines:

- **Dr Steven Watson** (University of Cambridge) explored generative AI and assessment challenges
- **Dr Carla Pastorino-Campos and Dr Jing Xu** (Cambridge University Press & Assessment) examined ethical principles in AI-powered language assessment
- **Dr Shashank Agarwal** (Google Research) presented industry perspectives on AI in educational feedback
- **Dr Alexandru Marcoci and Dr Roni Tibon** led discussions on automated marking systems through the OpRaise project
- **Dr Thomas Corbin** (Deakin University, CRADLE) addressed fundamental questions about AI assessment integration
- **Dr Isabel Fischer and Dr Neha Gupta** (University of Warwick) explored responsible AI use in collaborative assessment ecosystems
- **Dr Miri Zilka** (University of Cambridge) examined risks of automation in educational contexts
- **Dr Imogen Casebourne** (Digital Education Futures Initiative (DEFI), Hughes Hall, University of Cambridge) facilitated discussions on considerations of ethical issues about future assessment practices

Interactive workshops enabled participants to engage directly with current research projects and emerging technologies, fostering dialogue between theoretical frameworks and practical implementation challenges.

Current State of AI in Assessment

Technological Capabilities and Limitations

Workshop discussions highlighted that AI technologies are advancing rapidly, with tools being developed that aim to provide educationally aligned assessment solutions. Dr Shashank Agarwal from Google Research presented pedagogical principles, including key concepts such as inspiring active learning, managing cognitive load, and deepening metacognition. These illustrate industry efforts to move beyond purely technical solutions toward educationally informed applications. Similarly, Dr Carla Pastorino-Campos and Dr Jing Xu from Cambridge University Press & Assessment explored ethical frameworks for AI in language assessment, focusing

on how to harness AI technologies while maintaining the integrity, quality, and fairness of human-centred teaching and assessment practices.

However, the workshop showed significant tensions between technological capabilities and educational requirements. Dr Agarwal acknowledged that foundational models struggle with pedagogical instructions, highlighting the persistent gap between technological sophistication and educational nuance. This limitation reflects broader challenges identified throughout the day about the disconnect between what AI can do technically and what education actually needs.

The OpRaise project, presented by Dr Alexandru Marcoci and Dr Roni Tibon, shared preliminary findings from their pilot testing showing that individual AI model accuracy reached approximately 60%, while adjacent match accuracy ranged from 93-98%. While the higher adjacent match figure suggests promise, the 60% exact match rate indicates a substantial margin of error, especially in high-stakes assessment contexts where precision is critical. These findings prompted discussions about whether current AI assessment accuracy is sufficient for educational use, particularly in high-stakes contexts.

Workshop participants critically examined the assumption that technological advancement automatically translates to educational improvement. Speakers questioned whether the current focus on efficiency and scalability aligns with educational goals of fostering deep learning, creativity, and critical thinking. The gap between what foundational models can achieve and what pedagogical practice requires emerged as a central challenge requiring sustained attention rather than technical solutions alone.

The workshop also highlighted variability in AI tool performance across different educational contexts and student populations. Participants noted that what works in one setting may not transfer to others, challenging industry assumptions about scalable, universal solutions for educational assessment challenges.

Implementation Disparities

A critical finding across multiple presentations was the significant variability in how AI tools are being used across different educational contexts, raising fundamental questions about equity and access. The workshop challenged common assumptions about student engagement with AI technologies, with participants noting that assumptions about universal enthusiasm for AI among young people are often unfounded and may not accurately reflect the diverse experiences and preferences of students.

These implementation disparities create several interconnected challenges. First, unequal access to AI technologies risks creating new forms of educational inequality, where students with access to advanced AI tools gain advantages over those without. Second, the variability in implementation approaches makes it difficult to develop consistent quality standards or evaluate the effectiveness of AI assessment tools across different contexts. Workshop participants noted that successful AI integration requires careful consideration of local contexts, institutional capacity, and cultural factors, rather than relying on universal deployment strategies.

The challenge of implementation extends beyond technical deployment to include educator preparation, institutional policy development, and student support. Participants identified significant gaps between the pace of technological development and institutions' capacity to implement AI tools thoughtfully and effectively. This mismatch creates risks of hasty adoption without adequate consideration of educational consequences or proper support for all stakeholders involved.

The Regulation and Policy Gap

The most urgent challenge identified was the growing disconnect between technological capability and regulatory readiness. Perhaps most concerning was the identified disconnect where regulation, policy, and decision making considerably lag behind the reality of how generative AI is being understood and used in educational contexts. This creates what Dr Steven Watson described as a significant gulf between what is now possible technologically and what is actually happening in assessment practice.

This regulatory lag manifests in several problematic ways. Educational institutions find themselves making critical decisions about AI assessment tools without clear policy guidance or regulatory frameworks. The absence of sector-wide standards means that practices vary widely between institutions, potentially creating inconsistencies in how students are assessed and evaluated.

Workshop participants noted that current policy responses often treat AI as either a universal solution or a problem to be solved, rather than recognising the complex, contextual nature of AI integration challenges. The lack of nuanced policy frameworks leaves educators to navigate ethical and practical challenges without adequate institutional support.

Examples discussed included the absence of clear guidelines around consent for AI-powered assessment, inconsistencies in data protection practices when using AI tools, and uncertainty about how to ensure fairness when AI systems are used to evaluate student work. Participants noted that some institutions have developed individual

policies, but the lack of coordination across the sector creates confusion and potential inequities.

The workshop highlighted the need for policy development that moves beyond technical considerations to address fundamental questions about educational purpose, student agency, and the role of human judgment in assessment. Current regulatory gaps leave critical questions unresolved regarding accountability, transparency, and the long-term consequences of integrating AI in educational evaluation. Addressing this gap will require coordinated sector-wide policy development, informed by research, that balances innovation with accountability — a theme explored in the subsequent sections of this report.

Critical Concerns and Risks

Threats to Assessment Integrity

Without human oversight, AI could both generate and assess work — a scenario that workshop participants saw as a direct threat to the fundamental purpose of educational assessment. The most immediate concern involves scenarios where AI systems generate student work which is then assessed by other AI systems, effectively removing human judgment from the educational process. This threat to assessment integrity extends beyond concerns about academic dishonesty to fundamental questions about educational purpose. If AI systems are both generating and evaluating academic work, the traditional relationship between assessment and learning becomes deeply problematic.

Participants described this as “masking by output,” where significant AI input makes it difficult to assess a student’s genuine understanding and intellectual growth. This prompted calls for assessment design that foregrounds process and transparency, ensuring educators can see *how* students arrive at their work and preserve the authenticity needed for assessment to fulfil its diagnostic and developmental functions.

The integrity challenge is compounded by the sophistication of current AI tools, which can produce work that appears to demonstrate understanding and learning, but may primarily reflect AI capabilities rather than student knowledge. This makes it increasingly difficult for educators to distinguish between genuine student learning

and AI-assisted output, undermining the reliability of assessment as a measure of educational progress.

Participants noted that current approaches to detecting AI-generated content are insufficient and may become obsolete as AI technologies continue to advance. This creates an ongoing challenge for maintaining assessment integrity, which requires a fundamental reconsideration of assessment design rather than relying solely on technical solutions.

Dehumanisation of Education

Workshop participants expressed deep concern that excessive automation risks eroding the human relationships and skills that make education meaningful. The concern centres on the loss of human presence and the development of essential human skills that occur through meaningful educational relationships and interactions.

This dehumanisation manifests in several problematic ways. First, increased reliance on AI assessment tools may reduce opportunities for the kind of personalised feedback and mentoring that supports student development. The human capacity to understand context, emotion, and individual circumstances becomes diminished when assessment processes become primarily automated.

Second, workshop participants identified risks of normalising digital surveillance through AI-powered monitoring and assessment systems. The question of where to draw the line between ongoing digital feedback and surveillance became a central concern, with participants noting that excessive technological monitoring of students may create educational environments that prioritise data collection over human development.

Workshop participants also noted that students may be particularly affected by excessive technological measurement and recording of their educational activities. The long-term psychological and developmental effects of constant AI-powered assessment and feedback remain largely unknown, creating concerns about how such systems might shape student self-perception and learning approaches.

Third, the dehumanisation concern extends to the impact on educators themselves. As AI systems take over more assessment functions, there may be a corresponding erosion of professional judgment and expertise among educators. This could lead to a deskilling of the teaching profession and a loss of the human insights that make education meaningful.

Equity and Fairness Challenges

The integration of AI in assessment raises complex and multifaceted questions about fairness and justice in educational evaluation. Workshop discussions highlighted concerns about how AI systems handle student identity and privacy, with particular attention to the risk that AI-powered assessment judgments may follow students throughout their educational journey rather than being confined to individual courses or teachers.

The equity challenges manifest in several interconnected ways. First, there are concerns about whether AI assessment tools provide robust outcomes across diverse student populations or whether they risk reinforcing existing educational inequities. Workshop participants questioned whether current AI systems adequately account for different cultural backgrounds, learning styles, and forms of expression that students bring to their educational experiences.

Second, participants expressed concern about AI systems potentially promoting homogeneity by favouring particular types of responses or approaches to learning. Several participants emphasised that diversity in expression and approach should be recognised as an asset in assessment, and that AI systems must be designed to support and reward such variety. The risk of diminishing diversity in educational contexts has become a central concern, with questions about how to ensure AI assessment rewards and supports psychological and cognitive diversity, rather than pushing toward standardised responses.

Third, the workshop identified challenges around inclusivity and access. Unequal access to AI tools and technologies may create new forms of educational advantage and disadvantage. Students with access to sophisticated AI assistance may gain unfair advantages in AI-assessed contexts, while those without such access may be penalised.

Fourth, there are fundamental questions about consent and agency in AI-powered assessment systems. Workshop participants questioned who has the right to consent to AI assessment, particularly for younger students, and how to ensure that students and families have meaningful control over the technological systems used to evaluate their educational progress.

The equity concerns are complicated by the observation that AI systems may exhibit preferences for AI-generated content, creating potential feedback loops that favour students who use AI tools over those who rely primarily on their own capabilities. This raises questions about whether AI assessment can ever be truly fair or whether it inevitably advantages particular approaches to learning and expression.

Quality and Reliability Questions

Workshop participants raised fundamental questions about both the quality and impact of AI in educational assessment, as well as whether current AI assessment tools can achieve sufficient reliability across diverse student populations and educational contexts. The concern emerged that current approaches might be reinforcing existing inequities rather than addressing them, while simultaneously introducing new forms of unreliability into assessment processes.

The reliability challenges manifest in multiple dimensions. Technical reliability concerns focus on the consistency and accuracy of AI assessment tools across different types of student work and educational contexts. The OpRaise project, presented by Dr Alexandru Marcoci and Dr Roni Tibon, demonstrated variations in AI assessment accuracy, with individual model accuracy reaching approximately 60% while adjacent match accuracy ranged from 93-98%. While the high adjacent match rate suggests potential, the relatively low exact match rate signals a substantial margin of error — unacceptable in high-stakes contexts. This suggests that while AI tools may provide generally appropriate feedback, they may struggle with the precise evaluation of student work.

Educational reliability concerns centre on whether AI assessment tools actually measure what they purport to measure. Workshop participants questioned whether AI systems can adequately evaluate complex learning outcomes such as critical thinking, creativity, and deep understanding, or whether they primarily assess surface-level features that may not reflect genuine educational achievement.

The quality concerns extend beyond technical performance to fundamental questions about what constitutes good assessment practice. Participants questioned whether the efficiency gains offered by AI assessment tools come at the cost of educational quality, and whether the speed and convenience of automated assessment may lead to superficial evaluation practices that do not support student learning and development.

Workshop discussions also highlighted concerns about the trade-offs involved in AI assessment adoption. While AI tools may offer benefits such as consistent application of marking criteria and rapid feedback, they may also introduce biases, reduce attention to individual student needs, and diminish the role of professional educator judgment in assessment processes. Understanding and managing these trade-offs have emerged as a critical challenge for educational institutions considering the adoption of AI assessment.

Taken together, these concerns underscore the need for AI assessment approaches that prioritise human judgment, equity, and educational purpose over speed and efficiency.

Emerging Paradigms and Reframing Approaches

Moving Beyond Reductionist Models

In considering the role of AI in assessment, Dr Steven Watson cautioned against replicating the limitations of existing systems. Dr Watson offered a critical analysis of the dominance of standardised, reductionist models of assessment—systems that treat learning as a set of fixed traits to be measured and quantified. He argued that while such models may provide the appearance of objectivity and comparability, they inevitably oversimplify the rich, non-linear, and relational nature of real learning. In practice, these models risk prioritising what is easy to measure over what is educationally valuable, and can entrench rigidity and exclusion in assessment systems.

As an alternative, Dr Watson proposed the concept of assessment as adaptive coordination. This approach views assessment as a relational, evolving process in which learners, educators, and assessment tools are engaged in ongoing dialogue and mutual adjustment. The goal is to support learning as it develops, rather than to capture a static snapshot of attainment. In such a model, evidence of learning might be gathered from multiple, complementary sources, including portfolios, modular credentials, and dialogic feedback that allow for nuanced interpretation and contextual understanding.

Dr Watson emphasised that adaptive coordination is not a fixed method but a guiding principle for designing assessment systems that are flexible, context-sensitive, and responsive to the learner’s trajectory. Importantly, it shifts the emphasis from ranking and sorting students to supporting their growth over time.

However, participants also acknowledged that implementing such an approach alongside AI technologies introduces significant complexity. Adaptive coordination requires a sophisticated understanding of both human learning processes and the capabilities and limitations of AI tools. It demands systems that can integrate machine-generated insights with human judgement in a way that preserves educational values, respects contextual differences, and avoids the risk of AI simply reproducing — or even amplifying — the same reductionism it was meant to replace. This was one of several calls during the workshop to move beyond static, standardised approaches and to explore more dynamic, values-driven models of assessment.

Collaborative Assessment Ecosystems and Process-Oriented Assessment

A recurring theme in workshop discussions was that AI in assessment should be designed to enhance — not replace — human judgment. Multiple speakers emphasised that successful AI integration must maintain humans at the centre of educational assessment, with AI serving as a supportive tool rather than a replacement for human judgment. This involves ensuring that human insight, emotion, and contextual understanding remain central while leveraging AI capabilities to enhance, rather than supplant, human evaluation processes.

Dr Isabel Fischer and Dr Neha Gupta described this principle as “unwavering human accountability” and proposed reframing AI use as part of a collaborative assessment ecosystem in which students, educators, and technologies work together. This approach moves beyond binary thinking such as “AI-written = cheating” or “AI-marked = unethical” and positions AI as one contributor within a broader, interconnected environment.

Within such an ecosystem, they argued, assessments should incorporate process-focused elements—for example, reflective accounts of AI use, tailored visual outputs or data representations, and AI interaction logs—so that the role of AI is transparent and open to critical evaluation. They also emphasised a distinction between low-stakes and high-stakes assessment contexts: the former providing space to experiment with AI-supported approaches, and the latter requiring clearer boundaries and more stringent oversight to ensure fairness and reliability.

To support responsible practice, Fischer and Gupta recommended concise, one-page institutional principles for AI use in marking and feedback. These should:

- Require human oversight at all stages, with AI outputs never determining marks independently.
- Restrict use to approved, privacy-compliant tools, with policies against uploading identifiable student work without consent.
- Establish transparency as standard, with educators explaining how AI is used and inviting student feedback.
- Encourage reflection on how AI use affects perceptions of authenticity, trust, and fairness.

They also recognised human–AI collaboration as a valid form of assessed output. Rather than viewing AI involvement as automatically diminishing authenticity, Dr Fischer and Dr Gupta argued that assessments can explicitly value the interplay between human creativity, critical judgement, and AI-generated contributions. This,

they suggested, can be mutually beneficial—enabling students to produce richer work while refining AI systems through human input. This approach also values diversity in expression and process, ensuring that AI-supported assessment rewards varied forms of engagement rather than promoting homogeneity.

Participants noted that while the ecosystem framing offers a valuable conceptual model, it also raises complex questions about maintaining educational values and defining responsibility and authenticity in increasingly hybrid human–machine environments. As such, the collaborative ecosystem model offers both a practical framework for current policy development and a long-term vision for human–AI partnership in assessment.

Contextualised Solutions

One of the workshop’s strongest messages was that AI in assessment must be shaped to fit the context — not the other way around. A strong theme throughout the workshop was the recognition that assessment practices vary significantly across different educational contexts and require tailored responses rather than universal approaches. Participants emphasised that different academic fields face varying degrees of challenge when integrating AI assessment tools, highlighting the inappropriateness of one-size-fits-all solutions for AI assessment integration.

This emphasis on contextualisation reflects deeper understanding that AI integration cannot be separated from broader questions about educational purpose, institutional culture, and local values. However, the focus on context also creates challenges for developing coherent approaches across educational systems and raises questions about how to balance local responsiveness with broader educational goals and equity considerations.

Several participants stressed the need for AI understanding and AI literacy. Institutions and educators, they argued, should not implement AI tools without a clear grasp of how these systems operate, their limitations, and the implications for learning and assessment. Without this knowledge, the sector risks adopting tools that are misaligned with educational aims or that carry unintended ethical and pedagogical consequences.

Discussion also highlighted that diversity and disagreement are valuable assets in shaping AI policy and practice. Encouraging a range of ideological and practical perspectives fosters critical engagement, challenges assumptions, and helps to avoid homogenisation of thought — a particular risk when AI systems themselves may reward conformity.

Reframing Ethics

Dr Imogen Casebourne argued that if AI is to be integrated into education responsibly, ethics must be understood as a dynamic, long-term practice — not just a checklist of present-day rules. Dr Casebourne invited participants to take a broader and more nuanced view of ethics in the context of AI in education, arguing that ethical reflection should not be limited to immediate practical concerns or compliance with existing codes. Instead, she proposed considering ethics across time—examining the past foundations, present uses, and long-term futures of AI technologies. This temporal lens requires asking whether the origins of these systems are just, whether their current applications are ethically sound, and how decisions made today might shape educational values and norms decades into the future.

She also urged moving beyond rule-based ethics—which tend to rely on universal prescriptions or utilitarian calculations—towards relational, virtue-based, and speculative approaches. In this view, the ethical evaluation of AI in assessment must account for its effects on wellbeing, human flourishing, justice, and what Gert Biesta terms subjectification (the process by which a person becomes a unique responsible subject, capable of independent thought and action). Rather than focusing solely on whether research into the development of an AI system complies with a fixed set of rules (such as BERA guidelines), this approach additionally requires a consideration of whether it will in future contribute to the kind of educational relationships and societal outcomes.

Dr Casebourne also highlighted that what a society values in education is not static — curricula, assessment priorities, and broader social norms evolve over time. Future assessments may need to reflect very different priorities from those we hold today. This makes AI in education a moving ethical target: technologies adopted now could lock in assumptions and value hierarchies that later prove outdated or harmful.

Given this uncertainty, she emphasised the ethics of emerging technologies, drawing on thinkers such as Lucivero, Brey, and Danaher to argue for speculative yet rigorous ethical reflection. This involves developing tools and practices that allow us to explore plausible futures, evaluate the moral implications of each, and prepare for long-term consequences that may not be immediately visible. Such foresight requires imagination, but also methodological discipline, to ensure we are not simply reacting to the present but actively shaping an ethical trajectory for AI in education. This long-view ethical approach calls for sector-wide capacity to anticipate change, question embedded assumptions, and design AI-enabled assessment systems that can adapt to evolving educational priorities.

Pathways Forward

Embracing Complexity and Iteration

Dr Thomas Corbin characterised AI in assessment as a “[wicked problem](#)” — too complex for simplistic or one-size-fits-all solutions — offering important guidance for how the sector should approach implementation. His framework emphasises granting educators:

- Permission to iterate – to refine and adapt approaches over time;
- Permission to be imperfect – to accept that early attempts will have flaws;
- Permission to be different – to adapt methods to their own context;
- Permission to fail – to learn from experiments that do not succeed.

Dr Corbin argued that educators need to be allowed and encouraged to experiment with AI tools in their unique contexts to find approaches that work for them, rather than waiting for perfect solutions or universal guidelines. This means creating institutional cultures that support trial and error, accept that initial attempts may not succeed, and recognise that successful AI integration will look different across different educational settings and contexts.

This approach to complexity acknowledges that uncertainty and ongoing adaptation are inherent features of AI integration rather than problems to be solved. However, participants noted that institutional cultures and accountability systems often resist iterative approaches, creating tension between the experimental mindset needed for thoughtful AI integration and the stability and predictability that educational systems typically require.

This ethos of experimentation was reflected in Dr Jing Xu and Dr Carla Pastorino-Campos’s multi-stage assessment model in which AI-supported feedback is embedded at multiple points to deepen learning and foster self-reflection. This example illustrates a wider call to design new assessment models from the ground up, rather than simply applying AI to existing formats.

Human-Centred Collaboration

Underpinning all such innovation, multiple speakers emphasised the need to keep humans at the centre of educational assessment, with AI serving as a supportive tool rather than a replacement for human judgment. This involves ensuring that human insight, emotion, and contextual understanding remain central while leveraging AI capabilities to enhance rather than replace human evaluation processes.

The principle that Isabel Fischer and Neha Gupta described as “unwavering human accountability” emerged as crucial, with recommendations that AI use in assessment should involve human oversight at all stages and never fully determine marks. They also highlighted the value of recognising human–AI collaboration as a valid form of assessed output, in which the interplay between human creativity, critical judgement, and AI-generated contributions is evaluated as a whole.

However, participants acknowledged that defining appropriate levels of human involvement raises complex questions about expertise, efficiency, and the changing nature of educational work itself. The goal of keeping humans central must be balanced against recognition that AI integration inevitably changes human roles and relationships within educational systems.

Together, these perspectives point to an approach where AI-enabled assessment is developed iteratively, grounded in human judgment, and adaptable to the diverse needs of educational contexts.

Institutional Policy Development

Workshop participants identified the need for clear institutional frameworks governing AI use in assessment. Recommendations included:

- **Transparency requirements** – Educators should be explicit about AI use and actively seek student feedback.
- **Data protection** – Ensure compliance and avoid uploading identifiable student work without consent.
- **Stakeholder involvement** – Include students, children, and parents in consent processes and give them meaningful control.

Alongside these principles, participants called for sector-wide trust and accountability measures, including agreed metrics to evaluate educational and equity impacts, and alignment with policies such as JCQ and SECA to ensure consistency and fairness. Participants stressed that evaluation should encompass not only technical accuracy but also alignment with educational values and long-term goals.

Participatory Approaches

Building on clear institutional frameworks, participants emphasised the importance of participatory approaches that empower educators and enable experimentation in low-stakes assessment environments. This includes creating opportunities for testing

new approaches and fostering feedback between stakeholders to make AI and assessment more collaborative and adaptive.

The emphasis on participation reflects recognition that AI integration affects multiple stakeholders with potentially different interests and perspectives. However, participatory approaches also raise questions about how to balance different voices, manage conflicting priorities, and ensure that participation leads to meaningful influence over AI assessment development and implementation rather than tokenistic consultation.

Interdisciplinary Collaboration

Given the complexity of AI assessment, participants stressed that successful integration requires sustained interdisciplinary collaboration and recognition that technology companies designing educational tools should consult educators for their expert knowledge to ensure these tools best serve educational needs.

This call for interdisciplinary collaboration acknowledges that AI assessment challenges cannot be addressed through purely technical or purely educational approaches. However, participants also noted that meaningful interdisciplinary collaboration requires significant investment in building understanding across different fields and may require new institutional structures and incentives to support sustained engagement between diverse stakeholders with different priorities and timescales.

Research Priorities and Practical Implications

Immediate Institutional Needs

The workshop identified several areas requiring immediate attention from educational institutions. Assessment design must move toward new assessment paradigms that combine human judgment with AI augmentation while preserving space for human emotion, creativity, and critical thinking. This requires institutions to develop clear policies and training for staff and students that go beyond technical skills to include critical evaluation of AI capabilities and limitations.

Professional development emerged as a critical priority, with participants emphasizing the need for educator preparation that addresses both practical AI tool

use and deeper questions about educational purpose and value. This includes helping educators navigate the changing nature of their roles and the tension between being a knowledge authority and a learning facilitator, as identified by Dr Fischer and Dr Gupta.

Student agency must be maintained through approaches that ensure AI tools enhance deep learning rather than merely measure surface performance. This requires careful attention to preserving learner agency while thoughtfully integrating AI capabilities into assessment processes.

Research and Development Priorities

Workshop discussions highlighted urgent research needs examining the role of AI in automated marking and the reliability of such tools. Current studies suggest significant error margins and inconsistencies that require further investigation, particularly regarding how AI assessment performs across different student populations and educational contexts.

Research is needed on developing dynamic assessment and reimagined assessment approaches that move beyond thinking about how to use AI on existing assessment models toward imagining new kinds of assessment. This includes exploring multi-stage assessment incorporating workshops, reflection, and final submission as integrated components rather than discrete elements.

Ethical impact assessment represents a critical research priority, examining how AI assessment tools affect learners, teachers, tutors, and families beyond just students themselves. This includes attention to societal and environmental implications and how AI in education might influence wider systems, including work, equity, and sustainability.

Long-term Development Needs

The workshop identified the need for measures and metrics that assess the impact of AI tools and frameworks for ensuring transparency in AI assessment systems. This includes developing approaches to aligning sector policies with an understanding of AI's role while maintaining flexibility for contextual adaptation.

Investment in evaluation represents a crucial priority - greater investment in evaluating AI's role and reliability before widespread implementation, with particular attention to contextual appropriateness and long-term learning outcomes. This evaluation must go beyond technical performance to examine educational effectiveness and alignment with institutional values and goals.

Conclusion

The workshop revealed a field grappling with fundamental questions about the nature and purpose of educational assessment in an age of artificial intelligence. While AI technologies offer unprecedented opportunities for personalised feedback, efficient evaluation, and innovative assessment approaches, they also pose significant risks to educational authenticity, equity, and human agency.

Perhaps most importantly, the discussions suggested that successful AI integration in assessment requires moving beyond questions of technological feasibility to deeper considerations of educational purpose and value. As one participant observed, the challenge is "how can we redefine assessment? Can technology help us be better educators, rather than replace a process that is no longer achieving the goal it was set out to achieve?".

The workshop demonstrated that there are no simple solutions to the challenges posed by AI in assessment. Instead, progress will require sustained dialogue between diverse stakeholders, continued research into the impacts of AI assessment tools, and careful attention to the values and purposes that should guide educational evaluation.

Alongside practical and regulatory discussions, the workshop also invited participants to reflect on the deeper assumptions that shape how assessment and ethics are understood in an age of AI. Contributors emphasised that the integration of AI is not solely a technical or procedural challenge but an opportunity to reconsider what counts as learning, judgment, and educational value within complex, interdependent systems. This broader perspective encourages a shift from controlling technological change to engaging reflexively with how human and technological systems co-evolve.

The conversations initiated at this workshop provide essential groundwork for navigating these complex challenges with both technological sophistication and educational wisdom. Future developments in this field will benefit from the kind of interdisciplinary collaboration and critical reflection exemplified by this gathering.

As the field moves forward, the key insight from this workshop may be that the question is not whether to integrate AI into assessment, but how to do so in ways that preserve and enhance what is most valuable about human learning and educational relationships while thoughtfully leveraging the capabilities that AI technologies can provide. The central challenge is not whether AI should be part of assessment, but

how to ensure it strengthens, rather than weakens, the human relationships, creativity, and integrity at the heart of education.

This report was commissioned by Cambridge University Press & Assessment to document and synthesise the key themes, insights, and recommendations emerging from the workshop on AI and Assessment organised by the AI & Education Community at the University of Cambridge Faculty of Education. It draws on contributions from the day's discussions and presentations, and aims to represent the wide spectrum of perspectives shared by participants. The analysis highlights both areas of consensus and ongoing debate in this rapidly evolving field, offering a resource for educators, policymakers, and developers engaging with the future of AI in assessment.

The workshop and report were made possible through the generous support of Cambridge University Press & Assessment, the Accelerate Programme for Scientific Discovery at the University of Cambridge, and Google Research.