TEACHERMATIC

AVALLAIN GROUP

SAFE

Framework

A Pedagogy-First Model for Ethical AI in Feedback and Assessment

teachermatic.com

We'd really value your feedback on the SAFE framework. Please take a moment to share your thoughts using the form.

If you'd like to explore this in more detail, we'll be hosting a Deep Dive session on the SAFE framework on Wednesday 3rd June. You can sign up here.

Rationale

AI-supported feedback is emerging rapidly across the education sector, yet there is still limited clarity around how it should be used responsibly within assessment. Feedback and marking sit at the heart of academic integrity, professional judgement and institutional reputation, so it is understandable that schools, colleges and universities are cautious about introducing AI into such a sensitive area.

For many educators this remains uncharted territory. Teachers are often navigating AI use independently and may be unsure what is acceptable, where the boundaries lie, or whether they may inadvertently be breaching institutional expectations. Many institutions also do not have the capacity to develop detailed frameworks quickly and are looking for practical guidance and reassurance.

The TeacherMatic SAFE Framework was developed to provide that clarity. It offers a structured, principled approach to ensure that AI-supported feedback remains secure, educator-led, fair and transparently governed.

As TeacherMatic introduces its AI Assisted Feedback Generators, including the Advanced Feedback Generator, which enables structured feedback on uploaded student work, it is important that clear ethical guardrails accompany this capability. SAFE provides a proactive response to the questions institutions are already asking about data protection, academic judgement, transparency and intellectual property.

Introduction

The TeacherMatic SAFE Framework provides a pedagogy-first, institution-ready approach to the responsible use of artificial intelligence in feedback and assessment- ensuring that innovation strengthens, rather than compromises, learning outcomes, academic integrity and educator expertise.

It has been developed in response to the rapid emergence of AI-supported feedback across education and the growing need for clear ethical and operational guardrails in this area. The framework reflects both pedagogical best practice and emerging expectations around responsible AI use in assessment.

SAFE helps schools, colleges and universities adopt AI tools in ways that protect academic integrity, safeguard data and preserve professional judgement while enabling innovation and reducing workload.

SAFE ensures that AI-supported feedback remains secure, educator-led, fair and transparently governed.

SAFE stands for:

S	Safeguarding data and privacy
A	Augmenting professional judgement
F	Fairness and inclusion
E	Ethical and transparent practice

Note on Regulatory Compliance

TeacherMatic's feedback and assessment tools are developed to align with the "Limited Risk" tier of the EU AI Act. This positioning is based on the platform's core function: it interacts with teachers to generate supportive content but strictly avoids autonomous, high-stakes academic decisions (such as final grading or admissions).

By focusing on educator-led workflows, TeacherMatic naturally meets the Act's transparency and safety requirements:

- Transparency: The platform ensures that users are clearly informed when they are interacting with AI-generated content.
- Contextual Guidance: Educators are provided with the necessary framework to interpret and refine AI outputs.
- Usage Records: Internal logging supports institutional accountability and traceability of AI interactions.
- Human Oversight: The system architecture prevents autonomous grading, ensuring that teachers retain final authority, thereby avoiding the technical and legal complexities of "High-Risk" AI systems.

S Safeguarding Data and Privacy

CORE PRINCIPLE

Student work is sensitive academic and personal data. It must be protected with the highest standards of security and care.

What this means in practice

TeacherMatic:

- Uses AI models that do not train on user-submitted data
- Operates within secure, controlled environments supported by internationally recognised security frameworks
- Does not claim ownership of student submissions or generated outputs, and does not reuse submitted content beyond the feedback generation task
- Processes uploaded content solely for the purpose of generating feedback on the submitted work
- Aligns with UK GDPR, EU GDPR and equivalent international data protection standards
- Supports secure bulk upload and inline annotation workflows
- Provides role-based access controls and appropriate data retention and deletion processes
- Supports transparency through platform-level logging available to authorised organisational administrators
- Encourages institutions to follow good practice when uploading work, such as avoiding unnecessary personal identifiers and adhering to institutional data protection policies

Institutional Good Practice

Institutions embedding SAFE are encouraged to:

- Ensure the use of AI-supported feedback is explicitly addressed within assessment, data protection and related institutional policies
- Define and communicate clear data retention and deletion procedures for student work processed through AI-supported workflows
- Provide staff with clear guidance on appropriate data handling practices when using AI tools, including what can and cannot be uploaded
- Encourage staff to minimise or anonymise unnecessary personal data where possible before submitting student work to AI-supported systems

- Maintain oversight of how student work is processed within AI-supported workflows, including visibility of data flows and points of access
- Establish a clear and consistent approach to student consent for AI-assisted feedback, for example through enrolment forms or learner agreements, and be prepared to address student questions about its use
- Verify that any AI tools used do not claim ownership of uploaded student work and provide clear assurances that submitted content is not used to train third-party AI models

A Augmenting Professional Judgement

CORE PRINCIPLE

AI should support educators, not replace them.

What this means in practice

TeacherMatic:

- Ensures educators remain in full control of the final feedback output
- Positions AI-generated feedback as draft, advisory and supportive, rather than definitive
- Requires educator review, refinement and approval before any feedback is shared with students
- Does not automate final grading or academic decision-making; any AI-generated grades or performance indicators are indicative and require educator review, validation and confirmation
- Supports hybrid feedback workflows that combine AI-generated content with educator expertise and contextual knowledge
- Supports transparency by providing visibility of AI tool usage to authorised organisational administrators, enabling oversight and accountability

Alignment with Assessment Expectations

TeacherMatic's approach is designed to support the core principle shared across awarding bodies and regulatory guidance: that AI must not replace human academic judgement in assessment. AI may support feedback generation, but final academic judgement and marking decisions remain with educators.

It is important to note, however, that each awarding body maintains its own specific rules and requirements regarding AI use in assessment, and these differ materially across organisations. Institutions must verify the guidance of each individual awarding body they work with before implementing AI-assisted assessment workflows, and should not assume uniform permission or prohibition across the sector.

Institutional Good Practice

Institutions embedding SAFE are encouraged to prioritise the following commitments. Verifiable human oversight is the most consistently referenced governance requirement from regulatory and awarding bodies, and should be treated as the foundation of any implementation:

- Put in place verifiable human review mechanisms that require educators to actively confirm they have reviewed and approved AI-generated feedback before it is released to students, rather than relying solely on professional goodwill
- Reinforce that final academic decisions, including grading and progression, remain the responsibility of qualified educators at all times
- Provide clear guidance on responsible AI-supported feedback and marking practices, including expectations around review, editing and contextualisation
- Integrate AI use into existing moderation and quality assurance processes, ensuring that AI-supported feedback is subject to the same scrutiny as traditional marking practices
- Avoid policies or practices that enable fully automated grading or decision-making without meaningful human oversight
- Encourage the use of transparent annotation approaches (for example, colour coding or equivalent conventions) to distinguish between AI-generated, human-authored and hybrid feedback, making the role of the educator visible within student work
- Invest in staff induction, training and ongoing support to ensure educators engage critically and professionally with AI-generated feedback, rather than passing it on without substantive review
- Recognise that staff confidence and attitudes towards AI-assisted assessment will vary, and support adoption through practical demonstration, peer sharing and evidence-informed approaches

F Fairness and Inclusion

CORE PRINCIPLE

AI-supported feedback should promote consistency, inclusivity and criteria alignment.

What this means in practice

TeacherMatic:

- Aligns feedback with rubrics, learning outcomes and assessment criteria to support consistent and criteria-referenced practice
- Encourages inclusive feedback practice by enabling educators to refine and adapt AI-generated feedback to reflect learner needs, including SEND and diverse learning profiles
- Applies multi-layered content filtering to reduce the risk of harmful, biased or inappropriate outputs
- Promotes consistency across markers, departments and cohorts by applying the same criteria regardless of marking sequence or volume
- Helps reduce common forms of human bias in assessment, including relational bias, fatigue-related bias and the unconscious influence of demographic factors
- Recognises that AI systems can also introduce bias, and therefore embeds safeguards through rubric alignment, structured inputs, educator review and ongoing monitoring
- Requires educator oversight to review, refine and contextualise feedback, ensuring that any unintended bias is identified and addressed before feedback is shared

Institutional Good Practice

Institutions embedding SAFE are encouraged to:

- Combine AI-supported feedback with the rubrics, marking guidance and essential assessment information provided by each relevant awarding body. There is no uniform approach across awarding bodies, and institutions must ensure that unit specifications, assessment criteria and guidance documents are consistently applied when using AI to support feedback processes
- Monitor consistency across programmes and departments to ensure that AI-supported feedback aligns with institutional standards and maintains parity of learner experience
- Ensure inclusive assessment policies remain central to AI use, with explicit consideration of SEND, EAL and diverse learner needs
- Provide ongoing staff development focused on bias awareness, inclusive feedback practice and the effective use of AI as an assistive tool rather than a decision-maker

- Communicate clearly with students about AI-supported feedback, focusing on the benefits most evidenced in sector pilots: improved consistency, more timely feedback and greater depth of commentary. Positioning AI primarily as a workload reduction tool for staff is unlikely to build student confidence, particularly in higher education contexts
- Recognise that student perceptions of AI in assessment vary by context. Resistance may be higher in higher education and in disciplines directly affected by AI in the labour market. Where concerns arise, direct and transparent communication from educators is more effective than relying solely on institutional messaging
- Engage students proactively in understanding how AI-supported feedback works, including its role in reducing marker fatigue bias, mitigating conscious and unconscious bias, and supporting greater consistency. Communication approaches should be adapted to reflect the needs, expectations and concerns of different learner groups

E Ethical and Transparent Practice

CORE PRINCIPLE

Students and staff should understand how AI is used in feedback and assessment.

What this means in practice

TeacherMatic:

- Encourages clear and proactive communication with students about the use of AI-supported feedback
- Supports institutions in developing transparency statements that align with internal policies and sector expectations
- Avoids hidden, opaque or autonomous marking practices, ensuring AI use remains visible and accountable
- Embeds documented risk assessment, governance and oversight processes into the use of AI-supported feedback
- Provides mechanisms for monitoring, reporting and escalation to support institutional accountability
- Supports institutions in communicating clearly and honestly with students about how and why AI is used in feedback, including addressing the reasonable expectation of fairness where students may be required to declare their own AI use
- Promotes transparency as a core principle for building student trust, recognising that openness about AI use is more effective than reassurance alone

Institutional Good Practice

Institutions embedding SAFE are encouraged to:

- Provide clear, student-facing statements explaining how and why AI-supported feedback is used
- Ensure students are explicitly informed when AI has contributed to feedback on their work, making its role visible at the point of feedback rather than only in policy documents
- Ensure all AI use aligns with academic integrity regulations and institutional policies
- Communicate clearly with staff about expectations, boundaries and appropriate use of AI within assessment workflows
- Review AI-related practices regularly in response to evolving sector guidance, regulatory updates and awarding body requirements
- Maintain a documented record of where and how AI has been used within assessment processes, including which assignments have involved AI-assisted feedback, to support quality assurance, external review and compliance
- Verify the specific guidance of each relevant awarding body before implementing AI-assisted assessment, recognising that requirements differ materially and should not be assumed to be uniform
- Exercise caution when using AI tools for plagiarism or academic integrity detection. Current tools are prone to high false positive rates and carry a risk of incorrect accusations, with serious implications for student welfare and institutional trust
- Ensure that professional judgement, supported by contextual evidence, remains central to all academic integrity decisions

Implementing SAFE: Institutional Governance

SAFE compliance is an institutional commitment, not an individual one. Evidence from ongoing sector pilots demonstrates that AI-assisted assessment cannot be implemented safely at the level of individual teachers working in isolation. It requires organisational readiness, clear governance and active senior leadership support.

Institutions planning to embed AI-assisted feedback are strongly encouraged to establish a cross-functional working group that brings together senior leaders, middle managers, teaching staff, business support staff and student representatives to work through implementation together. This group should address three foundational areas before rollout begins.

Permissions

Students should be informed that AI tools may be used to support feedback generation and should provide consent. For institutions with upcoming intake cycles, the most practical approach could be a tick box on enrolment forms or learner agreements. For institutions wishing to begin sooner, a clearly worded Microsoft Form or Google Form can serve as an interim consent mechanism. Transparency with students about the purpose and benefits of AI-assisted feedback is not just a regulatory consideration; it is also the most effective way to build acceptance.

Policies

Existing institutional policies may not yet fully address the use of AI in assessment and feedback. Institutions should review whether their current AI, assessment, teaching and learning, academic integrity or other relevant policies adequately cover AI-supported feedback practices, or whether additional guidance is needed.

Any institutional approach should reflect the hybrid marking model, the human-in-the-loop requirement, and the institution's position on which assignment types and awarding bodies are in scope. Policies should also be reviewed periodically as awarding body guidance evolves and sector practice develops.

Procedures

Staff need clear, practical guidance on how to use AI-assisted feedback tools responsibly within their institutional context. Evidence from sector pilots, including the Jisc AI Assessment and Feedback Pilot, consistently shows that formative assessment is the natural and lower-risk entry point for AI-supported feedback. Institutions are therefore strongly encouraged to establish confidence, consistency and robust processes in formative use before extending into high-stakes summative assessment.

Any progression into summative contexts should be approached with caution and supported by clearly defined policies, robust governance and alignment with awarding body requirements and qualification-specific regulations. This phased approach allows staff to build familiarity, test the consistency and reliability of AI-supported feedback against professional judgement, and refine workflows before applying these tools in contexts where outcomes directly contribute to final grades and academic decisions.

Procedures should clearly define which assignment types are in scope (for example, formative versus summative, and by level or awarding body), how student work containing personal data is handled, the minimum review standard required before feedback is released, and how AI use is recorded for quality assurance purposes. Where institutions provide broad access to AI-assisted feedback tools, ensuring all staff have access to guidance, training and support is as critical as providing access to the tools themselves.

Positioning Statement

The SAFE Framework underpins TeacherMatic's approach to AI-supported feedback and assessment.

Unlike general-purpose AI tools that can generate unstructured or autonomous outputs, TeacherMatic's feedback workflows are designed to operate within clear pedagogical and governance boundaries.

AI-generated feedback is positioned as draft and advisory, never as an automated academic decision.

The SAFE Framework ensures that AI-supported feedback remains:

- Secure in how student work is processed
- Educator-led in how judgement is applied

- Fair and inclusive in how feedback is structured
- Transparent and ethically governed in how AI is used within assessment workflows

SAFE supports a hybrid model of marking where professional expertise remains central, and technology reduces administrative burden without compromising academic integrity or learner trust.

Ongoing Development

The SAFE Framework will continue to evolve as sector practice develops, informed by institutional feedback, regulatory guidance and ongoing research into responsible AI use in education.

TeacherMatic's ongoing participation in the JISC AI assessment pilot is central to this development process. Evidence from this pilot continues to shape how the platform is built and how guidance for institutions is framed. As comparative feedback, BTEC assessment, portfolio-based grading and VLE integration capabilities are developed, the governance and ethical considerations that accompany them will be reflected in future versions of this framework.

Institutions using SAFE-aligned tools are encouraged to share their own experiences and governance approaches with TeacherMatic, contributing to a growing body of shared practice across the sector.

Disclaimer

The SAFE Framework is intended as practical guidance to support institutions in adopting AI-assisted feedback responsibly. It does not constitute legal advice. Institutions with specific regulatory or compliance questions are encouraged to seek independent professional advice appropriate to their context.

Each institution remains responsible for its own policies, assessment practices and compliance obligations. TeacherMatic makes no representation that following this framework ensures compliance with any specific legal requirement, awarding body rule or regulatory standard. Awarding body guidance on AI use in assessment differs across organisations and continues to develop; institutions should consult the relevant guidance directly and not rely solely on this document for compliance decisions.

This framework reflects the state of sector knowledge and regulatory guidance at the time of publication. It will be reviewed and updated as AI governance in education evolves. Institutions should treat this as a living document and monitor updates from TeacherMatic, their awarding bodies and relevant sector bodies as guidance in this area develops.