Risk Assessment Matrix for Schools Implementing AI

Introduction

The following risk assessment matrix is intended to help schools identify, evaluate, and mitigate risks associated with implementing Artificial Intelligence (AI) in educational processes.

The matrix considers potential risks across various domains, including data protection, ethical considerations, and operational integrity. There is a particular focus on safeguarding and wellbeing issues, highlighting potential risks to student welfare and offers strategies to mitigate these risks effectively. Schools should amend the content of the matrix as necessary and consider the risk profile that is relevant to their own circumstances.

Risk Assessment Matrix

Risk Area	Risk Description	Likelihood (Low/Med/High)	Impact (Low/Med/High)	Risk Level (Low/Med/High)	Mitigation Measures
Data Protection and Privacy Breaches	Unauthorised access to sensitive data or personal information, leading to safeguarding concerns and commercial risk.	High	Med	Med	Implement strong encryption, regular audits, and GDPR-compliant data management policies and conduct regular privacy audits.
Cyberbullying	Increased potential for bullying through AI-mediated communication tools.	Med	Med	Low	Monitor AI communication tools, implement clear reporting mechanisms, and provide student support.
Over-reliance on AI	Over-reliance on AI tools reducing interpersonal interactions among students. Reduction in teacher autonomy and critical	High	Med	Low	Encourage collaborative learning activities and balance AI use with social engagement. Define clear

	decision-making by overusing AI tools.				boundaries for AI use and regularly review its impact on pedagogy.
Emotional Manipulation	AI systems unintentionally affecting student mental health through curated content.	Med	Low	Low	Monitor AI-generated content, involve mental health professionals, and promote media literacy.
Inappropriate Content or Conduct	AI exposing learners to harmful or unsuitable materials / behaviour	Med	Low	Low	Conduct rigorous testing of AI tools, apply effective filtering and monitoring and ensure human oversight.
Mental Health Impacts	Overuse of AI tools causing stress, anxiety, or dependency in learners.	Med	Low	Low	Monitor usage patterns, provide mental health resources, and set expectations on use of AI systems.
Bias and Discrimination	AI systems propagating biases that impact student wellbeing or inclusion. AI models producing discriminatory or biased outcomes.	Med	Low	Low	Regularly audit AI algorithms for bias and provide inclusive media literacy education and training.
Misuse of AI	Learners using AI tools for harmful, unethical or illegal	High	High	Med	Educate learners on responsible and appropriate AI use and

	purposes (e.g. nudification).				establish clear usage policies.
Misinformation	Creation or spread of harmful or misleading AI-generated content.	High	Med	Med	Educate staff and learners to verify AI outputs and establish clear policies for verifying content authenticity.
Digital Divide	Inequitable access to AI tools among learners from diverse demographic groups.	Med	Med	Med	Provide equitable access to AI resources and ensure alternative solutions are available.
AI Ethics Awareness	Lack of awareness among staff and learners about ethical implications of AI.	Med	Med	Med	Provide training and education on AI ethics and its responsible usage. Establish an 'Ethics in AI' group.
Data Accuracy	AI systems generating inaccurate or misleading recommendations.	High	Med	Med	Regularly validate AI outputs and involve human oversight in decision-making.
Legal Compliance	Non-compliance with laws regarding AI usage and learner data.	High	Med	Low	Understand legal requirements. Conduct legal reviews and consult experts on AI-related regulations.
Cyber-Security	Increased use of AI tools in cyberattacks	High	High	High	Strengthen cybersecurity protocols and

	targeting school systems and data.				educate staff and learners on safe online practices.

Likelihood and Impact Definitions

- Likelihood: The likelihood that the identified risk will occur.
    - Low: Unlikely to occur under normal circumstances.
    - Medium: Possible occurrence based on past trends or vulnerabilities.
    - High: Likely to occur without intervention.
- Impact: The severity of impact should the risk materialise.
    - Low: Minimal disruption with limited consequences.
    - Medium: Moderate disruption affecting key processes.
    - High: Significant disruption with severe consequences.

Action Plan

Schools may wish to create an action plan based on the findings of the risk assessment matrix. Your plan should consider a before and after approach to demonstrate the intended impact of your proposed actions. These actions might include (amend as relevant).

1. Appoint a safeguarding lead to oversee AI implementation and its impact on student welfare.(Head Teacher).
2. Deliver regular training for staff on online safety, AI ethics, and responsible use.
3. Audit the curriculum to identify where these AI related issues might be incorporated into existing or new curricular programmes.
4. Establish transparent reporting mechanisms for any safeguarding or wellbeing concerns linked to AI.
5. Collaborate with parents and carers to raise awareness about AI risks and best practices.

6. Conduct annual reviews of AI tools and their impact on student wellbeing, updating the risk matrix as needed.
