Responsible Artificial Intelligence Assessment: Supplier Guide

Guide for suppliers participating in the Responsible AI (RAI) Assessment process for the Australian School Sector.

Release:	2025.1
Date of this version:	9 Oct 2025
Author:	RAI Assessment Team
Document Version Number:	V1.4 Released
Location:	Latest official version available www.st4s.edu.au

Important information including disclaimer:

This guide is provided:

- for information purposes only and does not constitute advice
- on the basis that suppliers are responsible for assessing the relevance and accuracy of its content.

Education Services Australia Limited through its business unit the National Schools Interoperability Program (NSIP) has compiled this guide in good faith and has endeavoured to ensure that all material is accurate and does not breach any entity's rights at the time of its inclusion. However, the material may contain unintentional errors and is provided 'as is.'

Participation in the Responsible AI (RAI) Assessment process is voluntary. An entity which chooses to participate in the RAI Assessment process acknowledges and agrees that:

- the product must undergo a successful assessment through the ST4S process, before it is eligible for RAI assessment
- the RAI Assessment process and results depend entirely on the answers provided by an entity and the point of time at which such answers are provided
- the RAI Assessment of an entity may result in a recommendation to participating education authorities that such entity's product not be used until issues around responsible use of AI are remedied
- NSIP is conducting the RAI Assessment on behalf of participating jurisdictions for the purpose of ensuring consistency in responsible AI assessments.

To the extent lawful, NSIP:

- excludes all warranties in respect of the guide and the RAI Assessment process
- is not liable for any loss or damage (direct or indirect) resulting from the use of the guide or participation in or the results of, the RAI Assessment process
- will not be liable for any incidental, special or consequential damages of any nature arising from the use of or inability to use the guide or participation in the RAI Assessment process.

Links provided to other websites are provided for the user's convenience and do not constitute endorsement of those sites. ESA is not responsible for material contained in any website that is linked to from this guide.

If you use the links provided in this guide to access a third party's website, you acknowledge and agree that the terms of use, including licence terms, set out on the third party's website apply to the use which may be made of the materials on that third party's website. If this guide contains links to your website and you have any objection to such link, or if you have any questions regarding use of material available on or through this website, please contact us.

Contents

1. Introduction

Thank you for your interest in the Responsible AI (RAI) Assessment initiative. This guide is an important resource and contains detailed information on how services are assessed, what outcomes mean and the criteria that must be met to obtain a compliant outcome.

1.1 Purpose

This supplier guide provides guidance and information regarding:

- the assessment process
- the questions that make up the questionnaire
- the minimum and indicative responses to the questions and links to relevant industry standards
- the clarification process
- the assessment results and how they will be shared with participating member organisations.

1.2 Key terminology

Term	Definition
AIS	Australian Independent Schools, including their bodies and representatives across States/Territories within Australia.
Education authority	Government or non-government ST4S member organisations responsible for ICT guidance to schools and other compulsory sector education providers in a given jurisdiction e.g. Government education departments, independent schools, catholic dioceses
ESA	Education Services Australia Limited (www.esa.edu.au)
ST4S	Safer Technologies for Schools
RAI	Responsible Artificial Intelligence
RAI WG	Responsible Artificial Intelligence Working Group
NEDAG	National Education Digital Advisory Group (formerly known as the NSIP Steering Group)
NSIP	National Schools Interoperability Program (www.nsip.edu.au), a business unit of ESA

1.3 Background

- The school sector has long focused on ensuring that products and services used in schools are assessed for compliance with cybersecurity and privacy concerns. In Australia, this evaluation has been consolidated at a national level in the Safer Technologies for Schools (ST4S) program, run out of ESA on behalf of school authorities nationally and the New Zealand Ministry of Education. ST4S sets a national benchmark for cybersecurity and privacy with EdTech products and works with suppliers to help them achieve that benchmark in the process of assessment.
- The recent explosion in Generative AI functionality has focused attention on the risks posed by the new technology in all domains of use, including education. The ST4S assessment has already been updated to include an AI module focusing on privacy, security and safety, for products and services using AI.
- The unique nature of AI has meant that there is particular attention on ensuring the ethical and socially responsible use of AI, as new AI-equipped software is procured and used. This domain has become known as **Responsible AI**, and it encompasses notions of the following:
  - *Transparency:* users should be able to understand, to the extent practical, how an AI system goes about making decisions and generating output, and what the limitations of its functionality are.
  - *Fairness:* users should have confidence that the AI system minimises bias in its outputs and decisions, and accounts for the diversity and different expectations of its user base in its configuration.
  - *Accountability:* it should be possible for users to understand which parties are responsible for the AI system, and to challenge them about inaccurate or unfair AI outcomes.

- The Australian schools sector formulated the *Australian Framework for Generative Artificial Intelligence in Schools* to address the challenges posed by AI, in November 2023. Responsible AI is a core concern of the framework: of the 6 principles it nominates, *3, 4,* and *5* are the Responsible AI principles of *Transparency, Fairness* and *Accountability*, while *2 Human and Social Wellbeing* provides the overall goal of Responsible AI, to safeguard users and their rights as they use AI.
- The Australian Government commissioned ESA in 2024 to explore the feasibility of Responsible AI evaluation program for EdTech, covering those four principles of the Framework. The study determined that a Responsible AI evaluation for Australian edtech was indeed feasible, and that the domain of Responsible AI in general was rapidly maturing to address the new challenges.
- Throughout 2024, policy frameworks and legislation were continually emerging to support Responsible AI. Significant advances included the EU AI Act (August 2024), and domestically the Australian AI Voluntary Safety Standard (September 2024).
- Australian Responsible AI policy assumes that regulators in different domains will take responsibility for establishing compliance, including the proposed mandatory counterpart to the Voluntary Safety Standard. As there is no such regulator in the school sector, school authorities are responsible for establishing Responsible AI compliance.
- In 2025, ESA was funded to produce a standard for Responsible AI evaluation of Australian EdTech, under the oversight of school authorities and with feedback from them. This work concluded in June 2025, and the first version of the standard is published as this document.
- In the second half of 2025, ESA will be conducting a pilot program of Responsible AI evaluation, to test the evaluation standard and to consolidate the evaluation process. The standard will be updated in response to the pilot program findings. The evaluation process will be used, as with ST4S, to set a common benchmark for compliance.
- Subject matter experts from agencies and the non-government school sectors, meeting as the Responsible AI Working Group (RAI WG), will be steering work on Responsible AI standards and evaluation going forward.

1.4 Benefits of a coordinated approach

- Most schools and school system authorities have established local risk assessment teams or are planning to do so. School authorities can draw on the national evaluation to perform their own evaluations and procurement decisions.
- The anticipated benefits of a coordinated assessment approach are as follows:
    - Agreed standards and practices for ensuring transparency, fairness, accountability, and human and social wellbeing in AI technologies in schools are clearly communicated to all school communities and product suppliers.
    - School selection of online services is guided by reliable information about AI transparency, fairness, accountability, and human and social wellbeing.
    - Reduced cost, effort and time for education authorities in assessing and on-boarding AI-enhanced online services for schools.
    - Increased transparency and trust regarding the responsible use of AI in schools.
    - Reduced cost and time for suppliers to demonstrate compliance with national Responsible AI expectations.
    - Incentive for suppliers to comply with national and local Responsible AI expectations.

1.5 High level assessment process & prioritisation

The Responsible AI (RAI) Assessment process consists of 2 steps:

1. Eligible services are prioritised by the RAI WG. Each month a limited number of services can commence a RAI Assessment, contingent on ESA resourcing.
2. Prioritised services undergo a full RAI Assessment in collaboration with the RAI Assessment team.

To be eligible, a service must have already undergone ST4S assessment, including assessment against the AI module, successfully, with the assessment currently valid. Importantly, the RAI WG, in consultation with the NEDAG, is responsible for determining the assessment priority of supplier products and services.

1.6 Start-ups and Small Businesses – Quick Notes

Our approach to the Responsible AI (RAI) Assessment is to work with you to help you achieve a compliant outcome. Upon submitting an assessment, we review and engage with you to understand your service and to provide time to remediate any items as required.

Drawing on the experience of ST4S, we anticipate that organisations will make changes, update policies and improve testing to meet the criteria, and we will be supporting small businesses with non-compliant items achieve a compliant outcome.

Our recommendation is to first review the questions and then get in contact with us to understand what the next best steps are and how you can meet compliance. We find after a discussion with small businesses / start-ups that they are more comfortable with the assessment process and how they can achieve a compliant outcome. Many small businesses / start-ups will find they may meet criteria already; they are on the right path or that there are small changes that when implemented can meet the criteria.

Please also ensure you read the 'Support' section of this guide for more information on how we support organisations throughout the assessment process.

1.7 Large Organisations and International Companies – Quick Notes

The RAI Assessment is an initiative to establish a single assessment framework for Responsible AI, to reduce company effort in completing multiple assessments for school authorities and education departments across Australia. If your service is in use by schools in Australia and relies on AI, it is recommended you complete an RAI assessment.

An RAI assessment is similar to undertaking an ISO27001, SOC2 audit or other review. RAI is not simply a questionnaire to complete for a customer. Documentation and evidence must be lodged. It is important to ensure your CIO is aware your company is interested in participating in the RAI so they can support you in providing you with the documentation and evidence you require.

The RAI Evaluation is formulated for Australian small-to-medium EdTech vendors. Large-scale vendors will require adjustment to the evaluation, because of their different scale and confidentiality details, and differences around procurement processes.

AI model providers are not in scope of RAI evaluation because of the significant challenges of scale and detail, and confidentiality that such evaluation would pose. The access of general AI models directly by schools and students poses a range of risks beyond what an evaluation program like this can address, as their use is not contextualised by classroom use, nor are the expected education-specific guardrails in place.

2. Assessment process

A summary of the assessment process is described below. Further information on each of the assessment stages are broken down into sub-sections.

2.1 Overview

1. **Nomination:** Products and services submitted for assessment are verified by the RAI Team and then prioritised by the RAI WG once a month. We set a maximum quota each month.
    a. Services not suitable for an assessment (e.g. noted on the RAI Assessment Exclusion List in the appendix of this guide or lacking a current valid ST4S AI assessment) are not included.
    b. RAI WG members nominate services based on school demand and other factors. If a service is not nominated, it will roll over to the next month until a new framework or substantial change has been made.
    c. If a service is not nominated but the quota is not filled, the RAI Team may provide an opportunity to other assessments without any nominations. This is done in order of the submission date.
2. **Full Assessment:** Once prioritised, suppliers are invited to participate in the Full Assessment which requires the submission of the full questionnaire and documentation. An assessment officer is assigned to review your submission, trial the service and complete the assessment with you.
3. **Reassessment:** Required every 2 years to remain current, unless major changes are introduced to your service or a RAI WG member or ESA requests a reassessment.

At any time, a supplier may withdraw from the process, or the RAI Team may discontinue an assessment. Further information on withdrawing or discontinuing is described in this guide.

2.2 Full Assessment process

Suppliers that are prioritised for assessment may be invited by the Responsible AI (RAI) Assessment Team to proceed and undertake a full assessment.

Suppliers that are invited to the assessment process will be provided with a link to the RAI Assessment questionnaire and be asked the set of RAI Assessment control queries as represented in this guide. Suppliers are also required to provide supporting evidence and additional information throughout the assessment process.

2.2.1 Assessment steps

Stage 1: Submission

Suppliers must first lodge the Full Assessment using the online questionnaire provided by the specified due date. Extensions may be requested by contacting the Responsible AI (RAI) Assessment Team in writing and are subject to review.

Stage 2: Precheck and Evidence Review

An initial review of the service is conducted which includes reviewing the submissions against the minimum criteria, reviewing the website, technical checks and reviewing provided documentation and evidence.

Stage 3: Assessment and Report Creation

The RAI Assessment Team will continue with its detailed review of the service and begin drafting the RAI Assessment school report for the service. As part of this process, suppliers are provided with an opportunity to clarify responses, and additional information may be sought from the RAI Assessment Team to complete the review process.

Suppliers may also seek to undertake additional remediation, should they wish to reduce any additional risks and improve the outcome of their assessment. Additional time to remediate is subject to review by the RAI Assessment Team and may extend the time to complete an assessment.

Stage 4: Finalisation

Provided the service achieves a compliant outcome, the RAI Assessment Team will provide a copy of the draft report to the supplier to review and approve to be finalised. Handling feedback of the report is described in 2.2.5 of this guide. If a service does not achieve a compliant outcome, the RAI Assessment Team may produce another outcome

(e.g. a non-compliant outcome) or choose to discontinue an assessment. Further information on outcomes is further described in this guide.

If a Supplier chooses not to accept a report, or we are unable to finalise a report we may discontinue the assessment.

Remediation

There are cases where an item may not be compliant in the Full Assessment either due to genuine error, we assess an outcome differently to what the supplier may have expected, or we do not accept the relevant documentation etc. In this instance, we provide a period to remediate (generally 3 months maximum) subject to review by the RAI WG. If remediation items are substantial, or we do not believe an organisation may be able to remediate in time, we may discontinue an assessment and ask the organisation to return later when ready.

Progress Updates

Throughout the assessment process, the RAI Assessment Team records an overall assessment status to the RAI WG, NEDAG and other education authorities as described in this guide. This status may detail the service is 'pending submission', 'awaiting remediation' or another status determined by the RAI Assessment Team. The RAI Assessment Team may also discuss compliance matters with the RAI WG or issues relating to the assessment and how it is progressing.

Communication of updates, findings, outcomes etc to schools are a local matter and are decided by the education authority, RAI WG member and/or NEDAG as relevant.

Assessment Approach

The RAI Assessment process requires open communication and transparency. Technologies and topics of cybersecurity, privacy and online safety can be complex and how each supplier implements technologies and features to meet RAI Assessment criteria can differ.

Where a supplier has a different approach to meet a control (e.g. using a different technical service or solution), the supplier can provide additional information during the clarification process to support the assessment.

Where suppliers have missed a question or not provided sufficient detail, the assessment team may follow up with the submitting supplier to ensure a fair and accurate response is gathered and assessed.

Where a response cannot be obtained from a supplier or the RAI Assessment Team is satisfied there is a differing level of compliance (or non-compliance) an alternative finding may be made by the RAI Assessment Team to facilitate the completion of the assessment.

2.2.2 Release of findings to suppliers

Suppliers will receive a draft of the school level report which is generated based on the responses provided to the supplier questionnaire. Suppliers may also receive a spreadsheet containing questions on which the assessment team is seeking further clarification. Suppliers are asked to respond to the clarifications within the timelines as directed. Supplier responses to the clarifications and a commitment to rectify any risks resulting in a 'non-compliant' outcome may alter the school report.

Following the successful conclusion of clarifications, suppliers should expect to receive a final school level report. A copy of the final school level report will be provided to the supplier's nominated contact. The exception to this release timeline is where a supplier has received a non-compliant outcome, the assessment was discontinued, or another outcome was determined which did not result in a report being produced.

2.2.3 Findings outcomes

The assessment of a product or service results in one of the following outcomes:

- Non-Compliant
- Compliant, High Risk
- Compliant, Medium Risk
- Compliant, Low Risk

The overall assessment outcome is the highest risk level remaining after all available treatments have been applied. A 'Non-compliant' assessment outcome is assigned when a mandatory minimum standard is not met. The assessment outcome appears on the front page of the school-level report.

Other assessment outcomes include: discontinued, non-participating, or another outcome as determined by the Responsible AI (RAI) Assessment Team. These outcomes may or may not result in the creation of a report by the RAI Assessment Team.

2.2.4 What do findings outcomes mean?

In typical school settings, there is always some risk in using a product/service. Some products/services may receive a Medium or High rating simply because of the types of functionality that they offer (e.g. assessment or streaming of students). The overall assessment outcome highlights to schools that in using the product/service there are treatments that need to be applied (e.g., configuration, reviewing of logs). Assigning a Medium or High outcome to a product/service is intended to draw school users' attention to the fact that treatments need to be reviewed and implemented when using the particular product/service. Typically, besides removing the particular functionality in question, there is little or nothing a supplier can do to reduce the overall assessment outcome to Low.

Products/services which have fundamental compliance gaps or have failed to meet the minimum criteria may be determined as being 'Non-Compliant', the Responsible AI (RAI) Assessment Team may discontinue the assessment or determine another outcome.

Each education authority, RAI WG member and/or NEDAG may determine what advice or other information they provide to schools.

Acceptance of outcomes

Suppliers are advised that education jurisdictions, other RAI WG members and/or NEDAG may at their discretion choose to accept or reject an outcome, apply additional requirements and/or conduct their own assessment activities. Common examples include where local regulations require an education authority to conduct a local Responsible AI assessment, policies require additional information to be communicated to schools, or a RAI WG member may choose to only accept a certain risk level or services with data hosting that is within their jurisdiction.

Suppliers are encouraged to contact the education authority for further information on their requirements and policies.

2.2.5 Alternative findings and challenging findings

Responsible AI (RAI) Assessment criteria are designed to be worded generically, and we acknowledge that each service and supplier may implement criteria differently. The RAI Assessment Team reviews how criteria may apply in different circumstances, such as where you may be performing a local impact assessment, or who you may restrict access for to the product or service.

The RAI Assessment Team will maintain an internal register of previous decisions to ensure we apply findings and decisions fairly. Exemplars of decisions will be established as the program gets underway.

Alternative findings are not guaranteed and are made in consultation with members of the RAI WG.

Our recommendation to suppliers is throughout the assessment process is to provide as much technical information as possible to help us in our review. This allows your assessment officer to properly review and search through decisions and come to a fair conclusion.

Challenging Findings

As part of the development of the final school level reports, suppliers will have been provided a draft copy of the school level report and clarification questions. The final school level report should not be a surprise to the supplier as the outcomes are dictated by the guidance and criteria in this guide. If a supplier considers a school level report is not accurate, that supplier may lodge a request to have their report re-reviewed. To request a re-review, suppliers need to provide relevant details to the RAI Assessment Team in writing with a sufficient amount of information including technical information to assist the team in its review. Where applicable, we recommend providing technical information on the products you use on the cloud service provider and configuration settings you may have applied to help our technical team members review.

Any challenge to a finding or outcome is first reviewed by the RAI Assessment Team and presented to the RAI WG if necessary. Should a matter be referred to the RAI WG, a decision by a RAI WG member, the RAI WG and/or NEDAG is final.

It's important to note that the RAI Assessment framework references major standards, frameworks and criteria as set by government agencies and other members of the RAI WG including Catholic and independent school bodies. Simply disagreeing with a criterion or claiming you should not be required to meet the criteria without genuine explanation or information is insufficient for a challenge to occur.

In almost all cases, the RAI Assessment Team is able to resolve matters with suppliers without the need to engage with the RAI WG. Consultation with the RAI WG may still be necessary in high-risk use cases.

2.2.6 Re-assessment

Assessments are valid for 2 years unless withdrawn beforehand.

Subject to resourcing and prioritisation, suppliers may be invited to be re-assessed based on several factors, including time since original assessment, updates to the Responsible AI (RAI) Assessment standards, updates to the supplier product/service and/or occurrence of a breach or security incident.

Suppliers may also lodge a request for a reassessment.

A reassessment will generally always be necessary in the following circumstances:

- the organisation that owns the service changes for example by sale, license transfer or similar
- features or functionality changes (either new or a modification and/or removal) that results in the new categories being relevant to the assessment
- changes in the AI value chain used, including the foundation model provider, the other AI infrastructure providers, and the terms and conditions for the AI service
- a change in hosting location or infrastructure (e.g. you migrate regions of infrastructure from within the same cloud service provider or you migrate one or more components from one cloud service provider to another)
- changes to the relevant Responsible AI policy that warrant a new assessment such as changes in liability or in testing regimen
- adding or removing features/functionality related to artificial intelligence or changes in how your organisation or any other organisation may use data on the service (including where de-identified) for training or development of AI models.

2.2.7 Changing the school level report

The final RAI Assessment report can only be altered by the RAI Assessment Team. Any request to change a school report must be made in writing to the RAI Assessment Team.

Updates are not always a simple change to the report and outcome. For example, where a product/service is acquired by a new company, there are changes to hosting locations, features/functionality or the types of data collected, the RAI Assessment Team may determine that a reassessment is necessary.

2.2.8 Discontinuing an assessment

The Responsible AI (RAI) Assessment Team may at any time discontinue an assessment and do so at their sole discretion. Where an assessment is discontinued, the RAI Assessment Team will no longer conduct any assessment activities and formally close the assessment, noting the reason for discontinuing the assessment to the RAI WG and/or NEDAG.

Assessments may be discontinued for a variety of reasons. Some examples may include:

- The service being assessed is not suitable for an assessment or the RAI Assessment Team postpones an assessment to a newer framework. We may communicate an outcome of 'postponed' to the RAI WG along with a note on when we expect an assessment to begin.
- An excessive number of non-compliant outcomes are identified in the Full Assessment.
- A supplier is not being transparent, omitting important information or is misleading the RAI Assessment Team throughout the assessment process (notwithstanding genuine errors in submissions).
- A supplier refuses to provide evidence to a satisfactory standard.
- As part of the review, the RAI Assessment Team will send out clarifications to the supplier to seek more information or to recommend product or process changes to remove non-compliant risks. If no response is received or no action is taken on the required changes and the assessment period has gone over 6 months from the time the first pass clarifications were sent, the RAI Assessment Team may discontinue the full assessment.
- The RAI Assessment Team receives a request from a RAI WG and/or NEDAG to discontinue an assessment.

If an assessment is discontinued:

- the Supplier must note their product was discontinued along with the reason when approached by a school, education authority, RAI WG member and/or NEDAG, and
- a minimum waiting period of 3 months may apply to the supplier before it is eligible for consideration for a new assessment.

Further information on communication requirements regarding discontinued outcomes is described in this guide.

3. Sharing and use of full assessment reports, findings and outcomes

3.1 Distributing reports, findings and outcomes

The Responsible AI (RAI) Assessment Team provides assessment findings (including raw results and school level reports) to the NEDAG (typically the Chief Information Officer at each education authority) and the RAI WG (Chief Information Officer and/or nominated security and privacy representatives). The RAI Assessment Team may distribute findings and outcomes to schools directly; however, this is currently limited to Australian Independent Schools (AIS) and is subject to change. The process and timelines by which each education authority distributes findings is a local matter and is not managed by the RAI Assessment Team. In some education authorities, findings will be distributed to schools within days of release from the RAI Assessment Team, in others, schools need to make requests directly to their local education authority.

3.2 Sharing information

One of the goals of the Responsible AI (RAI) Assessment is for a national assessment framework across Australia, with collaboration from education authorities and other members of the RAI WG and NEDAG. A standardised assessment process reduces the requirement for suppliers in providing multiple transparency, fairness, accountability and human rights impact questionnaires.

When responding to the questionnaire and participating in the RAI Assessment process, suppliers should be aware that information provided throughout the assessment, status updates, assessment results, evidence, reports and any other information (e.g. including your contact details) may be shared with the RAI WG, NEDAG and other parties (e.g. Trusted Parties) as nominated by the RAI WG and/or NEDAG. This may include central department or sectoral staff and their schools and/or regional offices.

In addition, subject to approval by the NEDAG and/or the RAI WG, results may be distributed to other parties without prior notice or consultation with the relevant supplier. Examples include where a RAI WG is engaging with the Privacy Commissioner in their jurisdiction, a RAI WG is requesting advice or engaging with legal counsel or another government agency or department has requested to review the report.

3.3 Sharing of findings with Suppliers

Suppliers will be provided with a copy of their school level report. These guidelines are intended to provide a sufficient level of detail so that suppliers can effectively perform a self-assessment against the assessment criteria. However, where there are critical risks the Responsible AI (RAI) Assessment Team may contact suppliers directly to communicate any issues identified.

The RAI Assessment Team will not provide suppliers with the findings of other suppliers who have submitted responses.

Suppliers must not provide results, findings, and outcomes themselves to schools. Instead, suppliers should refer schools to their education jurisdiction, education authority or other contact as specified in this guide.

3.4 Supplier use of the findings internally

One of the goals of the Responsible AI (RAI) Assessment process is to encourage suppliers to improve transparency, fairness, accountability, and human and social wellbeing approaches in the design, build, testing, deployment, maintenance, configuration and end-user training regarding their product/service. Suppliers can continue to improve their products/services over time and are encouraged to continue to reference the RAI Assessment standards (as documented in this guide) as it is updated over time.

3.5 Guidance regarding supplier use of assessment outcomes

Suppliers receive copies of the final assessment reports with the following caveats and conditions:

1. Responsible AI (RAI) Assessment reports will be marked as ‘Not for commercial purposes’.
2. Suppliers must not provide the RAI Assessment report or any copies or extracts of it to anyone outside the supplier organisation (e.g. schools or school communities).
3. Suppliers may notify existing and prospective customers that they have participated in the RAI Assessment process and meet the minimum required RAI Assessment standards (against a specific version of the RAI Assessment standards) for the specific version of their product/service.
4. Suppliers must acknowledge and communicate with customers that an RAI Assessment outcome does not necessarily mean that the supplier is compliant with local State/Territory/Country or Non-Government sector requirements.
5. Suppliers must direct enquiries from schools regarding the provision of detailed reports to the relevant education authority (Government schools to the relevant State/Territory Department or Ministry of Education, Australian Catholic schools to their local State or Diocese office and Australian Independent schools to their State/Territory association) as listed on the final report.

6. Suppliers must not edit or modify their final or draft school-level reports in any way.
7. Suppliers must not claim that an RAI Assessment applies to other products, services, or modules offered by the supplier, or different versions of the product, service or module.
8. Suppliers must not publish, advertise or promote their specific assessment outcome (low/medium/high), or use or extract any part or portion of their RAI Assessment report. Communications to existing and prospective customers must be limited to the particular service version that has been assessed and the result and must indicate that this version aligns to a particular RAI Assessment standard version (compliance assessments are not enduring for all time).
9. Suppliers must not claim or imply that the RAI Assessment is an endorsement, recommendation, or approval of the product/service or a guarantee that the service is fit for purpose.
10. Suppliers must not publish in whole or in part the RAI Assessment results for another supplier's service.
11. Suppliers must notify the RAI Assessment Team if they come into possession of some or all of another supplier's RAI Assessment report or results.
12. If a supplier does not comply with the above usage conditions, the RAI Assessment Team may rescind/withdraw/modify that supplier's assessment outcome.
13. In its sole discretion, the RAI Assessment Team may rescind/withdraw/modify any assessment outcome at any time.

These guidelines will be updated from time to time. Please refer to the RAI website for the latest usage conditions.

Suppliers should direct Australian government school queries to the relevant educational jurisdiction. Contacts will be provided when the RAI working group is established.

Suppliers should direct Australian non-government school queries to the relevant authority listed below:

- Catholic and Independent Schools
  - Catholic Education – Contact the relevant local jurisdiction i.e. diocese, CEnet or commission.
  - Independent schools – Contact the local AIS operating in your State/Territory.

3.5.1 Requirements for non-compliant, non-participating suppliers or discontinued assessments

1. If approached by current or potential customers regarding the Responsible AI (RAI) Assessment process, suppliers must state that their outcome was non-compliant, non-participating, discontinued or another status determined by the RAI Assessment Team as relevant, note the reason and direct schools to the relevant education authority as listed above.
2. If you have published documents, articles or other information in relation to your participation (or prior participation) within the RAI Assessment program you must retract that information where it is reasonably practical to do so.

3.5.2 Disclaimer in relation to Supplier Guide

1. This Supplier Guide is provided for your information only and you are responsible for ensuring that its contents are current, complete and accurate before using it.
2. Whilst ESA has endeavoured to ensure that the Supplier Guide is accurate and up-to-date, the Supplier Guide is provided to you on an 'as is' basis and you use it at your own risk.
3. To the extent lawful, NSIP:
   - excludes all warranties in respect of the Supplier Guide; and
   - is not liable for any loss or damage however caused resulting from the use or inability to use the Supplier Guide or caused to any property as a result of the use of the Supplier Guide.

4.0 Support

Our goal of the Responsible AI (RAI) Assessment is to help suppliers obtain a compliant outcome at the end of the assessment process. Many organisations (including start-ups and small businesses) that choose to participate in the RAI Assessment are able to work with us on making improvements, clarifying items and overall achieve a compliant outcome at the end of the assessment.

The RAI Assessment Team supports suppliers to achieve a compliant outcome by:

1. Responding to an enquiry you have on the criteria. Just reach out on the contact form on our website.
2. Providing time to remediate (subject to approval) for items you have genuinely missed or where we may have a differing assessment outcome.
3. Clarifying responses with you and allowing additional information to be submitted throughout the process, particularly where you may implement a differing approach to meet a control.
4. Providing you with support materials and general guidance on how to meet compliance.
5. Meeting with you (phone, video chat etc) to discuss the criteria or your concerns.

Additional Support

If you require additional support, such as extended discussions with development teams on technical matters, additional reviews on documentation etc., we can provide this subject to availability of our team. We generally prioritise organisations who need our support the most, such as small businesses or start-ups that may not have access to a Responsible AI expert, not-for-profits or charities, organisations delivering services which process highly sensitive information (e.g. mental health data), and services which address high priority initiatives as determined by the Australian curriculum or government.

5.0 Terms and Conditions

A Supplier electing to participate in the Responsible AI (RAI) Assessment process or any other related RAI Assessment activity shall agree to and abide by the applicable terms and conditions.

5.1 Important information, disclaimers and conditions in relation to RAI activities

If you do not agree to any of the points below, you must not complete a Responsible AI (RAI) Assessment questionnaire or participate in the RAI Assessment process:

- Responses provided may be used to inform any contractual arrangements entered into by government departments, non-government sectoral authorities or individual schools.
- Please note that the RAI Assessment school-level reports resulting from participation in the RAI Assessment do not constitute an endorsement, approval or recommendation regarding the use of the product/service to which they apply, nor do they constitute advice regarding the quality or licensing of, or the decision to purchase or use a particular product or service. RAI Assessment outcomes are provided with no guarantee or warranty.
- The RAI Assessment process encompasses the entire solution, including services, applications, and other components that form the overall solution. If an application processes information but does not store it or communicate the information back to the organisation, it may still be within scope.
- You will be required to provide evidence at a later date to support your responses. Evidence is closely inspected and reviewed to ensure organisations are meeting the criteria within the RAI Assessment framework.
- This questionnaire is
  specifically designed to elicit detail of the product, service or solution to inform potential end-users of the product, to detail any potential risks and mitigations and to arrive at an overall risk rating.
- Participating stakeholders outside of the RAI Assessment Team may seek further detail from suppliers to address local Responsible AI needs at a future date.

- Engagement in the assessment process and /or completion of the questionnaire does not guarantee or indicate any intention to proceed with purchasing, licensing or procurement activities.
- Participation in any stage of the RAI Assessment process or otherwise in relation to any matter concerning the RAI Assessment process, will be at each supplier's sole risk, cost and expense. NSIP will not be responsible for any costs or expenses incurred by a supplier in preparing its response to the questionnaire or otherwise taking part in the RAI Assessment process or taking any action related to the RAI Assessment process.
- The RAI Assessment process is not an offer capable of acceptance by any person or entity or as creating any form of contractual, quasi contractual or any other rights based on legal or equitable grounds. Therefore, engagement in the RAI Assessment process and /or completion of the questionnaire does not constitute an agreement, arrangement or understanding between a supplier and NSIP, the assessment service or any stakeholders in the RAI Assessment.
- NSIP is not liable to any supplier or any other entity on the basis of any legal or equitable grounds including negligence or otherwise as a consequence of any matter or thing relating or incidental to a supplier's participation in the RAI Assessment process.
- The questions below directly relate to the requirements contained within the various and relevant safety and equity legislation, various Government Responsible AI frameworks and best practices in the industry across key principles of transparency, fairness and accountability with regard to AI. Supplier responses will assist in the assessment, mitigation and monitoring of the risks associated with their product/service.
- Any individual completing the RAI Assessment, or any other related RAI Assessment activities on behalf of the Supplier must:
    - be duly authorised by the Supplier's organisation to do so, and
    - hold express written permission from the Chief Executive Officer (CEO), Chief Information Officer (CIO), or another senior officer of the Supplier with comparable authority for making executive decisions and directing the company's overall strategy.
- The RAI Assessment framework sets a high standard that Suppliers may elect to meet. Suppliers who begin the RAI Assessment process but subsequently determine that they are unable or unwilling to remediate in order to meet the criteria, or who fundamentally disagree with the criteria or the high standard expected under the framework, shall promptly withdraw from the RAI Assessment process. Suppliers wishing to meet the criteria are encouraged to discuss any concerns with the RAI Assessment Team and may collaborate on a remediation plan to achieve compliance.
- Suppliers shall provide all necessary information to support the RAI Assessment process, which may include documentation, evidence, and/or access to a trial or demonstration account.
- An assessment officer from the RAI Assessment Team will be assigned to work with each Supplier throughout the RAI Assessment process. The Supplier shall comply with all instructions and requests made by the assigned assessment officer in relation to completing the RAI Assessment. Should the assessment officer require referral of any enquiry or matter to another department or individual within the Supplier's organisation, the Supplier shall promptly effect such referral.
- A Supplier Code of Conduct ('the Code') is included as an appendix to this guide. Compliance with and adherence to the Code is a condition of undertaking any assessments or related activities with NSIP concerning the RAI Assessment initiative. Suppliers found to be in breach of the Code may be removed from the assessment process, have their assessment discontinued, or be subject to any other action NSIP deems appropriate. Determinations of whether a Supplier is in breach of the Code shall be made at the sole discretion of the RAI Assessment Team. The Code may be updated periodically, and the most recent version will be published in this guide. Suppliers are responsible for remaining up to date with any changes to the Code. The RAI Assessment Team may note a Supplier's engagement and conduct during the RAI Assessment process, or related activities, to the RAI WG.
- It is a condition of participation that any correspondence provided by a Supplier to NSIP and/or the RAI Assessment Team may be referred or made available to the RAI WG and/or any member thereof for review.

This includes correspondence from third-party organisations or individuals acting on behalf of the Supplier, or under the Supplier's instruction.

- Any challenges, disagreements, or disputes relating to the RAI Assessment criteria or the assessment process shall initially be addressed by the RAI Assessment Team. If further escalation is required, matters shall be referred to the RAI WG and/or NEDAG, in accordance with the escalation procedures outlined in the Supplier Code of Conduct. Suppliers shall at all times comply with this escalation process.

5.2 Completing the questionnaire

- Suppliers will receive, via email, a link to complete a questionnaire for a specific nominated service/product. A survey access pin will be sent via text message to the nominated contact.
- All questions are mandatory, and suppliers will not be able to navigate between pages without first completing the questions on the page displayed.
- If at any time suppliers are not sure which product, module or component is the subject of the response, please contact the assessment team.
- If the supplier's service offers a 'for school use' and a 'for home use' version, please complete the questionnaire based on the 'for school use' version.
- If suppliers need to provide any attachments which are directly relevant to the question being asked (please do not provide advertising materials or lengthy documents) prefix the file name with the relevant question ID e.g. INT3-API Product XYZ).
- Suppliers will be able to partially complete the questionnaire and return later to complete it.
- Suppliers may choose to print a copy of their responses to the questionnaire prior to submitting.
- Suppliers can contact the assessment team if they have any questions or comments. We are here to help. Please review '4.0 Support' of this guide for more information.

5.3 Accuracy of responses to the questionnaire

In submitting the questionnaire, suppliers must:

- confirm all information provided in response to the questionnaire is true, correct, accurate, up-to-date, and not misleading in any way
- acknowledge that:
    - the Responsible AI (RAI) Assessment Team will rely on the information provided in response to the questionnaire to assess the service's compliance and provide guidance to stakeholders
    - incomplete, inaccurate, out of date or misleading information may result in the relevant service receiving an inaccurate or misleading report; and
    - agree to provide further information or evidence to support the questionnaire responses if requested.

5.4 Timeline

Timelines to submit the self-assessment questionnaire are included in the assessment information email sent to suppliers.

The Responsible AI (RAI) Assessment is a detailed audit and review process. Time to complete an assessment varies depending on the complexity of the service, the types of data being provided, priorities set by the RAI WG and other factors. Suppliers should allow at least 3 months from submission of the Full Assessment questionnaire.

5.5 Other requirements

Throughout the assessment process, the Responsible AI (RAI) Assessment Team may request a supplier to provide additional support to assist in reviewing the service. This may include requesting a demo or trial account for the service, requesting access to support materials (e.g. user guides and manuals) or a service's terms and conditions and Responsible AI policy if these are not publicly available etc.

5.6 Supplier conduct

A supplier code of conduct applies to all suppliers. Agreeing to the conditions within this guide and the supplier code of conduct is a requirement of the assessment process. View Appendix G for the supplier code of conduct.

5.7 Supplier communication requirements

Suppliers should be aware that the Responsible AI (RAI) Assessment Team communicates regularly with RAI WG members including staff at a State/Territory Department of Education, etc. We also monitor supplier communications including updates you may post to your website, changes to your privacy policy / terms of use etc.

What you must not do:

- Publish that you are engaged in the RAI Assessment process or suggest/imply you are conforming to RAI Assessment standards and criteria when you have not successfully completed a recent full assessment for the service which remains valid.
- Use RAI Assessment, NSIP or ESA brand, logos, colour schemes and other materials in any communications or content related to the RAI Assessment or program
- Produce your own RAI Assessment-like badge, images or content such as a 'RAI Assessment Certified' badge or 'RAI Assessment' followed by a tick symbol etc.
- Engage in misleading or deceptive conduct when communicating with a school, a government agency (e.g. a State/Territory Department of Education body) or another RAI WG member as to the status of your assessment.

You are welcome to ask us to review any publications, media statements or drafts before you publish. We are happy to review material to ensure that any information you publish aligns with our expectations.

5.8 Compliance with the Terms and Conditions

The Responsible AI (RAI) Assessment Team reserves the right to remove a Supplier from the RAI Assessment process, discontinue an assessment, or revise any outcome or decision in the event that any condition set forth above (or elsewhere in this guide or within section 5 of this guide) is breached, or if a Supplier fails to comply with the RAI Assessment Supplier Code of Conduct.

6. Assessment criteria

Updates to the Responsible AI (RAI) Assessment Criteria, Response Options & Minimum Standards

Given the rapid change to the underlying standards which the RAI Assessment criteria draw on, the RAI Assessment Team estimates that the RAI Assessment criteria (as represented in this document) will be updated every six months, with release likely occurring in January/February and June/July each year. In addition, the RAI Assessment Team may make interim updates to RAI Assessment criteria and/or this guide. This may include changes such as correcting the wording of a question and response options or adding/removing/adjusting criteria etc. Suppliers are advised to frequently check for the latest version of the RAI Assessment Supplier Guide as published on the RAI Assessment website.

Control Identifiers: The identifiers of controls in RAI are structured as follows:

- A letter and number representing the Framework principle that the control addresses: H2 for human rights and social wellbeing, T3 for transparency, F4 for fairness, A5 for accountability. (G for general and C for company do not correspond to the Framework principles, and do not have a principle number.)
- A second number representing the Framework guiding statement that the control addresses; e.g. H21 for controls addressing guiding statement 1 of human rights and social wellbeing, Wellbeing.
- An ordinal number for the control within the guiding statement; e.g. under Wellbeing, the controls are H211, H212, H213, H214

In this initial stage of the assessment, the controls are still in flux, with several controls reordered or removed, so there are gaps in numbering, and some controls appear out of sequence, with respect to how they were initially presented. (So H236 has been moved after H211, as it directly relates to it; T302 has been removed, so T301 is followed by T303.) It is more important for maintenance purposes for the identifiers to be stable, than for them to appear in order and with no gaps; so the identifiers will not be renumbered going forward.

Mandatory Controls: Controls that are mandatory to fill out are indicated with an “M” flag.

Related Standards: In the following listing of controls, we reference two standards with related controls, which users should be aware of when filling out responses:

- The Voluntary AI Safety Standard (VAISS, https://www.industry.gov.au/publications/voluntary-ai-safety-standard) is the Australian federal government guideline on Responsible AI in Australia, with voluntary guardrails applicable to all organisations in the AI supply chain. As a cross-industry framework, this is a reference for compliance with Responsible AI principles in general.
- The ST4S assessment, and specifically its AI module, has some overlap with the RAI Assessment, and is a prerequisite for suppliers participating in the RAI Assessment. Suppliers are encouraged to reuse text and responses they have already provided for ST4S AI, when addressing the RAI Assessment. We differentiate in the following between:
    - Close matches: the related ST4S controls can be used to provide all information required for this control, and suppliers can work out what to answer for RAI based on those prior responses.

        - We do not currently have a mechanism for copying responses from ST4S to RAI automatically, and some close matches still require interpretation and consolidation on the part of the supplier.
    - Priority matches (indicated with #): matches that involve responses that could lead to an evaluation returning non-compliant or high risk. These are higher priority than any other close matches.
    - Partial matches (indicated in italics): the related ST4S controls do not provide all information required for this control, but may contain some useful information for the RAI response.

Minimum Standards: The following questions include response options which, if selected, demonstrate a failure to meet the minimum standards set by the RAI Assessment. (These are typically worded as 'No action taken'.) Failure to meet these standards automatically produces a non-compliant outcome for your product.

- **General: G8, G9**
- **Wellbeing: H211, H212, H213, H214**
- **Diversity of Perspectives: H221**
- **Human Rights: H233, H234, H235**
- **Architecture: T301**
- **Information and Support: T311, T312**
- **Disclosure: T324, T327**
- **Accessibility and Inclusivity: F411, F412, F413**
- **Non-Discrimination: F431, F432, F433, F434**
- **Accountability General: A501, A502, A503**
- **Human Responsibility: A512, A513, A514, A515**
- **Reliability: A522, A528**
- **Monitoring: A531, A532, A535, A537**
- **Contestability: A543, A545**

6.1 Criteria – Company & Product Detail

Note that these questions are in common with ST4S, and answers will be copied across from the preceding ST4S evaluation of the product where practical.

Q	M	Question	Response Options	Notes	ST4S
C0	M	For which countries are you submitting an ST4S survey response?	A. Australia and New Zealand B. Australia only C. New Zealand only		C0
C1	M	Vendor name		Informational	C1

C2A	M	Vendor ABN		Informational (AU submissions)	C2A
C2B	M	Vendor NZBN		Informational (NZ submissions)	C2B
C3A	M	Registered Australian address of vendor		Informational (AU submissions)	C3A
C3B	M	Registered New Zealand address of vendor		Informational (NZ submissions)	C3B
C4A	M	Country in which the company is registered for Australian customers		Informational (AU submissions)	C4A
C4B	M	Country in which the company is registered for New Zealand customers		Informational (NZ submissions)	C4B
C5A	M	For Australian customers: Preferred vendor contact name Preferred vendor contact email Preferred vendor contact phone number			C5A
P1	M	Name of service			P1
P2A	M	Version of service If no published version number, use date of version.			P2A
P2B	M	Is the service free or paid?	A. Free B. Paid		P2B
P2C	M	For paid services, URL of pricing page			P2C
P2D	M	Are you the product or service's original developer, a re-seller or 'other'?	A. Original developer B. Reseller C. Other (please specify)		P2D
P2E	M	Do you warrant that you have the legal authority to submit this product or service for a Responsible AI assessment?	A. Yes B. No		P2E
P2F	M	Does your organisation outsource any development, maintenance or operation activities to another organisation?	A. Yes B. No		P2F
P3A	M	URL of service for Australian customers			P3A
P3B	M	URL of service for New Zealand customers			P3B
P4A	M	URL of Terms of Service/use for Australian customers			P4A
P4B	M	URL of Terms of Service/use for New Zealand customers			P4B

6.2 Criteria – General

Q	M	Question	Response options	VAISS	ST4S
G1	M	What is your product or service intended to be used for overall (not focusing on the AI component of the product)? (Use cases)		1.2.2, 1.2.7, 9.2.1, 9.2.2	P5
G4	M	Which of these intended user groups interact with the AI component of your product or service directly? (e.g. giving it prompts or instructions, or creating images?)	• Students • Parents • Teachers • Admins	9.2.2, 10.3.1	AI_G4
G3	M	Which parties who are not intended users of your product or service are affected by its use? (e.g. a recommender system is used by a teacher, but impacts students)	• Students • Parents • Teachers • Administrators	2.1.4, 10.3	*AI_G1#, AI_G4*
G5	M	Please select from the following the AI features or functions used in your product or service:	• Automated grading and feedback systems • AI-driven curriculum design and optimisation • Anonymised Analytics and Reporting • Analytics and Reporting with Personal Information • Educational chat bots • Non-educational chat bots • Generation of learning resources and content • Language processing for plagiarism detection • AI-assisted research and data analysis tools • Text translation • Image, video or audio generation • Content summarisation and reformulation • Recommender system	9.2.1, 9.2.4	AI_G2
G6	M	What benefits have you identified for using AI in your product or service?		1.2.2, 9.2.1	AI_G3

G7	M	What outcomes does your product or service produce using AI? These can include content, actions, and decisions.	• Images • Videos • Text e.g. blog posts, social media posts, newsletters, presentations • Audio or music • Quizzes • Emails • Infographics • Code, scripts or other web development content • Actions (e.g. send email, assign students into streams) • Decisions (e.g. fail student, run program)	1.2.2, 9.2.1, 9.2.2	AI_PF2, AI_PF3
G8	M	Do you have a risk management system applicable across the AI lifecycle? A risk management system draws on: documentation (policies, procedures, processes), risk registers and assessment reports, executive meeting minutes, resource allocation.	A. Yes B. No	2.1.1, 2.1.2, 2.1.3, 2.1.4, 2.1.5, 2.1.6, 2.1.7, 2.2.1, 2.2.2, 2.2.3, 2.2.4, 2.2.5	AI_T1, AI_GO1
G9	M	Have you completed any risk impact assessments across the AI lifecycle for this product or service relating to Responsible AI?	A. Yes B. No	2.2.1, 2.2.2, 9.2.3	*AI_T1, AI_GO1, AI_G8*

6.3 Criteria – Wellbeing

Q	M	Question	Response options	VAISS	ST4S
H211	M	Indicate the safeguards and mitigation strategies you may have implemented to ensure that your product or service cannot be used to generate or facilitate prohibited content and activities: • Creation of or exposure to obscene, degrading, and abusive content, including child sexual abuse material, pornography, and deep fakes • Creation of or exposure to dangerous, violent, or hateful content	• Robust pre-trained filters and guardrails: Integrate advanced content moderation layers that automatically detect and block obscene, degrading, or abusive content, including deep fakes and other harmful materials • Model fine-tuning and RLHF: Optimised to reduce generation of prohibited content. • Real-time content moderation: Employ AI-based classifiers that screen all generated outputs for	2.2.3, 3.2.1, 4.1.1, 4.2.1, 5.1.2, 8.1.2, 9.2.2	*AI_G1, AI_T4, AI_SF2, AI_SF3, AI_SF4, AI_SF5*

		• Access to the promotion of dangerous materials, including weapons, toxic substances, drugs. • Misuse of surveillance and biometric data	dangerous, violent, or hateful content, preventing their dissemination • Human-in-the-loop oversight: Combine automated systems with human review, where flagged content is evaluated by moderators to ensure accuracy and fairness in content blocking • Strict data curation practices: Use carefully vetted training datasets that exclude prohibited content, thereby reducing the risk of generating harmful outputs • User reporting mechanisms: Offer intuitive feedback channels that allow users to report inappropriate content, triggering prompt review and remediation • Access controls and authentication: Limit advanced functionalities to verified users (e.g. educators or administrators) to reduce the risk of misuse by unauthorised parties • Continuous monitoring and logging: Implement real-time monitoring and audit trails to detect and respond to any potential breaches or misuse of the system • Exclusion of high-risk features: Avoid integrating functionalities related to surveillance and biometrics unless they meet stringent privacy and ethical standards • Regular policy and system updates: Conduct periodic reviews of moderation strategies and update guardrails in response to emerging threats, regulatory changes, and user feedback • External ethics and peer reviews: Collaborate with external ethics boards and independent reviewers to ensure that all mitigation measures remain effective and aligned with best practices • No action taken • Not applicable		

H236		Which safeguards or risk mitigation measures are provided by third parties to address risks listed in H211?	A. Infrastructure provider B. Foundation model provider C. Open source third party D. Commercial third party E. Not applicable	2.1.5, 2.1.6, 8.1.1, 9.2.2	*AI_T11, AI_EV5*
H212		Indicate the measures you have in place to address or mitigate the risks associated with using AI to make automated decisions about students' educational futures (e.g. class streaming, admissions, assessment) and staff recruitment. Please include any relevant guardrails, documentation, alerts, or human oversight processes.	• Model fine-tuning and RLHF: Optimised to reduce risks around automated decision-making. • Human-in-the-loop oversight: All automated decisions are reviewed by qualified human experts before final implementation to ensure fairness and accountability • Robust guardrails and decision thresholds: Predefined parameters and triggers are in place to pause or flag decisions for further review if they fall outside acceptable criteria • Comprehensive documentation: Detailed records of the AI decision-making process—including data sources, methodologies, and limitations—are maintained and made accessible to stakeholders • Regular bias and fairness audits: Scheduled internal and external audits are conducted to detect, measure, and address any biases in the decision-making processes • Real-time monitoring and alerts: Continuous monitoring of AI outputs with automated alerts for anomalies or potential issues that require immediate human intervention • Stakeholder feedback mechanisms: Feedback channels are available for students, educators, and staff to report concerns or irregularities in AI-driven decisions, informing ongoing improvements • Continuous model updates and retraining: The AI models are periodically retrained with updated, diverse data to ensure that decision accuracy and fairness are maintained over time	2.2.3, 4.1.1, 5.1.1, 6.1.1, 6.1.2, 6.1.3, 6.1.4, 7.1.1, 8.1.2	*AI_G1*

			• Compliance with legal and ethical standards: All processes adhere to relevant educational, data protection, and anti-discrimination regulations, ensuring ethical application of AI • No action taken • Not applicable		
H213		Indicate the measures you have in place to address potential risks arising from the human–AI interface – such as anthropomorphising AI, emotional entanglement, or over-reliance – and any other emerging wellbeing concerns you have proactively identified? Please include any guardrails, documentation, alerts, or human oversight processes.	• Model fine-tuning and RLHF: Optimised to reduce risk of entanglement. • Human-in-the-loop oversight: Ensure that interactions flagged for potential emotional over-reliance or misinterpretation are reviewed by human experts who can provide appropriate support or intervention • Emotional safety guidelines and design: Incorporate built-in prompts, explicit disclaimers, visual cues, and educational messages that reinforce healthy user engagement and clarify the AI's role as a tool rather than a human-like companion • Real-time monitoring and alerts: Set up monitoring systems to detect signs of over-reliance or emotional distress, with automated alerts that trigger human review and follow-up • Comprehensive documentation: Maintain detailed records of AI interactions and user feedback to monitor for patterns indicative of emerging wellbeing risks, informing iterative improvements • Regular wellbeing assessments: Collaborate with educational and psychological experts to periodically review the impact of the human–AI interface and adjust risk mitigation strategies accordingly • User education and training: Provide resources and guidance on the intended use of AI, helping users understand its limitations and fostering	2.2.2, 2.2.3, 7.1.1, 8.1.2	*AI_G1, AI_T4, AI_SF2, AI_SF3*

			balanced, responsible engagement • Feedback channels for wellbeing concerns: Offer accessible mechanisms for users to report any discomfort or concerns related to the AI interface, ensuring prompt action and continuous improvement • Compliance with Human and Child Rights standards • No action taken • Not applicable		
H214		Did you follow human-centred design principles when designing your product or service? (e.g. IEEE 2089 Standard for age-appropriate digital services) Has that included ensuring that AI outputs are presented to children at an appropriate level?	• Adherence to formal design framework: We designed our product or service in compliance with a formal framework such as the IEEE 2089 Standard, involving educators, psychologists, and child development experts, and ensuring all digital services are age-appropriate • Adherence to informal design framework: We designed our product or service in compliance with an informal framework to guide our product development • Tailored AI content: Our AI outputs are customised to match different age groups – adjusting language complexity, content depth, and visual presentation accordingly • User testing with children: We conducted extensive user testing and focus groups with children to gather direct feedback on interface usability and content appropriateness • Teacher controls and settings: We offer features that allow teachers to set preferences and manage the types of content accessible to their children • Ongoing review and iteration: We have established a continuous review process with child development specialists to update and refine AI outputs as children's needs evolve • Transparent communication: We provide clear,	4.2.1, 7.1.1, 9.2.2	*AI_T1.2, AI_SF1, SFP1*

			age-appropriate explanations about how AI is used in our product or service, fostering trust among young users and their guardians • No action taken • Not applicable		

6.4 Criteria – Diversity of Perspectives

Q	M	Question	Response option	VAISS	ST4S
H221	M	Indicate how your AI system has been designed and trained to incorporate cultural, geographical and diverse perspectives.	• Diverse training data: We source and curate training data from a broad spectrum of cultural, geographical and demographic groups to ensure balanced representation. This can include multilingual datasets and localised content • Collaboration with domain experts: We work with cultural experts, sociologists, and community representatives to validate our data sources and adjust the AI system's outputs • Bias audits and evaluations: Regular audits are conducted to assess AI system performance across different cultural and geographical contexts, and to identify any disparities • Transparent documentation: We maintain detailed records of our data sourcing, training methodologies and bias mitigation strategies, making them available for review • User feedback integration: We actively gather and incorporate feedback from a diverse user base to improve representation and address any emerging gaps • No action taken • Not applicable	3.2.2, 7.1.1, 10.4.1	AI_T9, *SC4, AI_T1.4*
H222	M	Indicate how your AI system can be customised to align with the specific needs and values of different schools and school communities. Please detail the approaches and strategies you employ.	• Customisation modules: We offer built-in modules that allow schools to adjust interfaces, content delivery, and recommendation algorithms to suit their specific educational needs at the level of the school • Policy-based customisation: We offer built-in modules that allow schools to customise UI design, content delivery, and recommendation algorithms, alignment with local policies, regulatory standards, or values. • User-configurable settings: We offer built-in modules that	7.1.1, 9.2.2, 10.3.4, 10.4.4	

			allow schools to adjust UI design, content delivery, and recommendation algorithms to suit their specific educational needs at the level of individual users • Localisation and language options: Our system supports multiple languages, dialects, and region-specific content, ensuring cultural and contextual relevance • Flexible Integration and API support: Our AI system can seamlessly integrate with existing school systems, enabling custom data inputs and workflows tailored to each community. This can include adaptive learning by the AI system. • Collaboration with educators: We work proactively with school leaders and teachers to customise features and content based on local curricula, pedagogical methods, and community values • Feedback loops for continuous improvement: We maintain regular feedback channels with schools to iteratively refine and expand customisation features based on evolving needs • Pilot programmes and case studies: We offer pilot projects that allow schools to trial and adapt our AI system before full-scale implementation, ensuring a smooth and tailored integration • Not applicable		
H223	M	Indicate whether your organisation has a formal policy committing to diversity, inclusion and fairness, and how these commitments are integrated into your organisational goals for AI development and deployment.	• Formal policy adoption: We have a publicly available policy that explicitly commits to diversity, inclusion and fairness across all our operations • Performance metrics: Our organisational goals for AI development include measurable targets for diversity, inclusion and fairness, ensuring these principles guide all strategic decisions • Training and awareness: We provide comprehensive training on diversity and ethical AI practices for all staff, ensuring that our policy translates into day-to-day operations • Inclusive recruitment and leadership: Our hiring and	1.2.1, 1.2.2, 8.1.1, 10.2.1, 10.2.2, 10.2.3, 10.2.4	AI_T1.1

			promotion processes prioritise diversity, with dedicated roles and committees to oversee the integration of these values in AI projects • Regular auditing and reporting: We conduct periodic internal and external audits to measure our progress on diversity and inclusion, and we publicly report these outcomes • Stakeholder engagement: We engage with diverse community representatives, external experts and partner organisations to continuously refine our policies and practices in line with emerging best practices • We are in the process of adopting a formal Diversity, Inclusion and Fairness policy. • We do not have a formal Diversity, Inclusion and Fairness policy.		

6.5 Criteria – Human Rights

Q	M	Question	Response options	VAISS	ST4S
H231	M	Do you establish informed user consent for all use of AI? How is relevant information communicated to users?	• Clear direct consent workflows: A straightforward opt-in process that explains AI features in plain language and requires explicit acknowledgement from students (when they are able to directly), teachers, or parents • Consent via guardians or administrators: Permission from parents, legal guardians, or authorised school administrators, even for students who may be able to provide consent directly • Age-appropriate communication: Use child-friendly language, visuals, or interactive tutorials to help younger users understand the role of AI • Contextual disclosures: Display short, on-screen messages or pop-ups explaining how AI will be used at the moment it is activated • Policy and documentation: Provide a publicly accessible AI usage policy [and/or documentation] that outlines data handling, privacy measures, and consent	6.1.1–6.1.4, 6.2.4, 6.2.5, 9.2.2	AI_PR4, AI_PR9, *AI_D1, AI_PA2, AI_PR2,*

			protocols • Regular updates: Prompt users or their guardians to renew or review consent when significant changes are made to AI features • We do not establish consent for all uses of our product or service.		
H232	M	What options do you provide for users to opt-out of AI usage, and how do you communicate or facilitate this process?	• User interface settings: A dedicated toggle or menu where users can switch off AI features at any time • Administrative control: School administrators or parents can disable AI modules for specific classes or individual students • Helpdesk or direct support: Users can contact support to request deactivation of AI-related services • Granular feature opt-out: Ability to opt out of particular AI-driven functionalities (e.g. recommendation engines) while retaining others • Clear documentation: Step-by-step instructions on how to disable AI features, available in both digital and printed form • Users cannot opt-out of AI usage for this product or service.	5.1.2, 6.1.3	*AI_G1B, AI_A1, AI_T7*
H233		Indicate how you have involved students, teachers, and parents in the design process to address their needs, concerns, and expectations regarding AI use in schools, especially around diversity, inclusion, and fairness.	• Focus groups and workshops: Conducted sessions with students, teachers, and parents to gather direct feedback and co-design features • Surveys and interviews: Deployed structured surveys or interviews to identify concerns about fairness, bias, or usability • Pilot programmes: Launched small-scale trials in diverse school settings to gather real-world insights and refine the AI system • Advisory panels: Formed committees including teachers, parents, and possibly older students, to guide ongoing development • Iterative feedback cycles: Maintained an open channel for continuous feedback, ensuring new concerns are addressed promptly	7.1.1	*AI_T1.4*

			• No action taken • Not applicable		
H234	M	What measures do you implement to ensure that your AI system respects human dignity – particularly in how it engages with users respectfully, avoids manipulative or deceitful practices, and does not exploit or coerce users?	• Ethical design framework: Incorporate guidelines that prohibit manipulative features or deceptive interfaces • Consultation during design: Ensure representative users have been consulted on design decisions around their dignity as users • Use respectful language models: Use an AI model selected to minimise offensive or biased content, and apply filters for harmful language • Train and fine-tune respectful language models: Train and/or fine-tune an AI model to minimise offensive or biased content, and apply filters for harmful language • Transparent system prompts: Ensure the AI system's role and limitations are transparent, so users are not misled into thinking it is human or infallible • Consent and control mechanisms: Provide users with the ability to manage data usage and opt out of certain interactions • Regular ethical audits: Conduct periodic evaluations to detect potential manipulative or exploitative behaviours in AI outputs, e.g. Human Rights Impact Assessment • No action taken • Not applicable	2.2.2, 4.1.1, 6.1.1, 6.1.2, 6.1.3, 6.1.4, 10.4.2	AI_T4, AI_I1, AI_I2, AI_EV3, AI_SF4
H235	M	How do you ensure that your AI system respects worker and learner rights, specifically avoiding overwhelming users with excessive or unengaging tasks, information overload, or undue labour demands?	• Workflow and learning design: Design effort to anticipate how users can most effectively interact with the product or service • Task simplification features: Automate repetitive tasks (e.g. grading) without adding unnecessary complexity to teachers' or students' workloads • User-friendly interfaces: Design intuitive dashboards and workflows that reduce cognitive load and minimise confusion • Time-management controls: Integrate features that limit notifications or tasks outside normal working or study hours	2.2.2, 7.1.1, 10.4.4	

			• Workload monitoring: Track usage metrics to identify when staff or learners are being overburdened and adjust accordingly • Consultation with educators: Involve relevant stakeholders to ensure the AI system meets professional guidelines and labour standards • No action taken • Not applicable		

6.6 Criteria – Architecture

Q	M	Question	Response options	VAISS	ST4S
T301	M	Can you provide documentation outlining your system architecture, including AI components and non-AI components, a 'bill of materials' (statement of shared responsibility), and system requirements? This documentation should include what country the AI components are hosted in, and a visual representation where helpful.	A. Yes B. No	3.1.1, 9.1.1	AI_G6, AI_G7, AI_H1, AI_T1.2, AI_EV5

6.7 Criteria – General Transparency

Q	M	Question	Response options	VAISS	ST4S
T303	M	Please describe your product or service's AI deployment model. For example, indicate whether the product or service is hosted on a fully managed cloud platform, on-premises, or via a hybrid approach. In your response, explain the rationale behind this choice and detail how it supports Responsible AI (RAI) compliance and risk management measures.	• Fully managed cloud deployment (public Cloud/SaaS) e.g. AWS, Azure, GCP • On-premises deployment (in-house data centres, max control and data sovereignty) • Hybrid cloud deployment (mix of on-premises and public cloud) • Private cloud deployment (dedicated cloud environment, Virtual private cloud) • Multi-cloud deployment (multiple providers) • Edge deployment (on-device or local processing) • Ready-made AI (off-the-shelf, unmodified) • Customised third-party AI (pre-built	2.2.2	AI_G5, AI_G5A, AI_G5B

			model with tweaks, extended datasets) • Internally developed AI (in-house, proprietary model)		
T304	M	Please select which of the following best describes the AI value chain for your product or service, including any stages where you: • train or fine-tune models • refine or edit AI-generated content • validate, filter, or otherwise review outputs. Clarify which steps are performed internally and which (if any) are outsourced.	• Model training only (supply, curate training data and training of model, no post-processing of outputs) • Fine-tuning/customisation (fine-tune a third-party model using own data to align to educational context) • Output refinement (apply automated or manual post-processing to tailor raw AI outputs to meet quality and relevance criteria) • Human validation and filtering (educators/SMEs review and validate outputs pre-deployment) • Comprehensive involvement (involved at every stage of value train) • Minimal involvement (rely entirely on third-party AI solution for output generation, quality assurance, limited or no internal modifications)	2.1.5, 2.1.6, 3.2.1, 3.2.3, 4.1.1, 4.2.1	*AI_G5*

6.8 Criteria – Information and Support

Q	M	Question	Response options	VAISS	ST4S
T311	M	Do you maintain a formal Responsible AI (RAI) or acceptable AI usage policy?	A. Yes, we have a documented RAI/acceptable usage policy that covers all aspects of AI usage in our product or service. B. Yes, partial or in-development policy C. No, but under consideration D. No formal policy	1.2.1, 1.2.2, 8.1.1	AI_T1, AI_EV1
T312	M	How do you ensure users understand and apply your RAI policies, indicating the communication strategies and role-specific support resources?	• Included in Terms & Conditions: Our policies are embedded in our Terms & Conditions, which all users must accept • Dedicated user documentation: We provide a	6.1.1, 6.1.2, 6.1.3, 6.1.4, 7.1.1, 7.1.2	AI_PF9

			separate policy document or FAQ specifically about AI usage • In-product notifications or tooltips: We offer prompts or disclaimers within the user interface • Training sessions or guides: We conduct regular webinars, training modules, or guides that explain acceptable AI usage • User training and workshops – Live or recorded training sessions on ethical and responsible AI use • Dedicated support hotline or helpdesk – Direct access to experts for Responsible AI guidance • Automated guidance – In-product prompts, tooltips, or AI-driven assistance on responsible usage • Community support – Online forums, peer discussions, or open-source collaboration on best practices • No specific support provided – users are expected to determine responsible usage independently. • Communications addressed to educators • Communications addressed to parents • Communications addressed to administrators • Communications addressed to IT staff • Not applicable		

6.9 Criteria – Disclosure

Q	M	Question	Response options	VAISS	ST4S
T321		How do you manage the RAI risks associated with your product or service for teachers in the classroom?	• Minimal Risk: We believe our product or service poses no significant risk to teachers' roles • Positive Outcomes: We have identified potential positive outcomes, such as work efficiency • Identified Potential Risks, With Mitigations: We've identified potential workload or bias risks and have put measures in place (e.g. training, guidelines) to mitigate them • Ongoing Risk Monitoring: We actively monitor teacher	2.2.2, 7.1.1, 7.1.2	*AI_T1.1*

			feedback to adapt and refine how our AI features are used • No Formal Assessment: We haven't conducted a formal risk assessment for teachers yet • Not applicable		
T322		How do you manage the RAI risks associated with your product or service for students in the classroom?	• Minimal risk: We believe our product or service poses no significant risk to students' roles • Positive outcomes: We have identified potential positive outcomes, such as personalised learning paths • Identified potential risks, with mitigations: We've identified potential workload or bias risks and have put measures in place (e.g. teacher controls, content filters) to mitigate them • Ongoing risk monitoring: We actively monitor teacher and student feedback to adapt and refine how our AI features are used • No formal assessment: We haven't conducted a formal risk assessment for students yet • Not applicable	2.2.2, 7.1.1, 7.1.2	*AI_T1.1, AI_SF1*
T323		How do you manage the RAI risks associated with your product or service for administrators?	• Minimal risk: We believe our product or service poses no significant risk to administrators' roles • Positive outcomes: We have identified potential positive outcomes, such as administrative efficiency • Identified potential risks, with mitigations: We've identified potential workload or bias risks (including overreliance), and have put measures in place (e.g. disclaimers, support on interpreting data responsibly) to mitigate them • Ongoing risk monitoring: We actively monitor administrator feedback to adapt and refine how our AI features are used • No formal assessment: We haven't conducted a formal risk assessment for administrators yet • Not applicable	2.2.2, 7.1.1, 7.1.2	*AI_T1.1*

T324	M	How do you communicate to users the ways in which your AI system affects them or their work?	• Terms and Conditions • Clear documentation and disclaimers • In-product notification • User training or tutorials • No explicit communication • Students are made aware directly. • Educators are made aware. • Parents are made aware. • Openly • On Request	6.1.1, 6.1.2, 6.1.3, 6.1.4, 6.2.5, 6.2.6, 6.2.7, 9.2.2	AI_T1.4, AI_T7, AI_SF1
T325	M	Indicate how you inform users when system updates or contract changes might alter the AI's impact or usage conditions in your product or service	A. Re-consent or acknowledgement B. Formal update notifications C. Ad-hoc communications D. No notification process	6.1.4, 9.1.1, 9.2.2	AI_T7, AI_H2
T326	M	Do you provide comprehensive documentation about your AI system that is comprehensive and accessible to relevant stakeholders (e.g. users, parents, procurers, regulators), enabling them to assess compliance with RAI standards and guidance?	• Yes, publicly available: We have comprehensive RAI documentation accessible on our website or upon request. • Restricted or internal only: We provide technical documentation to regulators or specific partners only. • Minimal documentation: We share limited RAI information, focusing on user guides rather than compliance details. • No AI Documentation: We do not offer any RAI-specific documentation. • To students • To educators • To parents • To administrators • To procurers • To IT staff • To regulators	6.1.1, 6.2.5, 9.1.1, 9.2.2	*AI_T1.4, AI_SF1*

T327	M	Do you ensure that users are aware that they are interacting with AI? How?	• Explicit AI labelling: We label AI-generated content or interactions with clear markers like 'AI-generated' or 'AI Assistant.' • Onboarding or tutorials: We explain in our onboarding flow or terms & conditions that some responses or functionalities are powered by AI. • No distinction made: We do not explicitly state that users are interacting with AI.	6.1.1, 6.1.2, 6.2.2	AI_PF1, AI_SF1

6.10 Criteria – Explainability

Q	M	Question	Response options	VAISS	ST4S
T332		Did you design the AI system with interpretability in mind, ensuring it can be understood and trusted by educators and students (system interpretability)?	• External documentation: We include a dedicated 'How It Works' section in our user manual • In-app explanations: Our product displays contextual tooltips and visual cues that outline the reasoning behind AI outputs • Disclaimers only: We use disclaimers in the user interface to inform users that outputs are AI-generated and may have limitations • Multimodal approach: We combine in-app explanations with online FAQs and video tutorials • No formal explanations provided: We do not currently offer specific explanations for AI outputs • Not applicable	4.2, 6.1.2, 9.2.2	*AI_TI.4, AI_SF1*
T333		Indicate any varying levels of explanation you tailor to the user role or expertise?	• Fully customised: We offer simplified explanations for students and more detailed technical documentation for administrators and IT staff. • Partially customised: We have different documentation for educators and general users, with plans to extend role-specific materials. • Uniform explanations: All users receive the same level of explanation regardless of role. • Not applicable: Our AI system does not currently differentiate explanation levels. • To students	6.1.2, 6.1.3, 6.2.3, 6.2.7, 7.1.1	*AI_PF1A, AI_SF1*

			• To teachers • To parents • To administrators • To IT staff		
T335		Indicate whether you account for different sources of bias in the explanations you provide in T332 (e.g. training, RAG data, inherited upstream bias)?	A. Full disclosure, with suggested mitigations B. Full disclosure C. High-level disclosure D. No disclosure	3.2.2, 6.1.2, 6.2.5	*AI_PF1A, AI_PF9*
T336		What methods do you use to evaluate the effectiveness of your AI explanation materials?	• Regular user feedback: We routinely collect user feedback via surveys and focus groups to assess and improve our explanation materials. • Usability testing: We conduct regular usability tests and A/B testing to evaluate how well our explanations meet user needs. • Impact assessments: We perform impact assessments periodically to ensure our explanations are effective and adjust them based on the results. • No formal evaluation: We currently do not have a formal evaluation process for our explanations. • Not applicable	4.2.1,	AI_T4

6.11 Criteria – Accessibility and Inclusivity

Q	M	Question	Response options	VAISS	ST4S
F411		How does your AI system incorporate accessibility in its design, accounting for factors such as users with a disability?	• Adherence to established accessibility standards (e.g. WCAG) • Compatibility with assistive technologies (e.g. screen readers, Braille displays, alt-text) • Customisable interface settings (e.g. font size, contrast, text complexity,	7.1.1, 10.2.3, 10.4.2, 10.4.3	*AI_G1, P14, EV20*

			dyslexia-friendly fonts, spacing etc.) • Text-to-speech, speech-to-text, and other alternative input/output options, auto-corrective features, closed captions • Simplified navigation for users with cognitive or motor impairments • Involvement of people with disabilities in usability testing and feedback loops • No action taken • Not applicable		
F412		How does your system demonstrate inclusivity and adapt and/or meet the needs of diverse user capabilities or contexts?	• Personalised learning paths • Inclusivity features e.g. integrate text-to-speech, speech-to-text, closed captions, and keyboard navigation etc.; dyslexia-friendly formats, epilepsy-friendly formats • Multi-sensory engagement e.g. text, audio, visual aids, and interactive elements, gamification, multi-modal assessments, progress dashboard • Enables scaffolding and differentiation • Language support • Adaptive feedback • Inclusive design frameworks e.g. universal design principles • No action taken • Not applicable	10.2.1–10.2.4, 10.4.1, 10.4.2	AI_T1, *SC4*
F413	M	How do you ensure that your AI system treats all users fairly and equitably?	• Diverse and representative data sets: We collect and use training data from a wide range of regions, cultures, and user backgrounds, ensuring that no single group is over- or under-represented. • Bias identification and mitigation: We regularly audit our models and data pipelines for potential biases, and we run scenario-based assessments to simulate how various user groups might experience the system. We then implement strategies (e.g. re-sampling or re-weighting) to counteract the issues found. • Localisation and linguistic support: We support multiple languages and dialects, offer culturally relevant content, and adapt reading level and terminology for Cultural and Linguistic Diversity (CALD) communities. • Socioeconomic considerations: We explore options for low-bandwidth or offline access, and provide free or subsidised usage tiers to ensure the AI system is available to users with limited resources. • Continuous user feedback loops: We encourage feedback from all user groups and maintain dedicated channels for reporting fairness concerns, acting swiftly	2.2.2, 3.2.2, 4.2.1, 10.2, 10.3	AI_T1.1, AI_T4, *AI_HR1*

			on any identified issues. • Transparent policies and documentation: We make our approaches to data collection, model training, and fairness assessments clear and publicly available, enabling stakeholders to hold us accountable. • Regular external reviews: We partner with third-party auditors and community stakeholders to review our AI system's performance and fairness, obtaining unbiased insight and recommendations, e.g. Human Rights Impact Assessment. • Ethical governance: We have an internal Responsible AI committee overseeing the design and deployment processes, ensuring continuous alignment with fairness and equity principles. • No action taken		

6.12 Criteria – Equity of Access

Q	M	Question	Response options	VAISS
F421		What choices have you made to address difficulties with use of AI for schools and users in rural and remote areas?	• Low-bandwidth access • Offline and/or no-internet access mode (e.g. downloadable content as option, cached content) • Low-spec devices • Discounted pricing models • Scholarship, promotions, prizes • User testing, data collection of remote, rural, minority, disadvantaged communities • Solicit feedback and enable communication channels with underrepresented groups, or agencies representing them • No action taken • Not applicable	7.1.1, 9.2.2, 10.1.4

6.13 Criteria – Non-Discrimination

Q	M	Question	Response options	VAISS	ST4S
F431	M	Indicate the approaches or methods taken to mitigate or minimise adverse impacts (short- and long-term) of your use of AI on diverse user groups.	• Conducting focus groups with individuals from different linguistic, cultural, and socio-economic backgrounds to inform design decisions • Test and Integrating feedback loops at each development stage, ensuring that language localisation, accessibility options, and cultural	2.2.2, 4.1.1, 9.2.2, 10.3	AI_T4, EV2C

			sensitivities are addressed • Periodically reviewing user analytics and satisfaction data across different user segments to detect any emerging issues and update the AI system accordingly • Conducting bias audits on training data and model outputs • Implementing fairness metrics to track disparate impact • Incorporating diverse datasets from multiple demographic groups • Offering user feedback channels and rapid remediation processes • Engaging external ethics review boards or community stakeholders • Providing tiered or alternative access for marginalised communities • Integrating cultural, linguistic, and disability adjustments in the interface and content • Regularly retraining or updating models to correct identified biases • Other (e.g. additional design interventions or new accessibility features as needed) • No action taken		
F432	M	Indicate the processes and methods used to identify bias in your product or service both before and after deployment.	• Conducting pre-deployment audits by reviewing training data distribution, running fairness checks, and applying algorithmic bias detection scripts before production • Running pilot programmes or beta tests with demographically balanced user groups to spot potential bias or usability issues early • Monitoring post-deployment performance by analysing real-world metrics, user reports, and addressing any identified bias immediately • Monitoring post-deployment performance proactively through bias audits and algorithmic fairness checks at set intervals • Maintaining informal feedback channels (e.g. user forums, social media, direct email) to encourage open dialogue and guide continuous improvement • Measuring performance across demographic subgroups (e.g. gender, ethnicity, disability) to identify and track any disparate impact • Using established fairness metrics (e.g. demographic parity, equalised odds) to quantify bias issues • Employing external peer reviews or independent ethics committees for unbiased evaluation • Applying data augmentation or weighting techniques to improve representation of underrepresented groups	4.2.1, 4.3, 4.4, 5.1.2	AI_T4, AI_T11, AI_EV5, *AI_PF4#*

			• Collecting and incorporating user feedback, particularly from marginalised communities, to capture nuanced bias signals • Retraining or fine-tuning models when biases or performance gaps are detected • Ensuring transparent model documentation and accountability processes so that stakeholders understand how fairness is assessed and maintained • No action taken		
F433		Indicate which quantitative and qualitative metrics you use to evaluate for fairness in your AI system's outputs.	• Tracking error rates or accuracy across different demographic groups, measuring disparities, and investigating causes • Constructing balanced test sets that mirror diverse user subpopulations, monitoring performance differences with standardised metrics (e.g. F1 scores) • Applying general-purpose fairness metrics such as statistical parity, demographic parity, false-positive/false-negative rates, equalised odds, or equal opportunity • Comparing AI system outputs against expert or domain-specific benchmarks for a more holistic view of fairness • Gathering qualitative user feedback (e.g. direct testimony from educators, students, and parents) on whether recommendations feel equitable and representative • Conducting expert reviews via partnerships with accessibility specialists and diversity consultants to provide narrative evaluations on fairness • Collecting anecdotal feedback and documenting case studies from focus groups, diverse communities, or user interviews to uncover subtle, context-specific issues • No action taken	4.2.1	
F434		What measures and strategies have you implemented to manage or mitigate bias throughout the product or service's lifecycle?	• Data curation and augmentation: Proactively source diverse data sets; employ oversampling or augmentation techniques to ensure underrepresented groups are well-represented • Algorithmic interventions: Apply re-weighting or post-processing corrections to help equalise outcomes across different user groups • Fairness-by-design principles: Integrate fairness checklists and gating criteria at each development stage, making bias mitigation a standard practice rather than an afterthought	3.2, 4.2.1, 10.2.4, 10.4.1	AI_T4, AI_T11, AI_PF4, AI_EV5, AI_PF1A#, *AI_T1.1*

			• Disclaimers and documentation: Alert users about bias in UX disclaimers or in transparency documentation • Continuous evaluation: Regularly revisit and update model parameters, retrain with newer data, and monitor metrics to prevent previously resolved biases from resurfacing • Third-party audits: Invite neutral external organisations to review data practices, model assumptions, and system outputs for potential biases • Scenario-based testing: Develop use-case scenarios that represent different demographic or socio-economic backgrounds, then run tests to identify how the AI system responds under varying conditions • Policy-based constraints: Define clear rules or thresholds (e.g. disallowing certain flagged content or enforcing strict fairness measures) to override algorithmic decisions when bias is detected • Holistic data governance: Maintain documentation about data lineage, consent processes, and usage rights to ensure clarity and accountability throughout the data lifecycle • Regular stakeholder feedback sessions: Encourage ongoing conversations with teachers, parents, and students, particularly from underrepresented groups, to keep track of emerging issues. Feedback may additionally be sought from domain experts, particularly for high-impact decisions • No action taken		

6.14 Criteria – Copyright and Indigenous Cultural Intellectual Property

Q	M	Question	Response options	VAISS	ST4S
F441		Have you implemented any mitigations, monitoring or risk management planning to ensure that your product does not infringe copyright, moral rights, Indigenous Cultural Intellectual Property (ICIP), cultural rights and practices of Aboriginal and Torres Strait Islander people or protected cultural practice of any other culture?		2.1.6, 2.2.2, 3.1.5, 3.2, 3.2.6, 6.2.1	*PA3*
F442		Does the AI system reproduce third-party copyright material in outputs if the AI system uses retrieval augmented generation (RAG) or plugins for model alignment or attuning?		3.1.5, 3.2, 8.1.1	*PA3*

F443	M	Is any copyright material used as an input to the system (e.g. as data to train or fine-tune an AI model), whether that material is sourced and used by the vendor or a third-party? If so, please indicate the source of that material, the basis on which the material was used (e.g. with permission or licence of the copyright owner or in reliance on a copyright exception), and whether the vendor or a third-party undertook any relevant training or fine-tuning using that material.		3.1.5, 3.2, 3.2.3, 9.2.4	*PA3*

6.15 Criteria – General Accountability

Q	M	Question	Response options	VAISS	ST4S
A501	M	Within your organisation, indicate whether there are clear roles and responsibilities for managing the application of AI in your product or service (e.g. organisational chart descriptions of roles and responsibilities, job descriptions outlining competencies that are expected of staff).	A. Clearly defined roles exist (with organisational charts available). B. Roles are emerging/in development. C. Roles are not formally defined.	1.1.1, 1.1.2, 1.1.4, 6.1.4	AI_T1.1, AI_GO1
A502	M	For your AI supply chain, is there a statement of Shared Responsibility between your organisation and external providers, indicating who is responsible for what part of the process?	A. Yes, we have a formal shared responsibility statement. B. Partially, we are developing such a statement. C. No, such statements are not in place. D. Not applicable	8.1, 8.1.3	AI_GO2
A503	M	What training programmes have you implemented for your personnel around RAI?	A. Comprehensive training (e.g. regular workshops, online modules) B. Some ad hoc or initial training only C. No training programmes implemented	1.3.1, 1.3.2, 1.3.5	AI_HR1

6.16 Criteria – Human Responsibility

Q	M	Question	Response options	VAISS	ST4S
A511		AI systems may make decisions that negatively affect users. If your AI system makes decisions that may impact users, indicate whether these decisions are automated.	A. All decisions are automated with no human review. B. All decisions are automated with human review. C. Some decisions are automated; others require human intervention. D. All decisions remain human-controlled. E. Not applicable	5.1, 6.1.2, 6.2.2, 9.2.2, 10.3.1	*AI_T1.3, AI_G2*
A512		Does the AI system include human oversight mechanisms for AI-based outputs, actions and decisions to ensure meaningful intervention when necessary?	A. Robust human oversight is built in. B. Partial oversight is in place; improvements are planned. C. Such oversight mechanisms are not included. D. Not applicable	1.1.6, 4.4.4, 5.1.2, 5.1.7, 5.1.8, 10.4.4	AI_T1
A513	M	Indicate whether your AI system offers override functionality that allows authorised school staff to stop or modify its decisions. Describe any such mechanisms.	A. Through an admin override function accessible to school staff B. Override mechanism is limited or requires additional approval. C. There is no override functionality. D. Not applicable	4.4.4, 5.1.2	AI_G1B
A514		Do you provide advisory services, training, and ongoing support to schools to help them deploy and monitor your product or service in a way that maintains human control over AI-driven decisions?	A. We provide comprehensive deployment guidance, training and ongoing support. B. We offer initial guidance and training but limited ongoing support. C. We do not currently provide such support. D. Not applicable	5.1.7, 5.1.8 10.3.1	*AI_T1, AI_PF4*
A515	M	Does your product or service allow school staff to selectively block or restrict AI access for particular individuals? If so, please explain how this functionality works.	A. Access can be selectively blocked via school staff controls. B. Only through central administration C. Blocking access is not supported. D. Not applicable	5.1, 9.2.2	AI_A1, PF2

A516	M	What approaches have you taken to ensure that AI, as used in your product or service, respects individual autonomy? (Specifically: unproductive work, reduced opportunity to exercise initiative, insufficient engagement with users about what they want to achieve).	• Workflow and learning design: Design effort to anticipate how users can most effectively interact with the product • Task simplification features: Automate repetitive tasks (e.g. grading) without adding unnecessary complexity to teachers' or students' workloads • User-friendly interfaces: Design intuitive dashboards and workflows that reduce cognitive load and minimise confusion • Consultation with educators: Involve relevant stakeholders to ensure the AI system meets professional guidelines • The AI system is designed to promote user autonomy and we have conducted autonomy impact assessments. • The AI system responds to user autonomy concerns as they arise. • Some measures are in place; further evaluations are planned. • No formal measures or assessments have been conducted. • Not applicable	2.2.2, 5.1, 7.1.1	

6.17 Criteria – Reliability

Q	M	Question	Response options	VAISS	ST4S
A522	M	Do you address hallucinations in your AI system? (i.e. confidently stated but erroneous content) How?	• Model training and fine tuning using high quality, verified, domain-specific data to improve accuracy and truthfulness • Output filtering (e.g. confidence thresholds) and abstention mechanisms to avoid generating inaccurate information • Grounding (using RAG) and real-time fact-checking to validate outputs • Prompt engineering to avoid common hallucination misdirection	2.1, 2.2.2, 4.1, 4.2.1, 4.3, 6.2.1	*AI_T11, AI_PF1A, AI_PF4*

			• Output validation (self-consistency or ensemble methods) for cross-validation • Human oversight (where possible) • No action taken		
A523	M	Do you have defined acceptance criteria for AI system accuracy? What are they?	A. Yes, clearly defined criteria exist (details available in documentation). B. Criteria are defined but subject to review. C. No, acceptance criteria are not formally defined.	4.1.1, 4.2.1	AI_T11, *AI_PF1A, AI_PF4, AI_T1#*
A524		How reproducible are your AI system outputs: can you confirm that two consecutively run tests give the same result?	A. Yes, outputs are highly reproducible (e.g. >95% consistency). B. Reproducibility is variable. C. No, outputs are not reliably reproducible.	4.1.1, 4.2.1, 4.3	*AI_T1#*
A526	M	When variations occur in the AI system or its training data, indicate how you retest your outputs.	A. Yes, all changes trigger a retesting process immediately. B. Yes, all changes are tested as part of periodic testing. C. Retesting is done for significant changes only. D. No, retesting is not systematically performed.	4.2.1, 4.3, 4.3.1, 4.4	AI_L2
A528	M	Do you have formal processes in place to manage training data quality and document data provenance?	A. Comprehensive lifecycle management: We have formal, documented procedures for inventorying, validating, and disposing of training data. B. Inventory and validation only: We maintain a detailed inventory and regularly validate our training data, but our disposal process is less formal. C. Ad-hoc or partial processes: We manage training data on an as-needed basis without a fully formalised lifecycle process. D. No formal lifecycle management: We do not currently have a structured process in place for managing the training data lifecycle.	3.2.1, 3.2.2	*AI_T9, AI_T8*

6.18 Criteria – Monitoring

Q	M	Question	Response options	VAISS	ST4S
A531	M	Do you have documented processes for testing, monitoring, and continuous improvement of your AI system with respect	A. Yes, comprehensive and documented processes exist. B. Processes are in place but are not fully	4.1, 4.2.1, 4.3.1, 4.4,	AI_T1.3, AI_T4, AI_EV2

		to accuracy, reliability, and explainability (e.g. ongoing testing, benchmarking, proactive monitoring, red teaming)?	documented. C. No, we do not have formal processes.		
A532	M	Do you have documented processes for testing, monitoring, and continuous improvement of your AI system with respect to human rights, wellbeing, and fairness?	A. Yes, these processes are documented and regularly reviewed. B. Partially, with improvements underway. C. No, we do not have such processes.	2.2.2, 4.1, 4.2.1, 4.4, 10.2	*AI_T1.3, AI_T4, AI_EV2*
A533		Do you monitor for incidents in your AI system, including human rights concerns and inaccuracy?	A. Yes, incidents are monitored continuously, and a detailed incident register is maintained. B. Yes, incidents are monitored continuously. C. Monitoring occurs on an ad hoc basis. D. No, incident monitoring is not conducted.	4.1.1, 4.4, 5.1.4, 8.1.2	AI_I2
A535	M	Do you have a formal incident management plan in place that addresses incidents in your AI system as they occur?	A. Yes, we have a formal incident management plan and protocols. B. A plan exists but is not fully implemented. C. No, we do not have a formal incident management plan.	2.2.2, 4.2.3, 4.4.3, 8.1.2	AI_I2, AI_EV3
A536		Do you monitor outputs from your AI system post-deployment for bias, discrimination and lack of equity?	A. Yes, monitoring for these issues is routine. B. Monitoring is sporadic. C. No, these aspects are not monitored.	4.2.1, 4.4. 5.1.2, 5.1.4, 10.1.4	AI_L2
A537		Do you have defined acceptance criteria for your AI system outputs for bias, discrimination and lack of equity?	A. Yes, criteria are clearly defined. B. Criteria exist but require further refinement. C. No, such criteria are not established.	2.2.2, 4.2.1, 4.2.3, 4.4.1	*AI_SF4*
A539		Do you conduct regular system audits for compliance of your AI system against your established Responsible AI acceptance criteria?	A. Yes, we commission regular external audits B. Yes, internal audits are conducted regularly. C. Audits are conducted but not on a regular schedule. D. No, we do not conduct regular audits.	4.5, 4.5.1, 8.1.2, 9.2.2	AI_L2, *AI_PF7, AI_L1, AI_L1A, AI_L1B, AI_T11*
A5311		Do you establish clear responsibilities for schools to monitor the performance and conformance of AI in your product, and communicate concerns back to you?	A. Schools are expected to provide information back regularly, following a written agreement describing what kinds of monitoring are expected from the school, and frequency of feedback. B. Schools can provide information back to the vendor on an ad hoc basis, but there is no formal	4.5, 7.1.1	AI_T1.1, AI_PF4, *AI_PF9*

			agreement outlining expectations on the subject of monitoring or its frequency. C. No action taken		

6.19 Criteria – Contestability

Q	M	Question	Response options	VAISS	ST4S
A542		Is user feedback systematically used to refine and improve your AI system's outputs?	A. Yes, feedback is actively used to refine outputs and can be provided through multiple channels (e.g. in-app, email, online forms). B. Yes, feedback is actively used to refine outputs. C. Feedback is reviewed but changes are infrequent. D. No, feedback does not affect outputs.	4.4.2, 4.4.3, 4.4.4, 7.1.4, 10.1.4	AI_T1.5:, AI_PF4, *AI_PF9*
A543	M	Do you offer formal mechanisms that allow users and parents to challenge or appeal AI-based decisions? (Select all that apply.)	• Appeal or review process – A formal process for users/parents to request a review of AI decisions • Human oversight for disputes – A human decision-maker is available to override or explain AI outcomes • Transparency in decision-making – Clear documentation on how decisions are made and how they can be contested • User feedback and reporting tools – In-product options to flag, dispute, or provide feedback on AI decisions • Independent third-party audits – External review mechanisms to assess fairness and accountability • Legal or ethical review board – An internal body to handle AI-related complaints • No formal challenge mechanism provided – Users/parents have no official recourse to dispute AI decisions	5.1.2, 7.1.1, 7.1.2, 7.1.3, 8.1.3	*AI_PF6*
A545		Do you have a formal policy that addresses user complaints, the contestation of AI system outputs, and liability issues?	A. Yes, a formal policy exists covering complaints, contestation and liability B. A policy is in development C. No, such a policy is not in place	7.1.5, 9.2.1	*AI_T1.5*

6.20 Criteria – Evidence

Depending on supplier responses to prior questions, the following documentary evidence is required to be uploaded (system accepts PDF, .DOC, .DOCX). The following includes an explanation of how the evidence relates to individual controls, which is guidance for how the evidence will be used in assessment.

Q	Question	Individual controls	Relating to Question ID	VAISS	ST4S
E1	Architecture documentation, design specification: documents that outline how the AI functionality of your product or service has been designed, and how its architecture complies with RAI expectations. Ensure that your design documents cover, where applicable, stakeholder engagement, human-centred and child-centred design principles, and human oversight and override. Ensure that your architecture documents cover, where applicable, AI configuration, the AI value chain, user feedback and system retraining, diversity, and explainability. Indicate the country/ies where your AI is physically hosted. Include diagrams or visual aids where helpful.	DESIGN • Specifies the design methodology of AI in the product, including whether any specific human-centered or child-centred design approaches were used (H214), and the extent and forms of stakeholder engagement (H233) • Specifies what the AI is used to do in the product or service (G5) and why (G6), including what kinds of outputs the AI generates for users (G7) • Indicates the extent of human oversight over any decisions the AI makes, in interacting with the outside world (A511, A512), and what provision there is for human override (A513) ARCHITECTURE (T301) • Indicates which components of the product or service are AI, and which are not (T301), including what countries the AI components are hosted in (T301) • Specifies how the AI service is configured and hosted (T303), and how the arrangement of AI components enables value to be added, as an AI value chain (T304) • Specifies what guardrails or other components the product puts in place to address high risk and prohibited content (H211) • Indicates whether there is a feedback loop in the system from users to retraining or updating AI outputs (A542)	T301	3.2.1, 3.2.2, 4.1.1, 4.2.1, 5.1.1, 5.1.2, 7.1.1, 8.1.1, 9.2.4	AI_EV4

		• Mentions how the architecture and design of the product or service addresses concerns with diversity (H221), interpretability (T331)			
E2	Bill of Materials: breakdown of the different systems involved in your AI functionality, and the data they generate. Ensure the documentation includes any statements of shared responsibility (which party you are working with is responsible for what functionality), and whether external parties are responsible for any of the guardrails on high-risk outputs of your AI system.	• Indicates who is providing high-risk guardrails (H236) • Provides a clear statement of shared responsibility (A502)	T301	2.1.5, 2.1.6, 3.1.1, 3.2.1, 5.1.2	
E3	Risk impact assessment: a risk impact assessment should inform all the responses to questions around risk in this evaluation. It should also address explainability and communication with users as means of managing risk.	• Covers prohibited/high-risk output (H211) • Covers inappropriate human–AI interface (H213) • Covers automated decision-making (H212, A511) • Covers human dignity (H234), individual autonomy (A516), and worker/learner rights (H235) • Covers risks specific to teachers (T321), students (T322), and administrators (T323) • Covers remote/rural users (F421) and diverse user cohorts (F431) • Addresses explainability in the context of risk management, and how to ensure it is effective (T336)	G9	2.2.2, 6.1.2, 9.2.1, 9.2.2	
E4	Risk management plan: a risk management plan should address all the types of risk discussed in this evaluation. It should also address explainability and communication with users as means of managing risk.	• Describes the risk management system in place for development of the product or service (G8) • Covers prohibited/high-risk output (H211) • Covers inappropriate human–AI interface (H213) • Covers automated decision-making (H212, A511) • Covers human dignity (H234), individual autonomy (A516), and worker/learner rights (H235) • Covers risks specific to teachers (T321), students (T322), and administrators (T323)		2.1.4, 2.2.2, 8.1.1, 8.1.3	

		• Covers remote/rural users (F421) and diverse user cohorts (F431) • Covers identification of bias (F432) and ongoing management of bias (F434) • Addresses how risk and impact are communicated to users (T324) • Addresses explainability in the context of risk management, and how to ensure it is effective (T336)			
E5	Responsible AI policy: indicates how the organisation is committed to RAI principles. Should include discussion of diversity, inclusion of fairness; human-centred and child-centred design; and how RAI informs technical decisions around AI. Include documentation on roles and responsibilities in the organisation for managing AI, and for training personnel in RAI	• Indicates how development follows human-centred and child-centred design principles (H214) • Indicates organisational commitment to diversity, inclusion, and fairness, overall and specifically in the use of AI (H223) • Includes roles and responsibilities in the organisation for managing AI (A501) • Includes commitment for training staff in Responsible AI (A503) • Discusses how Responsible AI principles inform the organisation's technical decisions around AI, including AI architecture (T303) • Includes an AI Governance policy aligning with organisational objectives and values: policy document outlining responsible development and deployment, aligned business goals, ethical principles, adherence to regulatory compliance/requirements where possible (e.g. antidiscrimination), 12 month reviews of policy, meeting minutes, organisation chart, role descriptions. Could ask to be made available on request. (T311)	T311	2.1.1, 2.1.2, 3.1.1, 7.1.1, 7.2.2, 8.1.1, 8.2.1, 8.2.2	AI_EV1
E6	Incident management plan: indicates how the organisation monitors, records, and addresses incidents relating to Responsible AI as they occur.	• Indicates how incidents are monitored (A533) • Indicates how incidents are recorded in a registry (A534) • Indicates how incidents are addressed (A535)	A535	2.1.4, 8.1.1, 8.1.4	AI_EV3

E7	Test processes, test logs: indicates how the organisation tests and monitors for Responsible AI concerns, including explanation efficacy, bias, accuracy, and reliability. Include acceptance criteria, test schedules, and testing and monitoring methodologies.	TESTING • Indicates how effectiveness of explanations is tested (T336) • Indicates how and when bias is identified (F432), and with what kinds of metrics (F433). Necessary documentation could include fairness assessment reports and bias test results, data representativeness analysis, bias mitigation documentation, fairness monitoring logs, incident resolution reports, training data audit reports, fairness metric trending reports. • Indicates how AI is tested for accuracy and information integrity (A521, A531) and hallucinations (A522), with what acceptance criteria (A523) • Indicates how AI is tested for reproducibility (A524) • Indicates test schedule for AI (A525), including whether testing is done on variation to the system or training data (A526) MONITORING • Indicates how AI is monitored for accuracy, reliability, and explainability (A531) • Indicates how AI is monitored for human rights & wellbeing, and fairness (A532) • Indicates how AI is monitored for bias, discrimination, and lack of equity (A536), including the use of differential audiences (A538), and with what acceptance criteria (A537) • Indicates how AI monitoring is coordinated with schools (A5311) • Indicate whether AI is subject to regular system audits for Responsible AI compliance (A539)		4.1.1, 4.2.1, 4.3.1, 5.1.1, 5.1.2	AI_EV2C
E8	Data management policy, data provenance documentation	• Indicates the extent of data quality and provenance management processes (A528) • Indicates whether data provenance is documented (A529)	A528 A529	3.2.1, 3.2.2	

		• Addresses concerns around Indigenous & Cultural Intellectual Property (F441) • Addresses concerns around copyright and third-party content (F442, F443)			
E9	Terms and Conditions, usage policy, consent forms: indicates how informed consent is established for different user groups, whether users can opt out of using AI, and how user complaints and liability are addressed. Include how policies are communicated to users, and how users are kept up to date with policy changes.	• Indicates how informed consent is established for different user groups (H231) • Indicates whether and how users can opt out of using AI (H232) • Demonstrates how the Responsible AI policies of the organisation are communicated to users (T312) • Indicates how users are made aware of updates to the AI system policies or contracts (T321) • Indicates how user complaints and liability are addressed (A545)		6.1.2, 8.1.1, 9.2.1, 9.2.2	
E10	User documentation, product brochure: a combination of documentation indicating what the product or service does using AI, what classes of users interact directly or indirectly with AI, and whether the AI carries out automated actions or decisions. It should indicate the level of support offered to schools around Responsible AI usage and monitoring; whether AI functionality can be customised for different audiences; whether AI-based decision-making can be overridden; and whether access can be disabled for specific users. It may also indicate how users are made aware they are using AI; what disclaimers about AI quality they are exposed to; and how explanations of AI outcomes are presented to users.	• Describes the intended use of the product or service (G1), whether its intended users interact directly with AI (G4), and the groups it impacts on (G3) • Describes what functionality is carried out by the AI (G5), and what outcomes are produced using AI, including outputs, actions, and decisions (G7) • Indicates whether AI functionality can be customised for different schools and school communities (H222) • Describes what kinds of support around Responsible AI usage are provided to users by the organisation (T312) • Indicates whether and how users are made aware of when they are interacting with AI (T327) • Indicates how interpretability (explainability) of AI outcomes can be accessed by users (T331), and whether different classes of users access different types of explanations (T333) • Indicates how to override or stop any AI-based decision-making (A513)		6.1.2, 9.2.1, 9.2.2	AI_EV4

		• Explains how users can monitor the system to retain human control, and describes the extent of support offered to schools to that purpose (A514) • Indicates whether AI access can be blocked for specific individuals (A515)			
E11	Transparency documentation: indicates the risks exposure to users, suggested mitigations, can be used to determine compliance with the full range of Responsible AI principles.	• Documents the kinds of outcomes AI produces in the product or service, and their respective risks (G7) • Documents the risks and mitigations for prohibited/high-risk usage (H211), inappropriate human-AI interface (H213), and automated decision-making (H212) • Documents the risks and mitigations for human dignity (H234) and worker and learner rights (H235) • Describes what kinds of support around Responsible AI usage are provided to users by the organisation (T312) • Describes how users are made aware of the product or service having an impact on them (T324) • Indicates whether and how users are made aware of when they are interacting with AI (T327) • Indicates whether the product or service UX enables explanations of its decisions (T331), with different sources of bias explained (T335) • Indicates whether different classes of users access different types of explanations (T333) • Documents how AI design addresses accessibility (F411), inclusivity (F412), diversity (F413, F431), and rural and remote access (F421) • Documents the risks and mitigations around bias (F434) • Documents the risks and mitigations around automated decision-making and human oversight (A512, A513) • Explains how users can monitor the system to retain human control, and describes the extent of support offered to schools to that purpose (A514) • Documents the risks and mitigations around accuracy and reproducibility (F434)	T326	3.2.2, 6.1.2, 9.2.2	

		• Explains the range of mechanisms available for users to challenge AI decisions or outcomes (A543)			
E12	External audit: audit of the AI system by an external party regarding Responsible AI concerns	• External audit (A539)	A539	8.1.1	AI_EV5

In addition, other evidence may be requested or inspected throughout the assessment process. This includes information on a supplier's website, terms and conditions, privacy policies and other documentation or information.

In assessing and reviewing documentation requirements, the Responsible AI (RAI) Assessment Team makes considerations to:

- Content: Does the document contain the sections per the relevant RAI Assessment control as described in the table above. Documentation should contain specific and relevant technical information to the service being assessed.
- Quality: Does the document demonstrate a level of standard relevant to the expectation laid out in the control.

Documentation Requirements:

All documentation provided throughout the assessment must be in English, be an authorised and final copy by the organisation and contain the organisation's name and company number.

Verification and Validation:

The RAI Assessment Team may contact the author, certifying body etc to verify the authenticity of documentation, evidence, and other information.

Appendix A – Standards, Frameworks and References

Standard / Reference	Weblink
Australian Voluntary AI Safety Standard	https://www.industry.gov.au/publications/voluntary-ai-safety-standard
Safer Technology for Schools	https://st4s.edu.au/st4s-vendor-guide/

Appendix B – Storage and Processing of Information

The information within this appendix describes how information submitted as part of the Responsible AI (RAI) Assessment process may be stored, processed and handled. For information regarding how we share results, submission data and other information with other parties please refer to *'3. Sharing and use of full assessment reports, findings and outcomes.'*

Transmission of Information:

a. We are committed to maintaining the confidentiality and security of the documentation you upload. We employ industry-standard security measures to assist us in safeguarding supplier's information from unauthorised access, disclosure, alteration, or destruction.
b. The technology providers we use to deliver the RAI Assessment process to you have achieved cybersecurity assessments or industry certifications such as ISO27001, SOC2 etc. You can find out more by visiting their website listing in the subprocessors table below.
c. We use technologies such as HTTPS and file encryption to protect files as they are transmitted from suppliers to our devices and other services we use to store and process information.
d. You understand and acknowledge that while we take reasonable precautions to protect your information, no method of transmission over the internet or electronic storage is completely secure, and we cannot guarantee absolute security.
e. When transmitting information to us, it is your responsibility to ensure that you access the file hosting and transfer services in a secure manner. This includes applying protections such as installing and using antivirus and malware protection on your devices and verifying you are connecting to the correct website address.
f. We use services and other tools which may reside outside of Australia. This means your information may be transferred and potentially stored in other countries. Our preference is to select legal jurisdictions which have privacy and data protection laws that are comparable and/or provide stronger protections to personal information. Please refer to the sub-processor list in this guide for further information.

g. When providing documentation to us, you may distribute these via email however our preference (particularly for evidence documentation) is to utilise a direct upload to our questionnaire tool.
h. If you wish to provide limited access to documentation to us, you may discuss various options with your assessment officer. We may be able to review documentation and evidence with 'read-only' permissions on your cloud storage provider or access documentation from your trust portal or security centre. Should you wish to provide us to tools or services to access this information, it is your responsibility to ensure our access is setup correctly, securely, that you communicate credentials in an encrypted and secure manner to us, and that you appropriately handle our access or accounts by disabling access when the RAI Assessment review or activity is completed.

Storage and Retention:

1. We store the documentation you upload on company devices and various online services we use at our organisation and as part of the RAI Assessment (referred to as sub processors). These services are described within this guide and includes include such as Alchemer or Zendesk which we use to collect and process information.
2. Sub processors we use may store and process information differently to us. This may include storing information in different countries, encrypting files to a different standard and more. You can find out more information about how our sub processors process, store and secure your information by visiting their privacy policies and other published information (e.g. security documentation) the sub processor has published on their website.
3. We retain documentation and information for as long as we deem reasonably necessary (generally a minimum of 7 years) in order to:
    a. Fulfill our obligations to the RAI WG and facilitate business activities related to the RAI Assessment program;
    b. Ensure a RAI WG member (such as the Department of Education in a State/Territory within Australia) is complying with their archival and information keeping rules, legislation, regulations and policies;
    c. Conduct RAI Assessment activities and ongoing monitoring and compliance of services to the RAI Assessment framework, standards and requirements;
    d. Conducting audits, reviews and other activities related to the RAI Assessment program.
    e. Other business purposes relating to the RAI Assessment program such as improving the RAI Assessment framework and its criteria.
4. If documents or evidence is being submitted in relation to an assessment completed for a specific RAI WG member, then the evidence and documentation may be shared with them. Examples may include where you or your organisation has been invited to complete the RAI Assessment on behalf of, or by a government agency and we have noted within the invitation, by email or other methods that the evidence will be shared with the RAI WG member/s. Please check your original invitation email which will describe if this is to occur.

Sub Processors:

The RAI Assessment Team uses the following third-party services to deliver the RAI Assessment. These services may store or process your information, including personal information. Importantly, if your organisation lodges documentation to us, these documents may also include personal information if you have not redacted it. Please be mindful when sharing documentation and information to us and ensure you are authorised to disclose such information (particularly personal information).

Name	Link	Purpose	Personal Data Types	Storage Country
Alchemer	https://www.alchemer.com/	Facilitates online questionnaires for the assessment, Evidence File Uploads, and other data collection activities such as the feedback and consultation form.	Contact information (e.g., name, email, phone number), IP address, device information.	Germany*
Directus	https://directus.io	Relational database to store data surrounding assessment, including data lodged in Alchemer.	Contact information (e.g., name, email, phone number).	Australia
Retool	https://retool.com	Front end for assessors to interact with assessment. Processes and updates data entered in Alchemer, which will be used to produce the report.	Contact information (e.g., name, email, phone number).	Singapore
N8n	https://n8n.io	Workflow engine, used to generate assessment report based on data in Retool.	Contact information (e.g., name, email, phone number).	Germany*
Microsoft Office365 (O365)	https://www.microsoft.com/en-au/microsoft-365	Stores contact details of suppliers and stakeholders engaging with the RAI Assessment, stores assessment files and supporting documentation.	Contact information (e.g., name, email, phone number).	Australia
Zendesk	https://www.zendesk.com/	Customer support tool and portal. Maintains contact information, enables communication with suppliers and stakeholders via email, chat or other means. Hosts knowledge base articles.	Contact information (e.g., name, email, phone number).	Australia
Amazon Web Services (AWS)	https://aws.amazon.com/	Hosts the public website. Facilitates data integrations such as processing data and transferring information between our services and sub processors. Hosts our email for st4s.edu.au domains.	Contact information (e.g., name, email, phone number), IP address, device information.	Australia
Auth0	https://auth0.com/	Manages authentication for suppliers accessing RAI Assessment questionnaires, the knowledge base, supplier portal, and access into our other services.	Contact information, IP address, device information.	Australia
Google Analytics	https://analytics.google.com/	Facilitates analytics for the public website. Also provides engagement metrics for knowledge base articles.	IP address, device information such as browser type.	Australia

** The service provider does not offer hosting in Australia and this time. We have opted for a European Union member state due to strong privacy and data protection laws operating in this jurisdiction.*

Appendix C – Responsible AI (RAI) Assessment Excluded List

For a list of product categories not assessed by the Responsible AI (RAI) Assessment, please refer to the Safe Technologies for Schools (ST4S) Excluded List, published on the ST4S website. Excluded product categories are consistent between the ST4S and RAI Assessments and are excluded due to their specialised nature, the framework not yet having the controls to cover the features or functions, or because the service falls outside the primary educational context.

Appendix F – Definitions

Item	Definition
Harmful	Harmful is defined to include anything that is objectionable, illegal or unlawful, and restricted content made available to the wrong age group.
NSFW	Not Safe for Work is a general term used to describe any content that may be deemed inappropriate to create, view or access whilst in the workplace (and schools). NSFW content includes content that may be deemed inappropriate for younger audiences (e.g. persons under 18). Examples of NSFW content include nudity, excessively violent images, offensive material etc. NSFW content can also apply to text-based content. Examples may include AI services which may generate overly violent or sexual text stories or engage in conversation of this nature.

Appendix G – Supplier Code of Conduct

All suppliers participating in the Responsible AI (RAI) Assessment process must demonstrate due care and skill, remain transparent, adhere to the conditions within this guide, and refrain from misconduct. Misconduct matters may be referred to the RAI WG and/or NEDAG, alongside relevant information such as correspondence or internal notes.

A decision as to what constitutes misconduct and whether an organisation or individual has engaged in misconduct is at the sole discretion of the RAI Assessment Team. Examples are provided in this appendix.

Escalation Path and Resolutions

1. **Initial Decision**
   The assigned assessment officer makes the initial decision. Decisions of risk levels and compliance and pre-defined within the RAI Assessment framework. This includes the risk and treatment wording on reports and risk outcomes. Most cases are resolved by assessment officers. In the first instance, you should work with the assessment officer to resolve the matter and provide technical information to support their review.
2. **Program Manager Review**
   If necessary, the decision is escalated to the RAI Assessment Program Manager.
3. **Primary Member Review**
   If necessary, the matter is referred to the RAI Working Group member who either assigned the most nominations to your assessment, referred or invited

your organisation to participate in the RAI Assessment process (the *primary member*). In most cases, the primary member can resolve the matter. Their decision is final unless they choose to escalate the issue to the RAI Working Group, however they are not required to.

4. **RAI Working Group Review**
   If the matter is still unresolved, the primary member may choose to refer the matter to the RAI Working Group for a final decision by majority vote.

Examples of Misconduct

Misconduct includes, but is not limited to, the following:

1. **Plagiarism and Copyright Infringement**
   Using material belonging to another organisation or individual without proper authorisation. This includes on your website, privacy policy etc.

2. **Solicitation of Outcome**
   Attempting to influence an outcome whether by bribery, legal threats, bypassing the escalation process or other improper means.

3. **Bypassing Procedure or the Escalation Path**
   Lodging a complaint about the RAI Assessment process, outcome, or decisions to senior management at ESA instead of following the 'Escalation Path and Resolutions' process.

4. **Misleading or Deceptive Conduct**
   Providing false or misleading information during the assessment or related activities or making misleading representations.

5. **Being Uncooperative or Unprofessional**

   Being uncooperative throughout the assessment process or communicating in a manner we deem to be forceful, rude, inappropriate, aggressive, and/or unprofessional.

6. **Misrepresentation of RAI Assessment Status**
   By act or omission, misleading others about your status within the RAI Assessment and any other related RAI Assessment activities.

7. **Breaches of Terms and Conditions**
   Violating any condition specified in this guide, the declaration you sign when completing an RAI Assessment or form, or any other instruction issued by the RAI Assessment Team.

8. **Other Misconduct**
   Any other behaviour that the RAI Assessment Team reasonably believes amounts to misconduct.

What is not misconduct

1. **Failing to Meet Criteria**
    - Receiving one or more non-compliant items during the Assessment and committing to either discussing further with us or withdrawing from the process as you may choose. Our goal is to help you achieve a compliant outcome. Please see the 'Support' section for more details.
2. **Requesting a Re-Review or Escalation**
    - You may request a second opinion or escalate a decision if you disagree with an outcome (see 'Escalation Path and Resolutions'). A challenge must be accompanied with technical reasoning and explanation.
3. **Genuine and Honest Mistakes**
    - Errors made in good faith, genuine misunderstandings, or providing incorrect information due to an honest misunderstanding.
4. **Fair Discussion and Debate**
    - Expressing genuine disagreement with the criteria, provided you remain open to further discussion with the RAI Assessment Team and provide reasoning.
5. **Constructive Critique and Feedback**
    - Offering constructive feedback on the assessment process, criteria, or related RAI Assessment activities.

If you have any concerns regarding the code, please contact the RAI Assessment Team on our website or directly to our assessment mailbox.