



Department for  
Science, Innovation  
& Technology



Central Digital  
& Data Office

# Algorithmic Transparency Recording Standard

## *Guidance for Public Sector Bodies*

[v3.0 update]

### Contents

Overview .....	1
Using this guidance.....	3
Step 1: Prepare to complete an algorithmic transparency record.....	3
a. a. What tools does the Algorithmic Transparency Recording Standard apply to? .....	3
b. b. When should you complete your algorithmic transparency record? .....	4
c. c. How should you complete and submit your algorithmic transparency record? .....	4
d. d. Exemptions: what information should not be published?.....	6
e. e. Considerations before publishing.....	6
Step 2: Complete the algorithmic transparency record.....	8
f. Tier 1 Information .....	8
g. Tier 2 – Owner and Responsibility (2.1).....	9
h. Tier 2 – Description and Rationale (2.2) .....	10
i. Tier 2 – Decision making process (2.3) .....	11
j. Tier 2 – Technical Specification and Data (2.4) .....	13
k. Tier 2 – Risks, Mitigations and Impact Assessments (2.5) .....	18
Step 3: Upload and update your algorithmic transparency record .....	21
l. Updating records.....	21
m. Version Management .....	21
Annex 1: Glossary of Key Terms .....	22

## Overview

### ***What is the Algorithmic Transparency Recording Standard?***

The Algorithmic Transparency Recording Standard (ATRS) is a framework for capturing information about algorithmic tools, including AI systems. It is designed to help public sector bodies openly publish information about the algorithmic tools they use in decision-making processes that affect members of the public.

Transparency is a key component of the safe, fair, just, and responsible use of algorithmic tools. However, many public sector organisations are unsure how to be transparent when using algorithms to deliver services. The ATRS provides a clear and accessible means of communication, and contributes to a more effective and accountable approach to public service delivery.

### ***What is an ‘algorithmic tool’?***

An algorithmic tool is a product, application, or device that supports or solves a specific problem using complex algorithms.

We use ‘algorithmic tool’ as an intentionally broad term that covers different applications of artificial intelligence (AI), statistical modelling and complex algorithms. An algorithmic tool might often incorporate a number of different component models integrated as part of a broader digital tool.

When determining if the ATRS is applicable, the use case where the algorithmic tool is applied is typically more relevant than the complexity of the tool itself.

### ***How do I know if I should complete an algorithmic transparency record?***

The ATRS has been created for public sector organisations to communicate about the algorithmic tools they are using in a standardised way. There is more detail on scope contained within the guidance, but most simply it is most relevant for algorithmic tools that either:

- Have a significant influence on a decision-making process with direct or indirect public effect, or
- Directly interact with the general public.

Algorithmic transparency is an important enabler of building public trust in our increasingly data-driven world. However, we are conscious that certain information should not be openly shared with the public as it may cause harm, for instance, endangering health and safety, prejudicing law enforcement, prejudicing someone’s commercial interests, or allowing ‘gaming’ of certain processes. Section 3.4 of this guidance covers the exemptions framework we have developed for the Standard.

In the AI white paper consultation response published in February 2024, we announced that use of the ATRS will become a requirement for all central government departments, with an intent to extend this to the broader public sector over time. Beyond this mandatory requirement for central government departments, organisations in the broader public sector are encouraged to complete an algorithmic transparency record for every algorithmic tool used which meets our [scoping criteria](#). These records will be uploaded onto our [GOV.UK repository](#) where they will be accessible to the general public and interested stakeholders.

### ***Why does this matter?***

Algorithmic transparency enables public scrutiny and greater accountability of public sector decision-making processes involving algorithms, helping to fulfil the public's democratic right to information. Increasing public awareness and understanding of the use of algorithms in the public sector is also essential to building greater public confidence and trust both in the government and its use of technology.

Proactively providing information in this way also can also:

- support senior risk owners in departments to understand the algorithmic tools that are being deployed, and take meaningful accountability for their use;
- highlight good practice and innovative use cases of algorithmic technologies;
- help to identify potential problems with a given tool early and mitigate the risk of public sector organisations implementing poorly designed tools;
- reduce administrative burden on public sector bodies by pre-emptively answering questions which may otherwise be raised through Freedom of Information requests;
- benefit external suppliers of algorithmic tools by providing clarity around the transparency requirements involved in supplying them to public sector organisations.

## Using this guidance

Step 1 of this guidance supports you in preparing to complete an algorithmic transparency record. This includes advice about:

- how to determine which tools the Standard applies to;
- when you should complete your record;
- how you should complete and submit your record;
- how to determine whether particular information may be exempt from inclusion in a record;
- the risks you should consider before doing so.

Step 2 supports you when writing your algorithmic transparency record. It covers:

- how much detail is required in each section;
- how to navigate overlaps with other guidance and impact assessments;
- examples to illustrate how each section should be completed.

Step 3 describes the process for uploading your completed record, and how and when to update your record.

## Step 1: Prepare to complete an algorithmic transparency record

### a. What tools does the Algorithmic Transparency Recording Standard apply to?

To assess whether your tool is in scope, we would encourage you to reflect on the questions in the scoping criteria included below.

The Algorithmic Transparency Recording Standard is most relevant for algorithmic tools that either:

- Have a significant influence on a decision-making process with direct or indirect public effect, or
- Directly interact with the general public.

To decide whether your tool has a public effect, you might want to consider whether usage of the tool could:

- materially affect individuals, organisations or groups?
- have a legal, economic, or similar impact on individuals, organisations or groups?
- affect procedural or substantive rights?
- impact eligibility for, receipt of, or denial of a programme?

Examples of tools that could fall within the scope of these criteria are:

- A machine learning algorithm providing members of the public with a score to help a government department determine their eligibility for benefits (impact on decision making and public effect)
- A chatbot on a local authority's website interacting directly with the public which responds to individual queries and directs members of the public to appropriate content on the website (Direct interaction with the public)

Examples of tools that would likely not fall within the scope of the criteria include:

- An algorithm being used by a government department to transform image to text (e.g. used in digitisation of handwritten documents) as part of an archiving process (no significant decision or direct public interaction)
- An automated scheduling tool which sends out internal diary invites from a mailbox (doesn't have public effect)

The context of use of the algorithmic tool matters here. The same image to text algorithm above might be relevant if being used instead to digitise paper application forms for a government service (e.g. poor performance of the algorithm on some handwriting styles might well have an influence on success rates for individual applicants).

If you are using an algorithmic tool that does not strictly meet these criteria but you would like to provide the general public with information about it, you can still fill out an algorithmic transparency record.

If you have any questions about this, or about which algorithmic tools you should prioritise, you can reach out to the ATRS team at [algorithmic.transparency@dsit.gov.uk](mailto:algorithmic.transparency@dsit.gov.uk).

## **b. When should you complete your algorithmic transparency record?**

Algorithmic transparency records should be published and made publicly available when the tool in question is being piloted and/or deployed. However, it is highly recommended that you start to discuss and fill out the transparency record during the design and development phases to ensure your team can obtain all the necessary information to complete the record. This will also provide an internal record of tools that have been considered during the design and development phases. For tools that do not reach the piloting or deployment stage, you are not expected to publish the record.

Records for tools in the pre-deployment phase will be kept internally but not published. Only records for tools in the production phase will be published in the repository.

If you are considering completing an algorithmic transparency record, we would encourage you to get in touch with the ATRS team at [algorithmic-transparency@dsit.gov.uk](mailto:algorithmic-transparency@dsit.gov.uk). We will be happy to provide advice on the suitability of algorithmic tools for the transparency record, and can answer any questions you have about the process. When you are nearing completion of the record, and we will also help to provide feedback and advice.

## **c. How should you complete and submit your algorithmic transparency record?**

Before writing an algorithmic transparency record, you should assign an algorithmic transparency lead to oversee the completion of the record. The lead will need to collate information from across and sometimes beyond the organisation. This could include those involved in the operational use of the tool, and the data science team or suppliers involved in the design and deployment of the tool. The algorithmic transparency record may also require clearance from data ethics, legal, communications, and senior leadership teams.

### ***Who should be the algorithmic transparency lead?***

The following set of questions should help you determine who should take on the role of lead and who else should be involved in the process of filling in the template:

- Who is initially responsible for the delivery of the project on an operational level?
- Who has access to most of the required information, or connections to the teams holding the required information?
- Who has the time and resources to complete the template?
- Who has the necessary expertise (technical/legal/other) to fill in the record?

In central government departments subject to the mandatory requirement of the ATRS, there will be a named single point of contact (SPOC) to lead the algorithmic transparency work within the department. This SPOC is an individual who is best placed to obtain information from across the department on all algorithmic tools that have a significant influence on a decision-making process with direct or indirect public effect; or directly interact with the general public. They will be responsible for mapping algorithmic tools that are in scope and the completion of accompanying Algorithmic Transparency Recording Standard (ATRS) records. This SPOC will generally have an oversight and coordination role rather than creating all records for an organisation.

### ***What if a supplier holds relevant information?***

If your supplier holds information that you need to complete a record, we encourage you to ask from your commercial contact for the relevant details, explaining why you are asking for this information and why algorithmic transparency is important in the public sector. If the supplier is reluctant to share some information with you based on concerns around potentially revealing intellectual property (IP), it can help to walk the supplier through the questions asked in the Standard, explain how they are designed to provide a high-level picture of the tool and not compromise IP.

### ***Suggested process for teams completing a record***

We recommend you adopt the following process when completing and submitting a record, which can be adapted to suit your organisational structure:

- Identify a tool which is in scope of the Algorithmic Transparency Recording Standard, referring to this guidance and the ATRS team where needed.
- Identify a lead to complete a record for the tool identified to be in scope of the Standard. You should also identify who will need to clear the record before it is published. If you have any questions about the process, please email [algorithmic-transparency@dsit.gov.uk](mailto:algorithmic-transparency@dsit.gov.uk).
- The lead uses [the template](#) to start completing the record. They then reach out to all teams/organisations who will be required to provide input into the template. Those involved agree on a plan and timeline for completion which accounts for any necessary clearance processes. The lead sets clear expectations about what information will be required from different teams/organisations. It is particularly important that these requirements are clearly communicated to any suppliers involved as soon as possible.<sup>i</sup>
- The lead manages the process of collecting and collating the required information. This could be through a workshop-style session with relevant teams/organisations, individual meetings with the teams/organisations, or a sharing the ATRS template for teams/organisations to input relevant information. Some of the required information may be available through alternative means; for example, through supplier records, impact assessments, or from an organisation's website. The lead should check with relevant teams/organisations that information gathered through other means is up to date and that they are happy for it to be included in the record.

- e. The lead should ensure that all information is added to the template and that all sections are complete. Most teams find that writing an algorithmic transparency recording using the Standard once they have the relevant information takes approximately 3-5 hours.
- f. The lead should clear the final record with the relevant individuals and teams.
- g. The lead should submit the final record to the ATRS team by sending it to [algorithmic-transparency@dsit.gov.uk](mailto:algorithmic-transparency@dsit.gov.uk) for upload to the repository.

#### **d. Exemptions: what information should not be published?**

Though transparency about how the public sector is using algorithmic tools is useful and appropriate in most circumstances, there is naturally some need for caution to ensure that information is not published that would be counter to the public interest.

The ATRS has been designed to minimise likely security or intellectual property risks that could arise from publications, and therefore situations where no information can be safely published are expected to be unusual (e.g. in cases where the existence of a tool cannot be made public). In general, publishing algorithmic transparency records and redacting some fields, with a brief explanation of why this has been done is preferable to not completing or publishing a record at all. Where a public sector body decides that it is not possible to make any parts of the record public, completing a transparency record without making it publicly available is highly recommended. Filling in an algorithmic transparency record is a valuable exercise for teams to properly think through the relevant risks, impacts, and accountabilities related to a tool.

Regarding the mandatory requirement for central government departments, the final exemptions policy is currently in development. We are working with partners to ensure it is fit for purpose and sufficient guidance will be available. Any updates or changes will be made to this page in due course, and clearly communicated to stakeholders. The FOI Act 2000 already has a well-developed framework for exemptions which will form the starting point for this work.

#### **e. Considerations before publishing**

##### ***i. Operational effectiveness and gaming***

Some use cases for algorithmic tools include identifying potential risky applications for a service or highlighting possible fraud. In such circumstances, providing too much information about how an algorithmic tool works, or the specifics of the datasets it draws on, could compromise the operational effectiveness of the tool. For example, a malicious user might modify their behaviour to avoid triggering a fraud warning.

In most cases, such issues can be managed by being careful about the level of detail provided in the algorithmic transparency record, especially around the technical design or data used. Wider information, for example on how the algorithmic tool is used in the overall decision-making process may still be safe to release and relevant.

##### ***ii. Cybersecurity risks***

Some types of information about the development and operation of your algorithmic tool could increase the vulnerability to a cyber-attack.

You should involve the appropriate individuals and teams (both within your organisation, and from any relevant third-party suppliers) in discussions about what details to include and how to communicate the information. The scope of these discussions could include the following:

- An assessment of which technical details about the system architecture are appropriate to include in order to satisfy the purpose of the record, and which specific details may pose an unnecessary risk and should be omitted.
- Appropriate steps to mitigate predictable forms of attack that could follow from releasing information about your algorithmic tool and organisation. For example, it may be possible to reduce the risk of a targeted phishing attack, and thereby consequent risks of unauthorised access to your algorithmic tool, by including a team email address, rather than that of a named individual, in the 'contact email' field.

Broadly speaking, obscurity is a weak cyber security defence, and if a tool is deployed in a way where the information defined in the Algorithmic Transparency Recording Standard presents a cyber security risk then it is highly likely that there are vulnerabilities that need addressing regardless of levels of transparency.

### ***iii. Intellectual property***

You should consider whether some of the information requested could infringe upon the intellectual property rights of your organisation or your third-party supplier. This may be raised by a supplier as a concern around providing the information in this record. We have designed and tested the ATRS to only request information at a general level that should not present risks to intellectual property. However, if you or your supplier are concerned, it may be worth checking relevant legal or commercial agreements and involving appropriate specialists.

### ***iv. Public scrutiny and communications***

Before completing a record, you should consider the possibility that publishing information on your algorithmic tool may invite attention and scrutiny from the public and media. This is particularly true for more high profile use cases and where the use of an algorithmic tool has not been publicly disclosed before.

You can help to mitigate these risks by ensuring you provide clear information and an appropriate level of detail in your record. You can also ensure that your senior leadership and communications teams approve the record before it is submitted to the ATRS team.

Your communications team should also be prepared to respond to media requests. You may want to consider publishing something on your organisation's website (for example, a blog post accompanying the release of your record, explaining what the tool is and what motivated you in publishing the algorithmic transparency record. The ATRS team (algorithmic-transparency@dsit.gov.uk) is happy to provide guidance on this.



## Step 2: Complete the algorithmic transparency record

### Tier 1 Information

In Tier 1, you should give a basic description of how the algorithmic tool functions and why it was introduced into the decision-making process.

We expect that the primary audience for information included in Tier 1 will be the general public and other interested parties looking for a summarised version of how the tool functions and the role it plays.

#### *How much detail should I include in Tier 1?*

Tier 1 asks for a basic description of the algorithmic tool aimed at the general public. For this reason, your answers to each question in Tier 1 should ideally be no more than a couple of sentences describing how and why the tool works.

Imagine you are describing the tool to a member of the public with only a basic understanding of what an algorithmic tool is. Looking at the information in Tier 1, a reader should be able to understand in general terms what the tool does, how it works, and how it fits into the wider decision-making process or wider public service.

#### *Why is it important to add a team email address rather than an individual's email address?*

Adding a team email address rather than the contact details of an individual is important for business continuity and security purposes. When an individual leaves the organisation but the wider team remains, the email address will still be up to date.

Disclaimer: This is a fictional example constructed to illustrate how the Standard could be applied. It is not based on any existing algorithmic system, nor is it intended to be a fully accurate representation of existing school admissions processes.

#### **Example:**

**1.1 - Name:** Y Council secondary school place allocation tool

**1.2 - Description:** This algorithmic tool helps the council assign secondary school places to individual children.

Where schools are oversubscribed, priority for places is determined by a set of published admissions criteria, with some criteria specific to individual schools. Applying these criteria consistently across all of the applications and schools in Y Council's area is a complex and labour intensive process. The council receives a large number of appeals each year, and occasionally these demonstrate that errors are made.

To make the process more efficient, and reduce the number of errors, the council has developed an algorithmic tool that assigns each child to a school. The algorithm automates the application of existing admissions criteria, including individual student preferences, geographical radius, special needs (e.g. due to special educational needs), etc. It will enable

schools to utilise a wider range of admissions criteria in future, for example replacing or supplementing current measures of geographic distance with travel time by walking, cycling and public transport in line with the council's healthy streets strategy.

**1.3 - URL of the website:** [www.ycouncil.gov.uk/residents/children-education-and-families/school-admissions](http://www.ycouncil.gov.uk/residents/children-education-and-families/school-admissions)

**1.4 - Contact email:** [school-admissions@ycouncil.co.uk](mailto:school-admissions@ycouncil.co.uk)

## **Tier 2 – Owner and Responsibility (2.1)**

In Tier 2, you should give more detailed information about the algorithmic tool.

While Tier 2 will also be accessible to the general public, we anticipate that the primary audience for Tier 2 will be informed and interested parties, such as civil society organisations, journalists, and other public sector organisations seeking to better understand the tools being used across the public sector.

The owner and responsibility section details information about accountability for the development and deployment of the tool. Providing this information is important because it helps people to understand who is accountable for the tool and its use, and how they can find out more information.

### ***How much detail should I include in this section? (2.1)***

This section should clearly convey information about the organisation, team and senior responsible owner (SRO) with responsibility for the tool. We would expect each field to contain around 1-3 bullet points of information, though you are welcome to provide more detail if you think it is helpful to include.

If the tool involved a third-party supplier, the information should provide a clear explanation of what the relationship with the supplier was, though we do not anticipate that this would consist of more than a couple of sentences.

### ***I am not sure who the senior responsible owner for the tool is. What should I put in this field? (2.1.3)***

The SRO should be the person who is accountable for the tool. This would usually be a different individual to the person who delivers the project at an operational level. It should ideally be someone with accountability for the use of the tool in an operational context, not just for the technical delivery. In your answer, you should specify the role title of the SRO.

### ***Multiple external suppliers have been involved in the delivery of the tool through a multi-layered supply chain. What information should I provide about this in the 'external supplier role' field? (2.1.4.3)***

A procured tool can entail the involvement of multiple companies at different places in the supply chain. For instance, a public body could procure a tool from a company, which in turn procured the model and data from another company before integrating the model into a customisable tool.

Ideally, you should describe those different supplier relationships as clearly and concisely as possible, detailing which organisation was or is responsible for which part of the final tool that you are deploying.

**Example:**

**2.1.1 - Organisation:** Y council

**2.1.2 - Team:** School Admissions Team (Education division)

**2.1.3 - Senior responsible owner:** Head of Education Division

**2.1.4 - External supplier involvement:** Yes

**2.1.4.1 - External supplier:** AI Tools UK

**2.1.4.2 – Companies House Number** 083827744

**2.1.4.3 - External supplier role:** The algorithmic tool was developed by AI Tools UK. Their experts have worked together with education policy experts at Y council to configure the tool and develop rules according to the requirements of the council.

**2.1.4.4 - Procurement procedure type:** Open

**2.1.4.5 - Terms of access to data for external suppliers:** AI Tools UK have been provided with controlled access to school admissions data from previous years to enable the system to be developed and configured. This has been done in compliance with data protection legislation and all AI Tools UK staff with access to the data have been subject to appropriate vetting checks. Access to the data is only granted for a limited period of time while the tool is developed. Ongoing maintenance and operation of the tool is carried out by council staff.

## **Tier 2 – Description and Rationale (2.2)**

In this section, you should provide more granular detail about the tool, including its scope and an expanded justification.

### ***What should be included in the detailed description field? (2.2.1)***

This field is optional and is to be a longer form version of the description provided in Tier 1. Information provided in this field may duplicate information provided in other fields below, however, it is an opportunity to provide an overall description of the tool and an account of the logic, rules and criteria used by the algorithm(s) and tool. A good rule of thumb here would be that, from reading this description, individuals should be able to understand how the tool relates to operational level decisions.

You are also encouraged to include technical details about how the tool works where appropriate.

The example included below may be helpful to give an idea of the level of detail we would typically expect to see in this section. You can also see how other teams have completed this section [here](#).

**Example:**

**2.2.1 - Detailed description:** N/A

**2.2.2 - Scope:** This algorithmic tool has been designed to apply admissions criteria to automatically assign a school space to each child in the main annual admissions round. The tool provides an initial allocation, but some individual circumstances will continue to be dealt with manually, including in-year admissions outside of the usual cycle.

The purpose of the place allocation algorithm is to ensure the allocation of school places is time- and labour-efficient, and applies the admissions criteria accurately and fairly.

Many of the admissions criteria supported are fact-based criteria (e.g. sibling preference, looked-after children). However, the assessment of travel time to school is based on a complex proprietary machine learning tool drawing on a range of mapping and public transport data.

**2.2.3 - Benefit:**

- Improve efficiency in allocation of school spaces (savings of 4 weeks' work for 5 full-time employees)
- Reduce error rate in allocation
- Enable potential future changes in admissions criteria to enable them to better take into account prospective students' choices and circumstances, resulting in a fairer outcome
- Improve the council's understanding of children's choices and circumstances

**2.2.4 - Previous process:** The current process for applying admissions criteria involves a mixture of manual and automated steps, managed through a complex set of internally developed spreadsheets. The process is highly reliant on the knowledge and experience of a single member of staff and is not sustainable in the long term.

The limitations of this process restrict the ability of the council to improve admissions criteria to better align to the needs of children and schools.

**2.2.5 - Alternatives considered:** Retaining the current approach was considered, but rejected because of the significant risk of failure of the current system in the event of staff changes. A range of technical options, and options for sets of admissions criteria that could be supported, were considered through the development process, in consultation with schools and representatives from parent groups

## **Tier 2 – Decision making process (2.3)**

In this section, you should provide further information about how the tool is integrated into the decision process as well as oversight mechanisms.

### ***How much detail should I include in this section? (2.3)***

This depends on the complexity of the algorithmic tool, but as a general guide we would expect teams to provide around 150 words written for a non-technical audience. This should cover the key elements of how the tool is integrated into decision making processes and oversight mechanisms.

### ***What information should I include in the ‘appeals and review’ field if this is not applicable to the tool I am developing or deploying? (2.3.6)***

If no appeals or review process is necessary for your tool, include a short sentence explaining why you are not completing this section in the relevant field.

You should also be aware of Article 22 UK GDPR which states that ‘The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her’.

If the algorithmic tool you are completing a transparency record for falls within scope of the provisions of Article 22, for example, because it informs decisions that are ‘solely based on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her’, then you must complete this section.

For further guidance on this issue, please refer to the ICO’s guidance on [‘Rights related to automated decision making, including profiling’](#).

#### **Example:**

**2.3.1 - Process integration:** Decisions around the allocation of school places are made on an annual basis by the council, working closely with schools. The algorithm replaces the previous system used for managing these allocations, adding a significantly greater degree of automation to existing processes. It makes a recommendation for the allocation of places. Prior to the allocation being finalised, the admissions team will carry out a manual review of the allocation, consulting with schools where necessary, prior to communication of outcomes to children.

**2.3.2 - Provided information:** The tool provides a recommendation for an allocation of children to available school places. It also provides an explanation to the case officer of how the different factors contributed to a student being allocated to the assigned school. This includes a map-based view of home addresses of successful and unsuccessful applications to enable sense checking of geographic distance and travel time allocations.

**2.3.3 - Human decisions and review:** The automated process is reviewed every year to check its performance. During each period of allocation, the school admissions team sense checks the allocations made by the software, and individual schools are also given an opportunity to review their allocations. After the sense check, the allocation is approved by the responsible director and the notifications to students are released. Human officers can override the recommendation made by the automated process. These decisions are also fed

into the automated process, alongside information on appeals by students and the outcome of appeals.

**2.3.4 - Frequency and scale of usage:** Each year, the automated process makes approx. 3000 decisions, equivalent to the number of students to be newly enrolled in school.

**2.3.5 - Required training:** Each officer using the tool goes through an onboarding process that trains them on how to use the tool and troubleshoot. Training is also provided for schools interacting directly with the tool.

**2.3.6 - Appeals and review:** Existing appeals processes for school admissions will continue to apply, as set out on the council website. Students and their parents can appeal a decision within two weeks of allocation, stating the reasons for their appeal. Appeals will be handled by human officers, but are only likely to be successful if they identify that the admissions criteria have not been applied correctly.

## Tier 2 – Technical Specification and Data (2.4)

This section is split into three, with a focus on a different technical component in each subsection. The three components we consider are:

1. **The algorithmic tool.** From a technical perspective, we can consider the algorithmic tools as systems of software and hardware components which embody complex algorithms and allow them to be deployed by non-expert users.
2. **The models:** Models are the products of algorithms applied to training data. They are constructed to help analyse, explain, predict, or control the properties of real-world systems. We can consider them as functions that perform a task by processing input data and returning outputs.
3. **The data:** Data is the information used to develop models. Often a learning algorithm is applied to data to train a machine learning model. Data may also be used to validate, test, or verify model performance.

The Model and Data specification sections (2.4.2 and 2.4.3) are modular. Where an algorithmic tool contains multiple interacting models, each model should be accounted for in the Tool specification section (2.4.1) and accompanied by its own specific model and data sections. In many cases, only one model and data section will be required.

In addition to the three subsections, we include an optional **dataset card** to explore the data component in more detail. A dataset card can be completed for individual datasets and attached to a record as a linked object. The Data specification section (2.4.3) can be used to aggregate the information in dataset cards, so completing a dataset card may be especially useful where a dataset has been used to develop multiple models and tools.

### Tool specification (2.4.1)

In this section, you should provide the technical specifications of the algorithmic tool.

#### ***What should I include in the ‘system architecture’ field? (2.4.1.1)***

The system architecture field should broadly describe how your tool is organised and provide context for the rest of the technical section. System representations such as AWS diagrams are ideal for conveying this type of information in a concise way – they capture the primary components of your technology stack and how they interact. You should think about the end-to-end process by which your tool processes inputs, the digital services and resources that it uses and the environments in which system processes occur. Any models that you consider later should be mentioned in this field.

You can see a helpful example of the diagram of system architecture provided by the Department for Health and Social Care in their algorithmic transparency report for the QCovid tool [here](#).

**Example:**

**2.4.1.1 – System architecture:** Find a system architecture diagram on our Github repository. Link: [www.github.com/ycouncil/admissions-tool](https://www.github.com/ycouncil/admissions-tool)

**2.4.1.2 – Phase:** Production

**2.4.1.3 – Maintenance:** As the tool is only being used once per year, maintenance and review of the tool occurs yearly prior to the operation of the tool. In the future, the logic and rules employed by the tool as admissions policies may evolve over time, and the travel time model is updated to reflect more recent travel data.

**2.4.1.4 – Models:** The tool features two models.

- i. Z-NET: a proprietary model that estimates travel time to school, trained on a range of transport and telecommunications data via a deep neural network.
- ii. Y-Admissions: a logic-based model that automates the application of admissions policy over a large set of applicants. The travel time estimates produced by Z-NET are one of the input features to this model.

***Model specification (2.4.2)***

In this section, you should provide information about a model that is used within the algorithmic tool. Please make and complete copies of this section for separate models.

***How is ‘model architecture’ (2.4.2.6) different to ‘system architecture’?***

Whereas the ‘system architecture’ field shows how a model sits within the technical apparatus of the tool, model architecture refers to the arrangement of parameters, formulas and rules that process an input into an output. At a minimum, you should enter the type of model used (e.g. random forest classifier, convolutional neural network, transformer, etc.). If it aids understanding of the model, you are also encouraged to provide further details or provide a link to publicly available resources that offer further information.

***What kind of metrics am I expected to detail in the ‘model performance’ field? (2.4.2.7)***



Performance metrics will differ based on what type of method and tool you are developing or deploying. It is up to you to choose which metrics are most appropriate and useful to your project. Useful metrics to consider are: accuracy metrics (such as precision, recall, F1 scores), metrics related to privacy, and metrics related to computational efficiency. You should also describe any bias and fairness evaluation you have undertaken (i.e. model performance over subgroups within the dataset), and any measures taken to address issues you identify.

For more information about setting performance metrics, you may find [this GOV.UK Service Manual guidance](#) helpful.

***How can I find more information about how to identify and mitigate bias in the data, model and output of the tool?***

For more information about bias in algorithmic decision-making, see the RTAU (formerly Centre for Data Ethics and Innovation) [review into bias in algorithmic decision-making](#), especially Chapters 1 and 2. For more information about how to mitigate bias in algorithmic decision making, you may find it helpful to review the RTAU's repository of bias mitigation techniques which can be found [here](#).

Example:

**2.4.2.1 – Model name:** Y-Admissions

**2.4.2.2 – Model version:** v4.1

**2.4.2.3 – Model task:** To allocate a fixed number of secondary school places to a set of applicants.

**2.4.2.4 – Model input:** The model requires two inputs:

- i. A dataset containing information about school applicants.
- ii. A dataset containing the number and types of places offered by local state schools.

**2.4.2.5 – Model output:** The model outputs a dataset with school allocations for every applicant, along with brief written explanations for why each allocation has been given.

**2.4.2.6 – Model architecture:** Y-Admissions is an optimisation-based automated planning model. The model consists of:

- A set of rules that dictate how a fixed number of school places are distributed across a population based on relevant variables.
- An ordered set of objectives that specify the goal conditions for allocation. These include:
  - Maximisation of preferred choices
  - Minimisation of total time spent travelling to school
  - Maximisation of sibling-matching

**2.4.2.7 – Model performance:** AI Tools UK – who work with Y Council to develop and maintain the admissions tool – conduct regular verification tests to ensure the Y-admissions model is executable (i.e. doesn't produce errors) and valid (i.e. satisfies its objectives).



We also evaluate the model on *demographic parity*: an analysis of how independent preferred choice allocation is with respect to sensitive attributes. The model must have a parity greater than 0.85 for sex and religion attributes – results below this threshold prompts us to reconfigure our model. Graphs showing demographic parity both before and since the introduction of the tool can be viewed [here](#).

More broadly, we collect performance metrics on the number of decisions overridden by a human and number of successful appeals/challenges to an allocation.

**2.4.2.8 – Datasets:** The model has been developed with two types of historic dataset:

- i. Datasets containing information about school applicants.
- ii. Datasets containing the number and types of places offered by local state schools.

**2.4.2.9 – Dataset purposes:** Both types of datasets are used to test the robustness and fairness of the model (see *Model Performance*).

### ***Data specification (2.4.3)***

In this section, you should provide information about the data that has been used to develop a model. Please make and complete copies of this section for the data associated with each model.

#### ***How much detail should I include in the ‘data description’ field? (2.4.3.3)***

You should provide a high-level overview of the range and nature of the data used to develop the model. It could include basic facts about the subjects of the data, the coverage of the data with respect to location and time, and who is or has been responsible for data management. Aim for around two sentences.

#### ***Why do you include a ‘data quantities’ field? (2.4.3.4)***

The purpose of the ‘data quantities’ field is to sense-check proportionality of data in relation to the model task and complexity. Where a learning algorithm is applied to data, small datasets with few samples are more likely to yield underfitting models, while large datasets with numerous attributes may cause overfitting. In addition, too few samples may indicate insufficient representation of a target population, and too many attributes may indicate increased data security risks (such as de-identification).

#### ***What do you consider ‘sensitive attributes’ to be? (2.4.3.5)***

While we don’t prescribe a specific definition of ‘sensitive’, we encourage you to disclose:

- **Personal data attributes:** “any information which are related to an identified or identified natural persons” as defined by [GDPR Art. 4\(1\)](#).
- **Protected characteristics:** any characteristics it is illegal to discriminate against as defined in the [Equality Act 2010](#).

- **Proxy variables:** any information that may be closely correlated to unobserved personal attributes or protected characteristics. E.g. frequency of certain words in an application may correlate to gender, birthplace may correlate to race.

In certain cases, it might not be feasible for you to disclose all the sensitive attributes in the data. In this case, at a minimum, you should disclose the fact that you are processing sensitive data and add as much detail as appropriate.

***I'm concerned that sharing information about the variables and potential proxies could lead to individuals being made identifiable. What should I do?***

It is unlikely that the transparency report would lead to individuals being made identifiable as you are only being asked to provide a general description of the types of variables being used. If you are considering making the dataset you are using openly accessible and linking to it, you should comply with the relevant data protection legislation to prevent individuals from being made identifiable from the dataset.

This should also be considered as part of a Data Protection Impact Assessment (DPIA). For further guidance on completing DPIAs, please refer to the [ICO's guidance](#).

***What other resources are available to support me with completing this section?***

You may find it helpful to consult the ICO's AI and data protection risk toolkit, which can be found [here](#).

**Example:**

**2.4.3.1 – Source data name:** Y Council school applicants

**2.4.3.2 – Data modality:** Tabular

**2.4.3.3 – Data description:** These data provide information about school applicants which is relevant to Y Council's admissions policy.

**2.4.3.4 – Data quantities:** Each year, the dataset contains approximately 3000 samples with around 40 attributes each, including the estimated time it takes to travel to multiple schools by various modes of transport.

**2.4.3.5 – Sensitive attributes:** The sensitive attributes of these data are:

- Name
- Age
- Address
- Sex
- Religion
- Disability
- Parent or guardian name(s)

**2.4.3.6 – Data completeness and representativeness:** Datasets are always ensured to be complete before being processed by the model – no missing values are permissible.

As each dataset comprises the entire population of Y Council school applicants each year, so the data is representative of the target population by definition.

**2.4.3.7 – Source data URL:** N/A - the data contains personal data and cannot be made public.

**2.4.3.8 – Data collection:** The original school applicant data was collected retrospectively from existing schools and the body of students that enrolled in the year prior to the development of the school space allocation software for purposes of developing the tool. Ongoing year-by-year data is being collected from students and schools 3 months prior to the allocation being performed.

**2.4.3.9 – Data cleaning:** After the data has been collected by Y Council, any pre-processing and cleaning is performed by AI Tools UK. This includes, for example, scanning and correcting for duplicate observations and missing data, fixing structural errors, such as spelling mistakes or differing naming conventions.

**2.4.3.10 – Data sharing agreements:** A data sharing agreement for this project has been put in place between Y Council and AI Tools UK.

**2.4.3.11 – Data access and storage:** AI Tools UK have been provided with access to limited amounts of the council's data to enable the system to be configured. This has been done in compliance with data protection legislation and all AI Tools UK staff with access to the data have been subject to appropriate vetting checks. Access to the data is only granted for a 3-month period while the tool is configured and operated. Otherwise, data access for yearly input data is restricted to the Y Council's education department and schools. Data is stored in identifiable format for 4 years after which point it is anonymised.

## **Tier 2 – Risks, Mitigations and Impact Assessments (2.5)**

In this section you should provide information on impact assessments conducted, identified risks, and mitigation efforts.

***Am I expected to write a summary of the completed impact assessment like the DPIA if I provide a link to the full assessment? (2.5.1)***

If you are providing an openly accessible link to the full assessment, you do not need to provide a summary.

***What do you mean by 'risks' and what are the most common risks you would expect to see described in this section? (2.5.2)***

In this field we are asking teams to consider possible risks that they think could arise in relation to the use of the algorithmic tool, and any actions or processes they have in place to mitigate against those risks. There are a range of possible risks that may arise, and the nature of risks will vary widely depending on the context, design and application of individual tools. The categories of risk likely to be relevant to the use of the algorithmic tool are:

- Risks relating to the data
- Risks relating to the application and use of the tool

- Risks relating to the algorithm, model or tool efficacy
- Risks relating to the outputs and decisions
- Organisational and corporate risks
- Risks relating to public engagement

Please note this list is not exhaustive and there may be additional categories of risks that are helpful to include.

### ***What other resources are available to support me with completing this section?***

For further guidance on conducting risk assessments for public sector projects, you may find it helpful to look at HM Treasury and the Government Finance Function's [Orange Book](#), which is updated regularly.

You may find it helpful to consult the ICO's AI and data protection risk toolkit, which can be found [here](#).

#### **Example:**

##### **2.5.1 - Impact assessment:**

Assessment title: Data Protection Impact Assessment

Short overview of impact assessment conducted:

See summary and full assessment under the link provided

Date completed: 10th December 2021

Link: [www.ycouncil.gov.uk/children-education-and-families/schoolspaceallocation/dpia/](http://www.ycouncil.gov.uk/children-education-and-families/schoolspaceallocation/dpia/)

Assessment title: Equality Impact Assessment

Short overview of impact assessment conducted:

See summary and full assessment under the link provided

Date completed: 15th December 2021

Link: [children-education-and-families/schoolspaceallocation/eia](#)

##### **2.5.2 - Risks:**

###### *Risks:*

The tool's computation of travel time might not always accurately reflect real world travel times for individual children.

###### *Mitigations:*

The council has set out a clear definition of how travel time is calculated on our website, so that applicants can understand clearly how this will apply to them.

Individuals can access estimated travel times to schools via the tool prior to

	submitting their application, and hence have an opportunity to raise any queries prior to school places being allocated.
The data and tool could be accessed by unauthorised users	The tool has robust access controls and only a small number of users have the ability to access this directly. Data access for both the training data and yearly input data is restricted to the Y Council's education department and school board. AI Tools UK have been provided with access to limited amounts of the council's data to enable the system to be configured. This has been done in compliance with data protection legislation and all AI Tools UK and council staff with access to the data have been subject to appropriate vetting checks.
Misuse of personal information	All data is being handled in compliance with data protection legislation. A <a href="#">Data Protection Impact Assessment</a> has been conducted and signed off by the council.

### Step 3: Upload and update your algorithmic transparency record

As noted in Step 1, if you have completed a record for a tool that is currently in the pre-deployment phase, these records should be kept internally in your organisation but will not be published in the repository.

Only transparency records for tools that are in pilot or production phases will be published on the repository.

Once you have completed an algorithmic transparency record and obtained the necessary approvals for it to be published, you should follow the following process to upload your record:

1. Send your completed record to the ATRS team at [algorithmic-transparency@dsit.gov.uk](mailto:algorithmic-transparency@dsit.gov.uk).
2. The record will be reviewed and quality assured by a member of the ATRS team, who will discuss any questions or suggested amendments with the team submitting the record. This is not an assurance exercise for the tool itself, it is to ensure accessible language that can be understood by a general audience has been used throughout and all applicable fields have been completed. Suggestions and queries will be made on the record template in the 'Reviewer comments' columns – discussion can also take place over email or calls until all amendments have been made.
3. The ATRS team will upload and publish the final version of the record onto the GOV.UK repository alongside other algorithmic transparency records.
4. Submitting teams are encouraged to take this opportunity to publish or link to the record on their own websites if possible. This will increase transparency by making the information available to those using the organisation's website as well as those looking at the GOV.UK repository. The ATRS team can provide a markdown file of the finished record if useful.

### Updating records

The content of an algorithmic transparency record may become outdated or not reflect the latest version of a given tool if the tool is updated or refined. We recommend that teams should review their records every few months to ensure the information is still up to date. Teams are also free to request updates be made to their published records at any time by contacting the ATRS team.

### Version Management

The most recent version of the record will be available on the GOV.UK repository. Previous dated versions will be available in the archive on GOV.UK.

If teams have uploaded an algorithmic transparency record to the repository and the tool in question is no longer in use, teams can contact the ATRS team to update the record accordingly. Such records will still be available in the archive on GOV.UK.

Where amendments are made to the Standard, teams will not be asked to complete updated versions of the Standard for tools they have already recorded on, unless they are required to update the existing record for other reasons - for example, if the tool has changed. The version of the Standard which has been completed will be noted on the completed record.

## Annex 1: Glossary of Key Terms

**Algorithm:** An algorithm is a set of step-by-step instructions. In artificial intelligence, the algorithm tells the machine how to find answers to a question or solutions to a problem.

**Algorithmic tool:** An algorithmic tool is a product, application, or device that supports or solves a specific problem, using complex algorithms. You can develop a tool in-house or buy from a third party.

**Bias:** In statistics, bias has a precise meaning, referring to a systematic skew in results, that is an output that is not correct on average with respect to the overall population being sampled. In general usage, bias is used to refer to an output that is not only skewed, but skewed in a way that is unfair. Bias can enter algorithmic decision-making systems in a number of ways. For more information, see the review by the Responsible Technology Adoption Unit (formally CDEI) on [Bias in algorithmic decision-making](#).

**Data:** Data is the information used to develop models. Often a learning algorithm is applied to data to train a machine learning model. Data may also be used to validate, test, or verify model performance.

**Model:** A model is the product of an algorithm applied to training data. They are constructed to help analyse, explain, predict, or control the properties of real-world systems. We can consider them as functions that perform a task by processing input data and returning outputs.

**System architecture:** A description of the overall structure of the algorithmic tool and its technical components, including key software and hardware.

---