Office for Statistics Regulation Guidance

Guidance for Models: Trustworthiness, Quality and Value.

Contents

Summary

The Office for Statistics Regulation (OSR) provides independent regulation of all official statistics produced in the UK. Statistics are an essential public asset: we aim to enhance public confidence in statistics produced by government by setting the standards they must meet in the Code of Practice for Statistics through the pillars of Trustworthiness, Quality, and Value (TQV).

This document provides guidance on how the principles in the Code of Practice can help in designing, developing and using models to improve their Trustworthiness, Quality and Value (TQV). Misuse and a lack of transparency of models can undermine public confidence, especially in the statistics they produce and decisions they inform. This guidance has been created to cover both traditional statistical techniques, (e.g linear regressions), and newer techniques (e.g machine learning), when they are used to create outputs that inform decision making and/or public policy. In this guidance, tick box statements have been included so that you can apply the principles in your work.

Part I explores the planning of a model. It provides steps to ensure your model meets the pillars of Trustworthiness, Quality and Value before you begin development. The main factors which should guide your decision to use a model are the purpose of the work, the user need and the social context surrounding the work. By thinking through these three things, you can demonstrate the appropriateness of your chosen technique.

This section discusses the following questions:

- What is the question you are trying to answer?
- What is the user need?
- What ethical and legal issues do you need to consider?
- Are the roles and responsibilities clear?
- Does your team have the right skills?
- Is resource sufficiently prioritised?

Part II focuses on the steps you should take to best develop and use your model to serve the public good. Part of this is ensuring that users of the model, both directly and indirectly, are at the heart of any decisions made around model usage.

This section discusses the following questions:

- Is there data of suitable quality?
- What is the right type of model?
- Are there opportunities for collaboration?
- Is the model clear and accessible?
- How will model quality and performance be measured?

All of the questions above explore important principles under the Trustworthiness, Quality and Value pillars. Thinking through these questions will enable you to demonstrate these pillars in your work as well as champion the themes of transparency, coherence and collaboration in a practical sense.

This is the final version of this guidance. It supersedes version one, which was published in October 2021. This new guidance reflects feedback we received from external stakeholders on version one, as well as OSR's recent work on Securing public confidence in algorithms **and our** QCOVID case study**. We would, however, still welcome feedback so please get in touch with** regulation@statistics.gov.uk **if you would like to share your thoughts and use cases.**

Introduction

The Office for Statistics Regulation (OSR) provides independent regulation of all official statistics produced in the UK. Statistics are an essential public asset: we aim to enhance public confidence in statistics produced by government by setting the standards they must meet in the Code of Practice and the pillars of Trustworthiness, Quality, and Value (TQV). In doing this, we have found the pillars to be useful in areas beyond official statistics production: for example. in production and use of administrative data. This guidance takes this idea and highlights the use of TQV in the design and production of models both within and beyond the statistics landscape.

Here, a model is defined as a tool used to create statistics, or to extract meaning from data for decision making. In this guidance, it primarily implies a data science or statistical model but could also be used to refer to an analytical model, conceptual model, mathematical model or data-driven model.

The term algorithm may also be appropriate in places. An explanation of terms can be found in Annex A.

As a model encompasses data, systems, and techniques, it is important that these components are not considered in isolation from one another. That is, considerations around data cannot exist without considerations of techniques; and neither exist independently of systems. It is this holistic perspective that the pillars of TQV enhance.

What do we mean by TQV of models?

Trustworthiness – having the confidence in those producing models that are to be used in the public domain

Quality – Data and methods that produce assured model outputs

Value – Models that support society’s needs for information

Who is this guidance for?

This guide is for anyone designing, developing, changing, reviewing or using models who wishes to uphold the principles of Trustworthiness, Quality and Value in their work. We also present some considerations for those thinking about using or changing models in the future. Users of this guidance may be those who are involved in:

- the design, creation or use of models to generate statistics/data used to inform decision making
- producing or using models to create new, exploratory statistics or testing new models for current statistics
- informing public policy using outputs from a model
- referencing data/statistics generated by a model in the public domain
- reviewing and validating models used to generate statistics or data used to inform decision making

What is the aim of this guidance?

The three Code of Practice pillars – Trustworthiness, Quality and Value (TQV) – provide an excellent framework for models used to produce official statistics and beyond. We have sought to highlight areas where TQV can be useful when building models. If you are interested in making a public commitment to TQV please consider Voluntary Adoption of the Code of Practice. Voluntary application (VA) of the Code is for any producer of data, statistics and analysis which are not official statistics, whether inside government or beyond, to help them produce analytical outputs that are high quality, useful for supporting decisions, and well respected.

Why are we producing this now?

Traditional statistical techniques, such as linear regressions, have long been used to create statistics or generate data used to inform decisions. However, recently, newer techniques including machine learning are being used to inform statistics production (for example, Using traffic camera images to derive an indicator of busyness: experimental research, ONS) and to inform decisions (for example, How the Ministry of Justice used AI to compare prison reports, MoJ). These statistics and decisions have impacts on society and it's led us to think about public confidence in this space.

There have also been high-profile cases of models being used in decision-making within the public sector with mixed public acceptability. This was highlighted in our report on awarding GCSE, AS, A level, advanced extension awards and extended project qualifications in summer 2020. Within this report we identified 41 lessons, aimed at Public Bodies, to consider when building models in the future and we tested these lessons on a model which was more widely accepted by the public, the model used to predict who should be added to the Covid shielding list, also known as QCOVID. The lessons held up well and have been incorporated into this version of the models guidance (beta version).

What does this guidance not cover?

This is not a regulatory document for data science, machine learning or artificial intelligence (AI)

The aim of this document is not to say that we, at OSR, will be regulating all AI tools. Effective, and appropriate regulation of these areas requires cross-sector agreement of which OSR is part of and of which is still in progress. This guidance is a framework for how to think of and instil Trustworthiness, Quality and Value into your work on models.

This is not technical guidance for designing models

We do not provide technical or methodological guidance since we are not a technical organisation nor is this part of our role. If you require technical support, you should contact relevant organisations, some of which are listed in Annex B.

How to use this guidance

The best time to use this guidance is before you plan to develop or use a model: it focuses on instilling public confidence and trust, which need to be considered from the very start to ensure success. This guidance can also be used retrospectively to understand how to strengthen TQV in your model. The guidance is split into two parts: planning a model that serves the public good, and developing and using a model that serves the public good. To help you think about how your model can meet the highest standards of TQV, we have created tick box statements (☐) throughout this guidance which can also be found in a separate, downloadable guidance sheet.

This is an updated version of this guidance building on recent work we have done in this space on Securing public confidence in algorithms and our QCOVID case study. We would, however, still welcome feedback so please get in touch with regulation@statistics.gov.uk **if you would like to provide feedback.**

Part I: Planning a model that serves the public good

Models present many opportunities for developing and improving the production of statistics and data, while also assisting decision making and predictive modelling. However, there are risks associated with introducing a new model or making changes to a current one. Whichever you are doing, it's important not to do so without full consideration of the impacts, both the immediate impacts (from the model itself) and the impacts further down the chain (e.g if the outputs are used in other models).

Before developing, using, or changing a model, you should take the necessary steps to appropriately plan and carefully consider whether your planned approach is right for its intended use.

Take time to consider the following:

What is the question you are trying to answer?

Firstly, it's important to understand the question you are trying to answer and what you hope the product will be. This will allow you to go out and collect the resources you need to be able to answer the question effectively. Starting with the question helps to keep focus on the aim as you move through model planning and development and make sure you are providing the value that you intended. On the flip side, starting with the resources (e.g. data) and trying to get them to 'fit' to the product that you want is likely not to provide the value you hoped. A model might not always be the best way to answer a question. The following questions can help your thinking:

- what are your motivations for building or changing a model?
- can your intended approach directly meet the aim of the work?
- are there alternatives that also need to be considered? (e.g. is a model the only way to answer the question?)
- are you adding value to the topic area?
- what is the deadline of the project? Do you have time to consider all aspects of TQV before your model needs to be deployed?

After considering the above, your aims should be stated upfront and regularly assessed throughout its development. When you can foresee tension between the aims and potential uncontrollable factors (e.g. time pressures), be clear that those tensions exist, what those tensions are and what impacts they may have.

Checklist

☐ After considering your question and desired product, the chosen approach is appropriate and warranted.

☐ Time has been put aside to regularly assess aims and return to your question.

☐ Any foreseen tensions between aims and uncontrollable factors have been clearly communicated and are expected.

Pillars in Practice: Meeting the diverse needs of users

Let's say you are thinking of building a new model to predict house prices for the month ahead. Your model replaces an old process. There are 3 types of users for your model:

1. An analyst wishing to use the model for an innovative housing project.
2. A statistics producer who is reliant on your model's outputs to feed into their statistics
3. A homeowner who your output will affect

All these users will be looking at the model from a different lens and will have different requirements, concerns and thoughts. The analyst is likely be keen for the model to be developed as there is little risk and a lot of benefit to them, however the statistics producer is likely to be more cautious and may have more specific requirements that will allow the model to work for their statistics. The homeowner may have concern around the lack of certainty of a new approach and what that means to them when selling their house. Each users' thoughts are valid and missing out engagement on even one user type could leave your model vulnerable to mistrust.

What is the user need?

Users should be at the heart of any decision to develop a model or change a process they rely on by introducing one. Models can provide opportunities to address user needs in ways which may not have been possible before such as helping to fill data gaps and, if used in a current process, can also mean an increase in quality for users.

However, it should never be assumed that a model is the best approach and engagement with a wide range of users should always be sought before any decision is made. After this engagement, you should be open and transparent about your intended approach and clearly communicate any benefits and potential risks or trade-offs associated with your model. It is important to be realistic in model expectations when speaking to the user, especially about the scope and ability you expect the model to have.

Checklist

☐ A range of users have been engaged with and their requirements have been considered.

What ethical and legal issues do you need to consider?

This section has been written in collaboration with our colleagues in the UK Statistics Authority's (UKSA) Centre for Applied Data Ethics.

Issues of data quality, methodological limitations, transparency, bias, and fairness all relate to ethical issues that should be carefully considered during project design. There are also legal considerations, particularly around data protection. Although we, at OSR, don't create ethical or legal guidance ourselves we would like to emphasise how important these considerations are and signpost readers to relevant additional material.

The UK Statistics Authority ethical principles identify six key areas that should be explicitly considered to enable the ethical use of data. Models are no exception. These principles include public good, transparency, methods and quality, confidentiality and data security, legal compliance, and public views and engagement. When considering the use of a model, these ethical principles should be considered to ensure that the use of such techniques is ethically appropriate. An ethics self-assessment form can also be completed to assist in this process.

There are also important legal considerations when using data to build models. Depending on whether you are a data controller or a data processor, you will need to consider what context you are using the data in. For example, under the General Data Protection Regulation (GDPR), data controllers must consider the rights of data subjects which may include ensuring transparency of models built using their data. Data processors can only process the data as allowed by the data controller and therefore when exploring model capabilities it is important to not go beyond this. The Information Commissioners Office (ICO) has some useful guidance on AI and data protection.

Finally, fairness for all is a major ethical consideration which we'd like to highlight specifically. No model should be built so that the output is in any way unfair or discriminatory to a certain group or individual. This is where human oversight becomes essential as fairness as a concept is not something a model is capable of identifying itself.

Checklist

☐ For models using personal data. The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately. No data subject should be unfairly disadvantaged by the model.

☐The model and data used in the model are consistent with legal requirements such as Data Protection Legislation, the Human Rights Act 1998, the Statistics Registration and Service Act 2007 and the common law duty of confidence.

☐ When working with data you are clear what legal role you have in its use and comply with the requirements under the GDPR.

☐ Safeguards have been put in place to allow checks for model fairness.

Are the roles and responsibilities clear?

Within statistics the Chief Statistician, Head of Profession, or those with equivalent responsibility, should have sole authority for deciding on methods used for published statistics in their organisation. This chain of accountability should be extended to models so it is clear who is responsible for what aspect of its development and upkeep and who will be ultimately responsible for errors of judgement. This is particularly important for models where accountability of processes and outputs is necessary for data protection and in gaining trust.

An appropriately chosen individual (Chief Statistician, Head of Profession, or other 'accountable officer') should be responsible for any statistics, output or data created by the model.

This means the accountable officer is aware of the methodological choices that have been taken in designing the model and those that sit underneath them understand their role in making design decisions clear to the accountable officer. They should also be aware of the methods used for validating the approach and any potential limitations and risks of the model. A chain of model accountability allows anyone involved in the project to know who to go to if something goes wrong or an error is identified. Should any harm be caused by the outcome of the model, the chain of accountability helps the accountable officer identify the source of the issue and who is legally responsible for the consequences of such harm.

When a model is developed by an external partner, the chain of accountability should be established before model development is agreed. This scenario can make accountability less clear as the team developing the model may be different to the team implementing and/or maintaining the model. The developing team have a responsibility to effectively communicate all aspects of the model to the implementing team with particular emphasis on what the model should be used for, its risks and limitations. The implementing team have the responsibility to question and fully understand those methods, risks and limitations for the context in which they are implementing the model. Furthermore, when the relationship between the developing and implementing teams has ended, it is important the implementing team know enough and are skilled enough to maintain and monitor the model.

Contingency should always be built into the chain of accountability whereby if one team member leaves, another member understands enough about the model to take on their role. This is to avoid a situation where the team does not have the relevant skills to continue using the model for its intended purpose. This would have a detrimental impact on the trustworthiness and quality of the outputs and should be a considered when planning and designing a model.

For more information on roles and responsibilities in a project lifecycle please see HM Treasury's Aqua Book.

Checklist

☐ A clear chain of model accountability has been established with clear roles and responsibilities at every level, including the role of the senior leader.

If the developing and implementing teams are the same:

☐ There is clarity of roles and sufficient resilience and skill within the team to build and manage the ongoing maintenance of the model

If the developing and implementing teams are different:

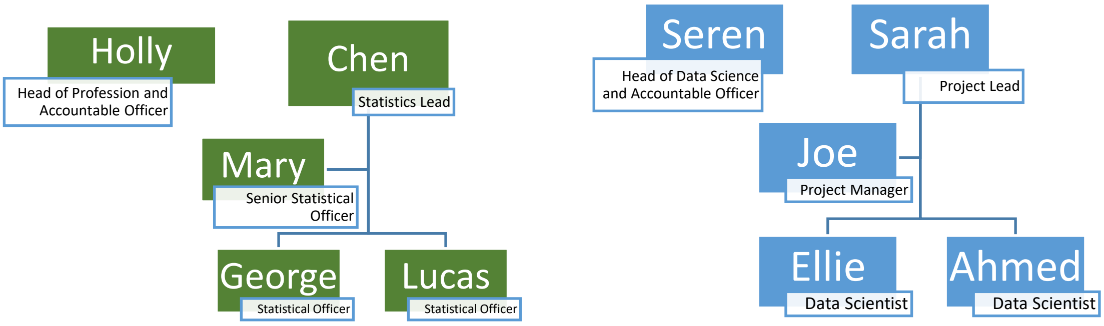
☐ Handover of knowledge, risks and limitations is given full priority when it is needed, and the skills needed to maintain the model are considered prior to handover.

Pillars in Practice: Building strong links in the chain

A statistics production team (Team A) asks a data science team (Team B) to build them a model to track the cost of leisure activities over the next 5 years. They intend to use the model to feed into their statistical output.

In this case, Team A is the implementing team and Team B is the developing team. You can see how they are structured below.

Team A *Team B*



Ellie and Ahmed design the model and feed their designs to Sarah for feedback. They then develop the model and feed their progress back up the chain to Joe and Sarah. Mary has been assigned as liaison for Team A so Joe regularly feeds back progress on the model to Mary. When the model is designed, Team B feedback their design decisions to Seren, the accountable officer, who can challenge the model design and assumptions. When Team B is ready to handover the model to Team A, all members of both teams meet to ensure all information is passed to the right people and training is provided by Team B. Holly is also able to challenge and understand decisions made in model design and development. The model is deployed by Team A but a dialogue remains between Mary and Joe to ensure knowledge stays up to date

This is one example of how two teams can work together to build trust and transparency when building a model.

Does your team have the right skills?

An important consideration in adopting any model is to assess whether the responsible team is sufficiently skilled. Some data science modelling techniques, in particular machine learning and artificial intelligence (AI), require a different skillset from traditional modelling techniques such as linear regressions. Likewise, knowledge of the skills required for more traditional modelling techniques are needed to avoid the potentially unnecessary use of more complex techniques. It is important to consider whether there is currently enough experience and knowledge in your team to ensure the expected standards are met. If not, consider investing time to train your team as it helps to develop skills and promotes innovation and improvement to current practices beyond just the model itself. Investment in the team also helps boost morale and contributes to career paths. If this isn't possible then specialist resource may need to be brought in from outside your team.

It is important to note that the team must have the appropriate knowledge and skills to manage both the implementation and maintenance of the model. In some cases, it may be advantageous to use external resource to build a model but if there is not enough knowledge and expertise to maintain the model (e.g. updating outdated data) then the model is unlikely to provide long term quality outputs.

Individuals may want to consider joining a relevant skills related profession or seek out accreditation to further build trust in their skills and knowledge.

Checklist

The team:

☐ has the knowledge and skills needed to develop the model (if appropriate).

☐ has the knowledge and skills needed to manage and maintain the model.

☐ has the knowledge and skills needed to successfully communicate the model and its outputs.

☐ has access to continuous development and learning.

☐ The benefits of building capability beyond the current work have been considered.

Is resource sufficiently prioritised?

Building a model can be resource intensive especially tasks such as data cleaning and model testing but this upfront allocation of resource can also lead to efficiencies in the longer term.

What's important is that any resource required to develop and maintain a model should be proportionate to the benefits and value arising from its use.

Models very often require not just upfront resource to build but also ongoing resource to maintain which is regularly underestimated when calculating resource needed. If this is the case with your model, it's necessary to consult your organisation's overall development plan and whether you believe these types of models will continue to be prioritised into the future. Implementing a model which does not have the resource to be maintained correctly could, depending on its use case, lead to more harm than good.

If you need to take resource from elsewhere, it is important to be open and clear about your plans and any re-prioritisation should be made clear to users and stakeholders.

Checklist

☐ The amount of resource required to implement the model has been estimated.

☐ The benefits outweigh the investment in resources to set it up and maintain.

☐ The required resource can be allocated to meet the project aims.

☐ Any reprioritisation required has been communicated.

Part II: Developing and using a model that serves the public good

When you have successfully planned your approach and are confident in your model's aims you should consider how best to develop and use that model to serve the public good.

While developing your model, collaboration and communication are key to demonstrating Trustworthiness, Quality and Value and you should consider these themes when thinking about the following questions:

Is there data of suitable quality?

Data is an essential part of any model and the focus **before** model building should be the quality and transparency around the underlying data. Getting data to a suitable quality and format is often the most timely step in model development but it shouldn't be rushed.

Models are only as good as the data that is fed into them and in the case of machine learning the model builds itself almost entirely on it; where there is bias in the data the model will learn this, where data is missing the model will assume this has meaning (and not simply missing data). Assuring quality of the data, therefore, is vital in any case but particularly so for these types of models.

Transparency around quality is just as important as assuring it. All parties need to be aware how the data will be used, the level of quality needed and where there are limitations, inequalities or biases ensure these are made clear. This transparency helps users to make informed decisions over its use.

Checklist

☐ It is known where the data came from, how it was collected and any limitations it has.

☐ Data suppliers and operational staff know how the data will be used and the level of quality required.

☐ Methods used to clean data have been made clear.

☐ Any limitations and inequalities that exist in the data are clearly communicated and their implications discussed.

What is the right type of model?

Choosing a suitable model is also important as different models are appropriate depending on the nature of the data, the type of problem and level of explainability needed. All these things should be considered and the thought processes behind the choice communicated clearly for those who wish to understand the rationale.

To find technical guidance on what model is applicable for your data and problem please consult 'A guide to using AI in the public sector' particularly the section named 'Assessing if artificial intelligence is the right solution'.

Public acceptability is related to how explainable your model is and the level of explainability should be appropriate for the context and purpose for which it will be used. Ask yourself the following questions:

- Will there be an expectation that users will want to know how the decision was reached or statistic produced?
- Will it impact decisions made about them?
- Will the outcome of the model provide such a public good that there is public acceptance that the model cannot be explained?

The last point makes reference to the fact that some of the best performing models can also be the least explainable so it may be relevant to think about priority with users.

What is the difference between explainability and interpretability?

Explainability is being able to work through the model at every stage and understand how it has come to deliver its output. Interpretability focuses more on the assurances that can be made that the model does what you think it does.

Full explainability of a model is easier for traditional statistical models such as a linear regression but often more difficult when you have a model that is building itself based on the data (e.g. machine learning models). Machine learning techniques to identify relationships and patterns in data at scale that are not easily detectable by humans which can make it difficult to describe how a machine learning model has reached a decision. With that said, you should always be able to communicate assumptions built into the model, any known biases and the uncertainty inherent in its outputs.

For some models, the sophistication of the models learned behaviour means that it may be impossible for you to understand how an outcome has been reached. These models are also known as 'black boxes' or opaque models. Black box models do not allow for transparency of model decisions. Without appropriate steps being taken to ensure they are interpretable, black box models could damage the trustworthiness of the statistics produced or decisions made, and you should be sure that a simpler, more explainable model could not have been used instead.

If a model is deemed to be difficult to explain, then it should be interpretable to meet the transparency requirements of TQV. This means having stringent quality assurance processes that can satisfy those who are accountable for the model that even though it can't be fully explained it is still behaving in an expected way given its inputs. Some examples of these quality assurance processes can be found in 'How will model quality and performance be measured?'

Checklist

☐ Reasons for model selection have been made clear.

☐ Any limitations and inequalities that exist in the model design are clearly communicated and their implications discussed.

☐ The users' needs for model explainability are known.

☐ It is known how the model is reaching its outcome or decision, and the result is reproducible.

Or

☐ If the model cannot be fully explained, the model is interpretable and fully quality assured (see ‘How will model quality and performance be measured?’).

☐ It is clear how changes to the inputs affect the outputs.

☐ It is clear to users that the model cannot be fully explained and reasons have been given as to why.

Pillars in Practice: Explainability vs Interpretability

Let’s say you have two hypothetical models in development:

Model 1: Uses data about household energy use to give each household a rating of energy efficiency which will provide them with advice on how to increase it.

Model 2: Uses personal health data to give a person a health rating which will affect their access to some health services.

For Model 1, you speak to households affected, some ask how the results will be used and when you explain it will only be used for giving advice they seem happy to accept the model without explanation. However, your senior manager still needs to know the model is fit for purpose and isn’t going to affect public trust in your organisation (a brand new home should not be expecting a low rating!). You decide to use a machine learning model but ensure one way you test the model is by changing the inputs many times and checking the output is as expected (interpretability).

For Model 2, you speak to members of the public and find out very quickly that this is a sensitive topic. You receive lots of questions around what data is involved and how the model will come to its conclusion. You know your model will need to be fully explainable to all types of audiences in addition to all quality assurance processes to gain public trust.

Are there opportunities for collaboration?

Collaboration can take many forms such as asking advice from domain and modelling experts to approaching teams who may have built something similar before.

You should collaborate with experts in both the type of model being used or developed and the subject matter which the data concern. You could do this by setting up a steering group of subject matter experts or run a workshop to gain feedback on specific aspects of your models design. This is to ensure any new insights drawn from the model are aligned with the experts’ understanding and the right type of model is deployed for the type of problem. It should also help you identify potential errors or bias that might already be built into a model's design while also offering opportunity for external and independent challenge.

By exposing your model in such a way, it could also offer opportunity to re-use or re-purpose those that already exist. There is lots of work being done on modelling and data access across government and beyond that will enable easier data and method sharing. It is possible someone else has already provided a solution to your problem or modelled something similar. The Central Digital and Data Office (CDDO) have recently launched the Transparency Data Standard which collects information on how government use algorithmic tools and the Integrated Data Service (IDS) launching at the end of 2022 aims to streamline data access and linkage for researchers. There are also government data science forums which can be used to get advice and track current projects.

Checklist

☐ Methods have been exposed to a wide professional audience to ensure appropriateness and opportunity for independent challenge.

☐ Similar models have been sought and/or considered to avoid duplication of effort.

Is the model clear and accessible?

This can be split into two main considerations:

Can those not involved in model development understand how it works?

Effort should be made to communicate with users in a way that is meaningful to them.

Explanatory material should exist and make clear why the model was designed, for what purpose it should be used and for what purpose it should not be used. It should also make clear the involvement of expert steering groups or relevant stakeholders in the development of the model. Full technical explanation should always be produced for those who wish to understand the technical detail and targeted at a more expert group. However, technical documents may not be appropriate for all users and therefore should not be the only method of communication. Consider making use of analogies and visual aids to communicate the model process while still providing access to the essential information. You should also ensure any code is fully annotated so a user can follow what it is doing. Best practice in this area is to use the principles of Reproducible Analytical Pipelines (RAP) which increase reproducibility and auditability which help bring trustworthiness in outputs.

When using a model in place of another process, communicate the strengths and limitations of the usual process so that all audiences can understand usual level of uncertainty. Please consult the Aqua Book chapter 5 ‘The importance and implications of uncertainty’ for more information.

It should always be good practice to make explanatory material available with the most assistive technologies.

Checklist

☐The documentation is sufficient to allow all types of users to understand the model and statistics or data produced.

Has access to all been considered?

Equality of access means access that is available to all and can come under four categories with regards to models:

1. Data used in the model

2. Outputs generated by a model
3. The code used to build the model
4. Supporting documentation

The default should be that all the above would be made available to all users, however, we understand the appropriateness of this can vary greatly depending on the context. If any of the above is deemed not appropriate to open and equal access this should be made clear, and an explanation given.

Open access platforms such as GitHub are recommended to make your model available and open to feedback and improvement. Platforms such as this make your model findable, accessible, inter-operable, and re-usable and give the option of public access to increase transparency of your approach.

You should also provide contact information for users who need to raise concerns or ask questions regarding accessibility.

Checklist

☐The model code, data and documentation have been made available and accessible to all (where appropriate).

☐ If the model code and/or outputs cannot be made available and accessible to all, an explanation has been given as to why.

How will model quality and performance be measured?

Quality assurances provide evidence that the model is fit for purpose and generating outputs that can be trusted but this doesn't mean that they will be. It may be tempting to think that once the technical processes required to quality assure a model (e.g. producing a performance metric) are completed the model will automatically be trusted by its users but technical assurances alone ignore the social elements. It is the combination of the technical and social elements that help ensure trust.

What this means in practice is that model performance and testing is carried out in a way that takes full account of its intended use and impact. For models that have a direct impact on individuals or groups this means testing the model for those groups and individuals specifically and checking whether any bias exists towards groups. The testing data should be previously unseen by the model and the quality criteria used to determine performance should be clear and chosen to reflect the problem at hand.

With that said, the social impacts of a model can only really be considered by humans. Human scrutiny is important, and it is advised that you take a diverse sub-set of individual outputs and manually check the route the model has taken to its output. There should be provision built-in for outputs/decisions made by the model to be challenged and time should be given to this process **before** a model's output is finalised. This final process builds trust by promising exposure of the model and allowing time for corrections and/or explanation before any significant impact is realised.

Please also consult the Aqua Book chapter 6 'Verification and validation' for further information.

Checklist

☐ Quality criteria used is clear and suitable to test model performance.
☐ The quality assurance process has been fully explained (including data used).
☐ The model has been assessed against the groupings that will be affected by the output and those of interest to the public.
☐ There is sufficient human oversight to consider the risks and impacts of the outputs from a social perspective.
☐ It is clear how the output from the model can be challenged and time is allowed for this process to happen.

Pillars in Practice: Making time to test and challenge

In models which feed into or make decisions relating to many entities it is very difficult to test every outcome and even harder to foresee every impact. Take the following hypothetical scenarios:

- A model which feeds into a UK wide statistical output (e.g. inflation) and that output feeds into 4 other statistical outputs.
- A model that gives a score for a child's reading age in England. This score affects a child's class placement which impacts educational outcome.

For both scenarios, even if quality assurance was thorough, the impact on every user of the statistics and on every child will only be fully apparent after roll-out. This is because 1) there are too many output entities to test and 2) there are too many interdependencies to test. It is recommended that in both scenarios a trial phase is implemented to allow the impacts on every user to be fully considered and any challenge to the model to be heard.

References

Office for Statistics Regulation, 2018. Code of Practice for Statistics

Data in Government Machine Learning Blog

Office for Statistics Regulation, 2021. Ensuring statistical models command public confidence.

Office for Statistics Regulation, 2022. QCovid® case study: Lessons in commanding public confidence in models

Office for Statistics Regulation, 2015. Quality Assurance of Administrative Data

Office for Statistics Regulation. Guide to Voluntary Application of The Code

Office for National Statistics, 2020. Using traffic camera images to derive an indicator of busyness: experimental research

Ministry of Justice (MoJ), 2019. How the Ministry of Justice used AI to compare prison reports.

UKSA, Ethics Self Reporting Tool

UKSA, Ethical Principles

UKSA, Statistics for Public Good

Information Commissioners Office (ICO). Guidance on AI and Data Protection.

The Data Protection Act.

UK Data Service. Importance of managing and sharing data.

HM Treasury, 2015. The Aqua Book: guidance on producing quality analysis for government

Central Digital and Data Office, 2019, Office for Artificial Intelligence, A guide to using artificial intelligence in the public sector

Central Digital and Data Office and Office for Artificial Intelligence, 2019. Assessing if artificial intelligence is the right solution.

Central Digital and Data Office, 2021. Algorithmic transparency data standard.

Government Analysis Function. Reproducible Analytical Pipelines.

Acknowledgements

This guidance was written by Emily Barrington, Elise Baseley, and Michael Hodge of the Office for Statistics Regulation. We extend our thanks to everyone who helped us to curate material for this guidance as well as provide detailed and helpful comments. Special thanks go to the following individuals who provided comments on this version of the guidance:

- Alec Waterhouse
- Anne Trefethen
- Arthur Turrell
- Catherine Seale
- Professor David Hand
- Sir David Spiegelhalter
- Louisa Blackwell
- Sarah Livermore
- Sonia Williams
- Tim Knight
- Shakeel Khan
- Simon Whitworth

As well as the following organisations who we have spoken to during this version of the guidance:

- The Alan Turing Institute
- Cabinet Office
- Centre for Applied Data Ethics, UK Statistics Authority
- Data Science Campus, ONS
- DEFRA
- Department for Business, Energy & Industrial Strategy
- Department for Work and Pensions
- HM Treasury
- Office for National Statistics

- United Kingdom Hydrographic Office
- University of Oxford
- Validate AI

Finally, thanks to our colleagues in the Office for Statistics Regulation who helped shape this guidance.

Annex A: Glossary

Here is a glossary of terms used in this guidance. This glossary includes terms referred in the Cabinet Office's Data Ethics Framework: glossary and methodology and the Code of Practice for Statistics.

Algorithm

A set of step-by-step instructions. Computer algorithms can be simple (if it's 3pm, send a reminder) or complex (identify pedestrians).

(Source: Matthew Hutson (2017) 'AI Glossary: Artificial Intelligence in so many words')

Artificial Intelligence (AI)

AI can be defined as the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence. AI is constantly evolving, but generally it:

- involves machines using statistics to find patterns in large amounts of data
- is the ability to perform repetitive tasks with data without the need for constant human guidance

(Source: GDS, OAI (2019) 'A guide to using artificial intelligence in the public sector')

Code of Practice

Shortened version of the Code of Practice for Statistics. Not to be confused with computer 'code', which relates to a collection of instructions that can be executed by a computer to perform a specific task.

(Source: Code of Practice for Statistics)

Data

In general, data can be understood as discrete values and statistics collected together for reference or analysis.

When we refer to data, we mean both data about people generated through their interactions with services, and also data about systems and infrastructure such as businesses and public services. Data can be operational (collected in the process of running services or businesses), as well as analytical and statistical.

(Source: National Data Strategy 2020).

Personal data means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

(Source: ICO, GDPR).

Data Science

Data science describes analysis using automated methods to extract knowledge from data. It covers a range of techniques, from finding patterns in data using traditional analytics to making predictions with machine learning.

(Source: Data Ethics Framework 2018).

Machine Learning

Machine learning is a subset of AI, and refers to the development of digital systems that improve their performance on a given task over time through experience. Machine learning is the most widely-used form of AI, and has contributed to innovations like self-driving cars, speech recognition and machine translation.

(Source: GDS, OAI (2019) 'A guide to using artificial intelligence in the public sector').

Model

A model, in computing terms is a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

(Source: US Department of Defence).

The word covers a broad range of uses. It may be used to refer to an entire system, such as a statistical model, mathematical model or conceptual model, or a component part of an AI system trained on a set of data, i.e. an AI model.

Official Statistics

Statistics produced by crown bodies, those acting on behalf of crown bodies, or those specified in statutory orders, as defined in section 6 of the Statistics and Registration Service Act 2007.

(Source: Code of Practice for Statistics)

Public Good

Defined in the Statistics and Registration Service Act 2007 in terms of the Authority's statutory objective to promote and safeguard the production and publication of official statistics that serve the public good. This includes informing the public about social and economic matters; assisting in the development and evaluation of public policy; and regulating quality and publicly challenging the misuse of statistics.

(Source: Code of Practice for Statistics)

Statistics

A collection of measures about a particular attribute compiled from a set of data. Statistics are used for making generalisations or inferring conclusions about particular attributes, at an aggregate level, for example, about a particular subset of the population.

(Source: Code of Practice for Statistics)

Traditional statistical techniques

A set of techniques that are considered traditional/classic in the field of statistics. Techniques include: the univariate t-test, ANOVA, ANCOVA, Pearson's r, Spearman rho, phi, and multiple regression; the multivariate Hotelling's T2, MANOVA, MANCOVA, descriptive discriminant analysis, and canonical correlation analysis.

(Source: Bruce Thompson (2013) 'Overview of Traditional/Classical Statistical Approaches')

Annex B: Relevant Organisations for Technical Guidance

One of our recommendations in Ensuring statistical models command public confidence is that "the cross-government Analysis and Digital functions, supported by the Centre for Data Ethics and Innovation should work together, and in collaboration with others, to create a comprehensive directory of guidance for Government bodies that are deploying these tools." This will undoubtably help those who use models to create statistics or inform decision making. Some relevant organisations for technical guidance on models may be:

- for guidance on best practice for data protection-compliant AI visit the Information Commissioner's Officer (ICO) Guidance on AI and data protection
- for general guidance on artificial intelligence please contact the Ada Lovelace Institute or the Office for Artificial Intelligence, or visit The Alan Turing Institute's frequently asked questions page. If you wish to contact the Turing with a specific enquiry, please email communications@turing.ac.uk.
- for a guide for the responsible design and implementation of AI systems in the public sector, see The Alan Turing Institute Understanding artificial intelligence ethics and safety
- for information on data ethics related to the use of models within research and statistical contexts, please contact the UKSA's data ethics team or visit the UKSA Centre for Applied Data Ethics. Also read the Central Digital and Data Office Data Ethics Framework. This will help to ensure that the use of models is ethically appropriate in a given context
- for practical guidance on modelling techniques, please contact the Office for National Statistics Data Science Campus or consult GSS Best Practice Team's Quality Assurance for Coding in Analysis or Research guidance (the Duck Book)
- for wider guidance on producing quality analysis for government refer to HM Treasury Aqua Book
- HMRC have an internal government guidance document called the 'Predictive Analytics Handbook' which is also a useful resource