Ethical Considerations in the use of machine learning for research and statistics

Alice Toms, UK Statistics Authority

Alice.toms@statistics.gov.uk
Data.ethics@statistics.gov.uk

UK Statistics Authority

- Independent body at arm's length from the Government.

- 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.

Strategy


UK Statistics Authority

UK Statistics Authority
Five year strategy
2020 to 2025

Statistics for the public good

Informing the UK.
Improving lives.
Building the future.

Our Mission

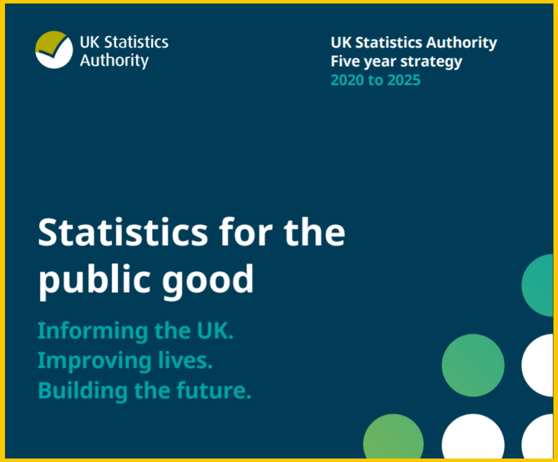
High quality data and analysis to inform the UK, improve lives and build the future.

UK STATISTICS AUTHORITY

Our ambition

To be recognised leaders in the **practical application** of **data ethics** for **statistics** and **research**.

UK STATISTICS AUTHORITY



UNECE Machine Learning Workstream: Data Ethics

Establishment of a set of **ethical principles** to provide a clear framework to **enable ethical use of Machine Learning** for research and statistics.

Creation of **applied** guidance and/or an accessible tool to **empower statisticians** to apply data ethics principles to their work, so that ethical risks can be identified and mitigated against.

UK Statistics Authority's [ethical principles]

Public Good	The use of data has clear benefits for users and serves the public good.
Confidentiality, data security	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.
Methods and Quality	The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.
Legal Compliance	Data used and methods employed are consistent with legal requirements such as Data Protection Legislation, the Human Rights Act 1998, the Statistics and Registration Service Act 2007, public equalities duty and the common law duty of confidence.
Public views & engagement	The views of the public are considered in light of the data used and the perceived benefits of the research.
Transparency	The access, use and sharing of data is **transparent**, and is communicated clearly and accessibly to the public.

Review Findings

- There are already a number of **great resources available** to help researchers think about these issues. A handful of these provide applied guidance, as well as high-level principles.
- There is **significant overlap** in the guidance that is already available across the international community.
- The majority of ethical guidance considered as part of this review had a broader **focus on use of AI**, as opposed to a sole focus on ML.
- Clear agreement that **sound and quality methods**, and **transparency** are vital considerations for the use of AI applications in research and statistical production.

Developing our Machine Learning ethics guidance

Identifying user needs

- Scoping review of literature
- Lack of available guidance in this area

Co-creating guidance

- Engagement with range of stakeholders
- Iterative development with colleagues

Release of open early draft

- Feedback from wider user community
- Further iterative development
- Review approach & application

Why is ethics important?

Reduce potential harm to all individuals involved in the research

Helps maintain public acceptability around the production of research and statistics

Enables researchers to efficiently access and harness data that supports the production of statistics for the public good.

Promote a culture of ethics by design

The potential for bias

- In many circumstances, bias in machine learning is a result of cognitive human error **– problems are introduced by those who design or train the systems.**

- For example, the training data may be incomplete, or unrepresentative of the correct population.

- Bias though does not always exist as a result of human error, for example, it may be introduced into data sets or models as a consequence of previous societal norms.

- Sample bias
- Algorithmic bias
- Prejudicial bias
- Observer bias
- Exclusion bias

Possible Mitigations:

- Take time to reflect on possible biases that may find their way into the data you are using. Reflexivity is key!
- It may be useful to discuss your project with colleagues who are not involved in the project as part of an independent review.
- Ensure that training datasets are as representative of the correct population as possible. This may help counteract sample and prejudicial bias.
- You cannot expect your model to learn what a cat looks like, if you only feed it images of dogs!
- Where possible, the algorithms and data sets should be tested and validated to mitigate any possible biases, and systems should be systematically monitored

TAKE CAUTION WHEN REMOVING SENSITIVE ATTRIBUTES FROM TRAINING DATA FOR THE PURPOSE OF MITIGATING BIAS.

This could make the problem worse as correlated attributes still existent in the data may still reflect the bias you are trying to erase.

Transparency and Explainability

Transparency

- Relates to the obligation of researchers to ensure that the decisions they make about their data, analysis, and methods, are openly and honestly documented and communicated in a way that allows others to evaluate them.

Explainability

- Used to define machine learning techniques that as humans, we are adequately able to <u>understand, trust and manage</u>.

Advice and Mitigations:

Everyone interacting with the project should be able and encouraged to ask questions at any point within the research process, and this should be communicated in a user interface, or similar document.

If the choice is taken that an opaque machine learning algorithm is to be used over a more explainable one, researchers may need to consider how this can be communicated to a non-specialist audience. All human decision-making processes should be audited and easily reviewable.

It may be useful to discuss your project with independent colleagues as part of “challenge sessions”.

It is always helpful to put yourself in the shoes of other stakeholders when you design any project, and this can be particularly helpful when thinking about how to communicate complex ideas (such as machine learning algorithms and systems) to non-expert audiences.

COMMUNICATING WITH NON-EXPERTS WHEN USING MACHINE LEARNING TECHNIQUES TO ENSURE TRANSPARENCY

WHAT INFORMATION MIGHT YOUR AUDIENCE FIND HELPFUL?	WHAT INFORMATION MIGHT YOUR AUDIENCE NOT FIND HELPFUL?
Take a minute to consider what *you* would like to know about a project if you were approached by someone who wanted to tell you about their machine learning research. Consider this from the perspective of the audience you are trying to communicate with…	Take a minute to consider what information you may not find helpful if you were approached by someone who wanted to tell you about their machine learning research from the perspective of the audience you are trying to communicate with…
MEMBERS OF THE PUBLIC*	
1. What is machine learning? 2. Why did you choose to use machine learning over other methods? 3. What is the aim of the research? 4. Why is the study important, and what will you do with the results? 5. Were there any limitations to your research, or the machine learning methods that you used? 6. How did you access the data and how will it be used?	1. What will the typical lay person learn from being presented with an algorithm? Are they likely to understand it, or is there a more understandable way of presenting this to a lay audience? 2. Too much information! It may put users off if there is too much information, or if the information presented is hard to read. How can you present the information in a way that is concise and easy to read?

Accountability

Researchers must be able to justify the method(s) that they have chosen to use. This should be clearly documented.

It is the collective responsibility of those who develop and train models, and those who deploy pre-existing models to ensure that the use case of a model is clear, and that models are not used beyond their intended use.

Using pre-trained models may be particularly problematic if the model lacks transparency, as it may be harder to identify the processes used to train the model, and existing biases.

Whilst developers cannot stop others from using their models, to remain accountable, **anyone making and training models should try to be explicit in communicating the intended use of a model**, so that models are not created that are then used by others in the wrong way.

Possible Mitigations:

If you play ***any*** role no matter how big or how small in the design or development of a machine learning project, it is your responsibility to understand what your role is, and the policies relevant to your role.

You should be accepting of scrutiny!

Ensure accountability by design, by having governance processes to ensure human oversight at the appropriate organisational level throughout design and implementation.

It is vital that a model is audited at regular intervals, and that this is well documented. Sufficient time should be built into the project plan to allow for this.

A continuous chain of human responsibility should be established through the full lifecycle of the machine learning model. All human involvement and oversight should be transparent, documenting all roles and actions taken to ensure human responsibility can be identified.

Confidentiality

Has data minimisation been appropriately considered? Only the data that is required should be stored and used, and any unnecessary data should be deleted once it has been determined that it is appropriate to do so.

Have you considered whether it is appropriate to anonymise your data, and if so, what the most appropriate method(s) of anonymisation will be?

Have you ensured that your data is being safely stored?

Our guidance:

- Ethical considerations in the use of geospatial data for research and statistics
- Considering public good in research and statistics: Ethics guidance
- Considering public views and engagement regarding the use of data for research and statistics
- Using data from third parties for research and statistics: High-level ethics checklist
- Ethical considerations in the use of machine learning for research and statistics

Contact us: **Data.ethics@statistics.gov.uk**

Or visit our website:

https://uksa.statisticsauthority.gov.uk/data-ethics/