High-level ethics checklist for machine learning

☐ 1. THINK

Step back from the project and think hard about the work you are about to undertake and where ethical issues might arise. In particular what might you do that:

- breaks confidentiality or privacy
- discriminates or disadvantages
- is biased or
- misrepresents the truth

Try and think about those things that might go wrong and take steps to avoid them – using the advice provided in our guidance and by seeking further support from colleagues, or the UK Statistics Authority Data Ethics team, if necessary.

Consider in particular the UK Statistics Authority's general ethical principles below:

☐ 2. Public good

Have the benefits of using machine learning for a project been clearly documented? For further guidance on communicating public good, see our Public Good guidance.

☐ 3. Methods and quality

Is machine learning the most suitable method to use? Have the limitations of machine learning data/technologies been considered?

☐ 4. Transparency

Has transparency in the collection, use, retention and sharing of the data being used and produced been considered?

☐ 5. Legal compliance

Has relevant regulation been considered in relation to the dataset used, both in the UK and if necessary, overseas?

☐ 6. Public views and engagement

Have potential public views regarding particular uses of machine learning data across different contexts been considered?

☐ 7. Confidentiality and data security

Have appropriate mechanisms to maintain confidentiality of datasets been considered?

To assist you in applying these ethical principles to your work, we recommend that you use the [UK Statistics Authority ethics self-assessment tool](#), which breaks each principle down into smaller items.

Consider these points more specific to machine learning in a statistical context:

☐ 8. The potential for bias

Have you considered the potential for bias, which could arise from any of your data?

Have you scrutinised your training data for potential biases, and considered the potential for your own conscious or sub-conscious biases to be reflected in the data?

If potential bias is identified, ensure that this is documented to enable informed interpretation of results. What are the potential implications of this bias, and how can this be minimised?

☐ 9. Transparency and explainability

Who are the key stakeholders who need to be considered when communicating your project, and what are likely to be their main questions and concerns? Assuming that you had no knowledge of the project, what would you like to know about how your data is being used to provide outputs that inform decision-makers?

Are you able to explain what data is being used to train the algorithm, and what you expect to get from the data afterwards?

Have you enabled an open and transparent system to allow stakeholders to ask questions throughout the research process?

Would another researcher be able to reproduce your results with the information available to them?

☐ 10. Accountability

Has human accountability been built into the project from the design phase? Are there structures in place to enable accountability?

Has a chain of human responsibility been established, with each stage of the project's lifecycle being documented to show the human oversight?

Has time been put aside, throughout the lifecycle, to account for an audit of the machine learning model?

If you have created a model yourself, has the intended use of the model been clearly communicated to ensure that it is not misused?

If you are using a pre-existing model, have you ensured that you are using it in the way it was intended?

☐ 11. 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?

Might your data, system, or results be re-used outside of their original context and purpose in the future to the disadvantage of individuals, groups or communities? What can you do to try and protect against this possibility?

If you require any further support in addressing ethical issues in your planned project or incorporating ethical principles or would like to provide feedback please contact the UK Statistics Authority Data Ethics team. Visit our webpages for further information on the support services that we offer.