Machine Learning Engine Principles

Introduction

Top tips
The principles

The BBC's Machine Learning Engine Principles (MLEP) and tools represent a way of **incorporating BBC's values and priorities into the technology we build.**

They will ensure that we **avoid common pitfalls** around Artificial Intelligence (AI) and Machine Learning (ML).

This checklist is intended as a set of prompts to turn our ML **principles into practice.**

It contains a list of questions for teams to work through and review. It has been developed by an inter-disciplinary group, drawing from best practice within the BBC and the industry.

Top tips:

The intention of the checklist is to **make thinking happen.**

It is a **self-audit tool** - and up to you and your team how you use it.

We recommend:

Using the ‘Scoping your ML project’ section as early as possible e.g. in the discovery phase.

Building the checklist into the development of your project - it is not intended as a review tool or something to fill out at the end.

Saving a record of your checklist responses.

Values

Reflecting the BBC's values

The BBC's ML engines will reflect the values of our organisation; upholding trust, putting audiences at the heart of everything we do, celebrating diversity, delivering quality and value for money and boosting creativity.

Audiences

Our audiences

Our audiences create the data which fuels some of the BBC's ML engines, alongside BBC data. We hold audience-created data on their behalf and use it to improve their experiences with the BBC.

Clear explanations

Audiences have a right to know what we are doing with their data. We will explain, in plain English, what data we collect and how this is being used, for example in personalisation and recommendations.

Responsibility

Editorial values and broadening horizons

Where ML engines surface content, outcomes are compliant with the BBC's editorial values (and where relevant as set out in our editorial guidelines). We will also seek to broaden, rather than narrow, our audience's horizons.

Taking responsibility: review, security and fairness

The BBC takes full responsibility for the functioning of our ML engines (in house and third party). Through regular documentation, monitoring and review, we will ensure that data is handled securely. And that our algorithms serve our audiences equally & fairly, so that the full breadth of the BBC is available to everyone.

Human in the loop

ML is an evolving set of technologies, where the BBC continues to innovate and experiment. Algorithms form only part of the content discovery process for our audiences, and sit alongside (human) editorial curation.

1.

Scoping your ML project

Intended ML application

1.1 What will this application of ML do and why is it being created?

1.2 Have you considered using potential alternatives to ML? If not, why is ML appropriate or essential to your project?

1.3 Which of the BBC's public purposes does the project support? Please give reasons for your choice.

1.4 How does this project represent value both to the BBC and (if your project is external facing) to the BBC's audiences?

1.5 What are the desired/expected outcomes?

Impact

1.6 Who will be affected by the deployment of this system? Will it have an impact on any audience-facing services? If so have you considered the editorial policy aspects?

Risks, opportunities and consequences

1.7 What are some of the potential limitations, issues, or risks that could arise from your project?

1.8 Have you considered whether any groups could be negatively impacted because of the use of your application? How might you mitigate this?

2.

Planning your ML project

Public service outcomes

2.1 How are you defining what success (e.g. intended outcomes) looks like? How will you evaluate the effectiveness of your approach in line with this?

2.2 What is your process for logging, reporting, and escalating issues? Who are the key people who must be contacted in the event that unanticipated risks and issues arise?

2.3 Is your application built for a product that has a younger audience or does it directly affect children? If so, how have you ensured that you conform to BBC guidelines relating to children and included this in your DPIA?

2.4 Is your application built for a product that deals with sensitive topics? If so, how have you ensured that you conform to BBC guidelines relating to harm and offence?

Ensuring fairness and equality

2.5 Is your team multi-disciplinary? Do those that provide feedback - such as testers - bring different perspectives?

How have you sought out diversity of thought (e.g. in your choices, of data sources, design process, functionality, UX)?

What specific areas of expertise and lived experience are important to your project (e.g. beyond technical)?

What measures will you put in place to ensure the perspectives of these relevant groups are taken into account?

2.6 What have you done to understand the potential impact of the system you are developing on people with protected characteristics (age, disability, gender, race, religion/belief, sexual orientation)?

Sign-off

2.7 Who is responsible for go-live sign-off, and is the sign-off process documented?

3.

Working with editorial values

If your project has editorial consequences then…

3.1 Have you considered how editorial stakeholders can be involved throughout your project?

3.2 What editorial insights would be useful to know? e.g. editorial priorities, audience observations, production workflows, public service offers.

3.3 Do you have a way of documenting editorial decision-making? e.g. via an editorial decisions log?

User insights

3.4 What user insights would be useful to know? e.g. audience preferences, issues which trigger user complaints, niche audiences.

3.5 What implications do these user insights have for editorial decision-making, e.g. via business rules and algorithmic weightings?

Content insights

3.6 What content production systems are used, and what impact does this have on your data engineering?

3.7 Should you conduct a data/content audit to gain editorial insights? This might include details of the production workflows, identifying edge cases etc.

3.8 What implications do these content insights have for editorial decision-making, e.g. via business rules and algorithmic weightings?

Editorial principles

3.9 Have you considered if/how the following editorial principles should be reflected in your project:

- Provide an experience that is impartial and reflects editorial integrity.

- Provide an experience that is in the public interest.

- Provide content that meets editorial and legal obligations.

- Reflect editorial judgement and sensitivity when providing content that is challenging.

- Protect vulnerable groups and the privacy of contributors.

Provide users with good quality content (metadata, fits the users context).

- Provide breadth and/or depth depending on the user's preference and context as well as editorial priorities.

- Doesn't violate any other editorial guidelines.

4.

Other relevant BBC processes

The following BBC processes or sources of expertise might also be relevant to your project:

4.1

Is other domain expertise required for this project (HR, UX, audiences etc)?

For example, you would need to consult HR if designing a project for staff to opt into. Or audience analytics teams can provide insights into audience behaviour and preferences.

4.2

Legal status of the data

- Have you checked the status of the data (e.g. personal, identifiable) and the legal basis for which the data is/was processed?

- Are you compliant with GDPR principles around data sourcing, use, minimisation and retention?

- Have you kept a record of your Data Protection & Impact Assessment (DPIA), if applicable?

4.3

Have you carried out a threat model review with InfoSec? This will cover topics including:

- The effect of malicious inputs intended to confuse or disrupt the model.

- Security controls in your non-production environments.

- Whether your model could be used to make inferences about an individual e.g. reveal their identity based on the data being processed.

- The InfoSec classification of the datasets that you are processing.

- The impact on other areas of the BBC if your service were to be compromised and a data breach occurred.

Refer to ML/AI Infosec Guidelines for further detail and examples.

4.4

If any procurement is involved, have you contacted Commercial Rights and Business Affairs?

5.

Data and privacy

5.1

Is your work aligned with the BBC privacy promise, BBC privacy practices and the BBC's commitment to privacy by design?

5.2

In order to understand what data is being used:

- Have your data source origins been documented (i.e. how and why it was collected)?

- Have all your transformations and/or modifications been documented? Documentation here includes but is not limited to: well-structured, accessible, and legible source code to perform transforms; data schemas; and system diagrams.

5.3

If using personal data: could you achieve your aims with non-personal data or minimised data? If not, are your aims clearly stated and justified?

5.4

Data Quality and Fairness:

- What are you doing to ensure quality?

- What are you doing to counter possible errors and sources of unfair bias in the data?

- How are you trying to minimise the effect of unfair biased data? If you have fixed groups, have you taken appropriate steps to fairly represent those groups (this may include minimum representation samples or other weighting techniques)?

6.

Training and testing a ML model

6.1 Have you asked for domain expert advice in your feature engineering process?

6.2 (having thought about the impact of your ML application on different groups as per 2.6) Have you carried out appropriate and reasonable tests for bias? (e.g. tests for disparate error rates).

Do these require any changes to the model, or data sources? (5.4)

6.3 Have you considered how particular types of use or less predictable use of the system could affect its expected performance, and have you carried out appropriate testing for this at this stage?

6.4 How will you monitor if the model is giving unexpected or incorrect results?

6.5 Can you fix the model quickly if it breaks? For example through retraining or rolling back to a previous version of the model?

6.6 How easily and quickly can you retrain and redeploy the model?

7.

Model documentation & transparency

7.1 Has this ML project been added to the BBC AI+ML registry?

7.2 Have you sufficiently described what your model does and documented how it was created? e.g. via a data science decision log.

7.3 Do you have an explanation or visualisation of the model that can be used to effectively communicate – in plain English – its purpose and how it works ?

Have you considered how to do this for:

- Your users (whether internal or external).
- Your ML project stakeholders across the BBC (whether direct - e.g. product, or more widely, e.g. Quality, Risk & Assurance division).

8.

Life cycle management & monitoring performance

8.1

Have you got a plan to monitor and review the continuing validity of your model and its live performance, including what to do if it is not performing as expected?

Does your plan include consultation with the relevant users (the audience, stakeholders, domain experts and others who use the model output) to ensure that it is working as expected?

8.2

What is a sensible cadence for review of the results? Quarterly, monthly, weekly, prior to large release, as needed and/or when triggered by performance metrics?

8.3

Given input will change over time, how will you be monitoring the output when new data is added?

8.4

What is the defined process to decommission the ML system if this is required?

9.

Checklist review

9.1

What important changes (or revisioning / redeployment of the model) would trigger a MLEP checklist review? Are there particular sections or questions you should revisit? If so, when?

9.2

Would it be helpful to get peer review of your checklist responses?

BBC

Machine Learning Engine Principles

The principles:

- Reflect BBC values
- Improving audience experience
- Transparency: explaining in plain English
- Responsibilty, security & fairness
- Editorial values & broadening horizons
- Human in the loop

The checklist:

1. Scoping your ML project
2. Planning your ML project
3. Working with editorial values
4. Other relevant BBC processes
5. Data & privacy
6. Training & testing a ML model
7. Model documentation & transparency
8. Lifecycle management & monitoring performance
9. Checklist review