TAP
Technology Advisory Panel
IPCO

AI Proportionality Assessment Aid

April 2025

Background

This aid presents a practical framework to guide the proportionality assessment of an AI model. It does not represent IPCO policy.

Purpose

The purpose of the paper is to provide a practical framework for guiding the proportionality assessment of AI models, specifically focusing on data-driven AI models. So, it is mainly directed at the latter, though it could be applicable in the former, if proportionality was a concern. It outlines key questions to consider throughout the AI model lifecycle—concept, development, operation, and exploitation of results.

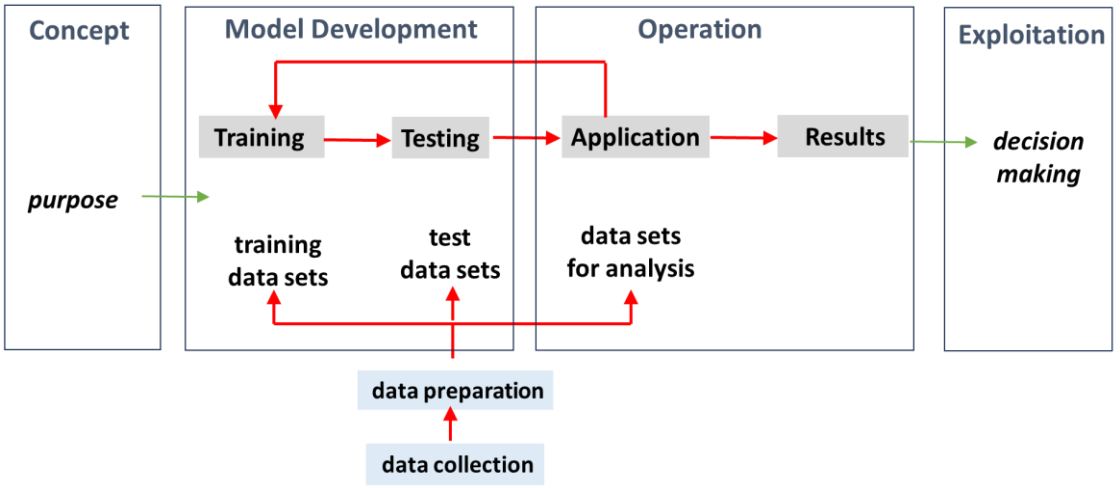
This is not the legal test of proportionality under the IPA; rather, it is designed to ensure that AI models are the appropriate tool for the task at hand. The paper is not specifically focused on IPCO's use of AI in its regulatory functions, nor is it restricted to public authorities' use of AI models in the exercise of investigatory powers. Instead, it provides a general framework for assessing AI models across various contexts.

The aid consists of questions to consider during each of the four stages in the model lifecycle: concept, development, operation, and exploitation of results. It builds on the recent CETaS (Centre for Emerging Technology and Security, The Alan Turing Institute) structured framework for assessing privacy intrusion[1], but there are additional considerations, and it is organised according to the lifecycle stages. It is oriented towards data-driven AI models.

Not all questions will be applicable in every context and there are no correct or incorrect answers; rather, the questions serve as a guide to factors that should be considered when making an assessment. The aid may also serve as a guide for model commissioners, designers, developers, testers, and those involved in governance.

[1] A. Janjeva, M. Calder, M. Oswald. Structured Framework for Assessing Proportionality of Privacy Intrusion of Automated Analytics. CETaS Research Reports, May 2023.

Figure 1: **Four stages of AI model lifecycle: concept, development, operation, and exploitation. Data preparation and collection processes may also involve use of AI models.**



One concept may lead to the development of several models, which in turn may be applied in several different operations. Conversely, an operation may employ several different models.

Proportionality considerations at each stage

Concept

1. Is this a new ***capability***, a step change in scale, a new variant in a family of capabilities, a minor improvement of a current capability, or a replacement of a current capability?

2. Why is an ***AI approach*** more appropriate than any other automated approach (to achieve a similar purpose), e.g. why is the past a good predictor of future behaviour and why is a rules based approach not suitable?

3. What are the ***constraints*** and how would they typically be balanced with ***urgency/potential harm***, e.g. what are the impacts of false positives and false negatives, and is the capability appropriate when the expected urgency/potential harm is high and/or low?

4. What ***biases are unavoidable*** for the purpose and what are their impacts, especially on different communities; how is this made transparent?

5. What is the policy for ***frequency of re-training/testing***, who has responsibility for determining it and for checking compliance?

6. Could a ***higher level of intrusion at training/testing*** lead to ***reduced intrusion at operation***?

7. What biases, constraints, and parameters (internal and external) could affect performance, and at which stages, e.g. training, testing, and operation, and ***who is authorised to change parameter settings and when***?

Model development – training and testing

This information should be available from the *model card* and *warrantry*.

All data

1. What are the ***sensitivities*** (including those that can be inferred) and can any be reduced without affecting purpose?

2. What is the degree of ***human inspection***, at any stage, why is it required, and could it be reduced or should it be increased?

3. What is the justification of data ***volumes***, could they be reduced (without affecting required statistical measures) and/or how is excess data treated?

4. What are the ***granularities***, ***biases***, *and* ***adequacy***, are they justified, do they align across training, testing, and operation, and can granularities and bias be reduced, or integrity and adequacy improved?

5. What is the ***provenance*** and ***integrity*** of the data sets, who has responsibility for their assurance and are the ***retention*** and ***deletion*** policies adequate and easy to implement?

6. Is there any ***differential intrusion*** (that is not related to purpose) to communities with protected or sensitive characteristics.

7. What is ***access control***, how is the data protected from ***loss*** and ***corruption,*** and what are mitigations?

8. If there are ***3rd party suppliers*** of data sets or pre-trained models, what are the assurances about data provenance and integrity and more generally supplier competence and transparency of process. If there is in house fine tuning, is there a possibility of data leakage to the 3rd party?

Training data

9. What is the quality of any ***labelling*** and are any additional sensitivities revealed that are not necessary for the purpose?

Testing data

10. What are the selected characteristics and their justification, which ***statistical measures*** are employed, why are they appropriate, and do any raise concerns?

Training Process

11. Is there ***process transparency*** and are there any concerns about manual interventions or random training algorithms?

Model operation

1. Is this an ***appropriate use of the model***, including balancing the impact of false positives and negatives, and alignment of granularity and integrity of the data for analysis with that used in model testing and training, with the operational urgency?

2. What is the ***provenance*** and ***integrity*** of data for analysis and the ***retention***, ***deletion,*** and ***access control*** policies for both data for analysis and results, how is the data protected from ***loss*** and ***corruption,*** what are the mitigations, and will the processes hold up to evidential scrutiny?

3. Is delivery of results ***staged*** or ***prioritised*** and can operation be halted at various stages to reduce unnecessary production of results?

4. To what extent are results ***understandable*** and ***actionable***, at any stage?

5. To what extent is ***human inspection*** of the data for analysis or the results required and is it achievable, e.g. is the volume or rate of production is too high?

Exploitation

1. What are the ***people***, ***reports***, and ***systems*** that have access to the results and what training do people have that ensures they make appropriate use of the AI and understand its limitations.

2. ***Model chains*** - are results from the model ingested directly, without human intervention, as ***training*** data for another model, **re-training** data for this model, ***analysis*** data for another model, or into any ***data sets*** that may be subject to future collection processes for this model?

3. What are the ***reporting processes*** following from adverse outcomes that derive from the results?

Overall evaluation of proportionality – example questions to consider

1. Are ***volumes*** and/or (possibly inferred) ***sensitivities*** of data sets minimal?

2. Is any ***human inspection*** of data minimal and feasible?

3. Are ***biases*** and ***uncertainties*** justified, minimal, and transparent?

4. Is ***testing adequate*** (e.g. statistically meaningful), for all appropriate characteristics, and are any statistical concerns ***balanced*** by the ***urgency*** of the operation?

5. Are the data sets appropriately ***sourced*** and ***curated*** and is ***data management***, across the whole lifecycle, adequate and transparent?

6. Is ***intrusion*** on ***communities*** minimal and related to purpose?

7. If ***further*** data collection or another AI tool is ***triggered***, is human judgement involved?

8. Are there robust mechanisms for implementing change based on ***feedback*** from ***error*** or ***adverse outcome*** reports?

9. Are all the people involved ***suitably trained*** and ***certified*** and will all processes hold up to ***evidential scrutiny***?