



Department of Telecommunications
Telecommunication Engineering Centre
K.L. Bhawan, Janpath, New Delhi – 110001
(Convergence & Broadcasting Division)



No.: 4-3/2022-C&B/TEC/AI-Fairness

Date: 22.02.2022

Subject: Framework for Fairness Assessment of Artificial Intelligence/ Machine Learning Systems by TEC - building public trust in AI/ ML Systems #AIforAll

Artificial intelligence (AI) and Machine Learning (ML) applications are increasingly being used in all domains such as Healthcare, Agriculture, Smart Cities, Smart Homes, Finance, Defence, Transport, Logistics, Natural Language Processing, Surveillance, and so on. Various Government Organisations are also using AI/ ML Systems for delivery of public services and e-governance.

Researchers have shown that many such Systems deployed worldwide were having unintended biases. Bias in AI/ ML Systems raises various ethical, social and legal issues. While the users expect these systems to be fair in their outcomes, a biased AI/ ML System prefers a certain demography while discriminating against others. When AI/ ML Systems are used for e-governance or by the judiciary, checking for their fairness would become a legal requirement. Therefore, one important requirement of Responsible AI is that the AI/ ML Systems should be unbiased or fair.

National Digital Communications Policy-2018 mandates for synergising deployment and adoption of AI. National Strategy for Artificial Intelligence #AIforAll, and the Approach Documents for India, released by NITI Aayog, establish broad ethical principles for design, development and deployment of AI in India and provide an essential roadmap for the AI ecosystem, encouraging adoption of AI in a responsible manner in India and building public trust in the use of this technology, placing the idea of 'AI for All' at its very core.

With the aim to build public trust in AI/ ML Systems, TEC is working on Voluntary Fairness Assessment of AI/ ML Systems. As these systems are being used in all aspects of our lives, the initiative of TEC would benefit every citizen of the country. Startups and MSMEs in particular and even large enterprises will benefit as their products will be more credible and acceptable if they are assessed and certified by a neutral government agency such as TEC. It would also act as a benchmark for fairness, so Researchers, College students and other individuals, Government organisations and all segments of the society would benefit at large.

Accordingly, TEC is initiating stakeholder consultations and invites suggestions for framing procedures for assessing fairness for different types of AI/ ML Systems. Your valuable inputs on the matter as per the enclosed template may be emailed to avinash.70@gov.in and adic1.tec@gov.in by 08-03-2022 please.

This issues with the approval of Sr.DDG & Head TEC.

(Avinash Agarwal)

DDG (Convergence & Broadcasting)
Telecommunication Engineering Centre
Department of Telecom, New Delhi
Email: avinash.70@gov.in

To:

1. All stakeholders

Copy for kind information to:

1. CEO Niti Aayog
2. Secretary Telecom/ Agriculture/ DST/ Defence/ Electronics & IT/ MoHUA/ MoRT&H
3. Member (Services)/ Member (Technology)/ DG (Telecom) DoT
4. Principal Scientific Adviser (PSA) to the Government of India
5. Chairman, TRAI/ Chairman DRDO
6. CEO Prasar Bharati
7. DG ICMR/ DG CDAC/ DG BIS
8. DG NTIPRIT
9. Directors of all IITs

Framework for Fairness Assessment of Artificial Intelligence (AI)/ Machine Learning (ML) Systems by TEC

Template for submitting inputs

(Inputs may be provided as per the following template)

1. Types of AI/ ML Systems

There can be different ways to classify AI/ ML Systems. Based on the machine learning algorithms used, they can be classified as supervised, semi-supervised, unsupervised and reinforcement learning systems.

Question 1: For assessing fairness of various AI/ ML Systems, whether classification based on machine learning algorithms used is suitable or any other classification is suggested?

2. Supervised learning systems

Supervised learning systems learn from labelled datasets and are used to classify data or predict outcomes of unforeseen data accurately. They are typically used for various predictions and forecasting, face detection, handwriting/ signature recognition, etc. Some applications could include identifying fraudulent benefits claims, CCTV surveillance, social media sentiment analysis, spam detection, weather forecasting, stock price predictions.

Question 2: For assessing fairness of various supervised learning based AI/ ML Systems, whether a single assessment procedure is sufficient or further sub-classification is required?

3. Types of possible Biases in supervised learning systems

An AI/ ML application could have different types of biases. Also, some type of biases might be typical to the types of AI/ ML Systems. Some types of bias are selection bias, measurement bias, recall bias, observer bias, exclusion bias. The biases could be due to datasets used for training, algorithms and/ or usage of the System.

Question 3: What are the various types of unintended biases that should be taken into account while assessing fairness of various AI/ ML Systems?

4. Fairness Metrics/ parameters

Fairness of various AI/ ML Systems is usually measured using various metrics. These metrics provide a mathematical definition of fairness. Some commonly used fairness metrics are demographic parity, equal opportunity, equal mis-opportunity, average odds. Many fairness metrics are mutually exclusive;

Question 4: Which fairness metrics/ parameters are appropriate for measuring bias/ assessing fairness of supervised learning AI/ ML Systems?

5. Standardising procedure for fairness assessment of supervised learning AI/ ML Systems

While developers may test their AI/ ML applications using various fairness metrics and their own internal processes, there is a need to standardise the assessment process. The standardised process could be used by the developers for self-assessment as well as by third-party auditors for assessment and certification.

Question 5: What could be the step-by-step procedure for assessing and rating a supervised learning AI/ ML Systems for fairness?

6. White-box vs Black-box testing

Ideally, the assessing agency may expect access to the training datasets and the implementation details of the AI/ ML System. This may be considered as White-box testing. There might however be situations when the training datasets and/ or implementation details are not available with the assessing agency. This may be considered as Black-box testing.

Question 6: Whether any separate procedure is required for black-box testing of supervised learning AI/ ML Systems for fairness?

7. Procedures for handling code and training datasets by the assessing agency

When the assessment is to be carried out a third-party then it is important the standard procedures are in place for safe handling of sensitive material such as code, datasets, etc.

Question 7: What are the best practices for handling code, training datasets, etc. by the third-party assessing agency?

8. Semi-supervised, unsupervised and reinforcement learning and other type of systems

Besides supervised learning systems, other types of AI/ ML Systems could be Semi-supervised, unsupervised or reinforcement learning systems. Unlike supervised learning systems, the unsupervised learning systems use unlabelled data to identify patterns. Semi-supervised learning systems are a mix of the supervised and the unsupervised learning systems. Reinforcement learning generally learns from new situations using a trial-and-error method.

Question 8: Whether the fairness assessment procedure framed for supervised learning systems would also be applicable for assessing other types of AI/ ML Systems or different procedures need to be framed?

9. Any other inputs

The stakeholders may provide inputs on any other aspect that has not been covered in this template. These could include uncovered points, challenges foreseen, standard testing tools, etc.

Question 9: What are the additional considerations for framing fairness assessment procedures for various types of AI/ ML Systems?