
# Paradigm of Ethical AI

An ADAz Data & AI Whitepaper

by **Debraj Das**

ADVANTAGE AZURE AZ Data & AI

# Paradigm of Ethical AI

## Context of Ethical AI

Artificial Intelligence is on use almost in every business functions across industries to the extent feasible and the analysts envisage that to be used more rigorously in the days to come. Eventually with evolving Technology Capabilities, especially in Machine Intelligence space, the reach as well as impact of Artificial Intelligence in Enterprise as well as Consumerized functions would be multifold. As Artificial Intelligence (AI) is playing a parallel role in Decision Making and Corresponding Actions in most cases, it is necessary to ensure that an AI-driven Decision or Action is always be **Ethical** to all its users. Like ethics of a Human is sought in a Strategy Making, similar way in Machine Intelligence also the touch of Ethical Sense is on demand. To understand it better, let us go through some real life scenarios.

**Scenario 1 – Discrimination in Recruitment Process**

An Enterprise analyses the history Data of their own Recruitment Events and from the Market using Machine Algorithms to ensure better Return of Onboarding Expense. Amongst various shortlisting or rejection criteria they identify and use in Recruitment Automation, one was to avoid pregnant women or even those who are likely to become pregnant in a year's cycle to get rid of productivity and cost hazards during their pre and post maternity period. Eligible and well qualified Women candidates found themselves discriminated here. Machine Learning (ML) guidance was deliberately handled here to cause this discrimination, but can the company, in search of better profit and competitiveness, be held responsible for this while facing Legal challenge by the discriminated women?

**Scenario 2 – Poor People kept out of Health Insurance Coverage**

Two multinational Insurance Companies in Mexico are using Machine Learning to maximize their Efficiency and Profitability. They are mining data for ML Feed from shopping history to recognize patterns associated with high risk customer and charging them more premium for same coverage. Due to this, Poorest and Sickest people are unable to afford access to Health Insurance Services which appears to have implications on human rights to fair access to adequate healthcare. Such Machine Learning output may trap defaulters by-nature one side, but cause severe impact to real poor people as well. Should ML be held responsible here or is it a scarcity of data (e.g. Income Info here might help)?

**Scenario 3 – Allegation of Racism against Bank**

Bank uses Machine Learning to study the history of Repayment of Loan when deciding the mortgage approval amongst several applicants. Applicants not shortlisted files lawsuit against the bank raising the complain of racism against them. Bank declares that Machine Algorithm was blinded to access the race details of applicants. However, the data of Loan defaulters incidentally had majority of black people having some other common properties from their demographics (e.g. residing in poverty stricken area, history of defaulting) which were picked to reject their applications. So is Machine Learning responsible here to create the racial discrimination in true sense? Was it intentional? How AI / ML are to defend themselves

from such allegation? Similar issue found in US Court's Reoffending Criminal Policy following Northpointe algorithm.

Some more incidents shows similar vulnerability in usage of AI / ML is disparate fields like –

- Use of Abusive Words by Microsoft Tay BOT which was designed to learn from Twitters' behaviour
- Tagging Afro-American users as Gorilla by Google Photos
- Amazon Echo auto-blasted music in absence of Resident resulting a police raid
- 3D-Printed Mask beat the Face ID on iPhone X a week after its launch
- Facebook BOT Alice and Bob developed own cryptic language to converse with each other

The challenge in the days to come would be containing the evolving AI / ML reach within Ethical Boundary. While **Robo Ethics** deals mostly with ethics related to Hardware driven Robotic Structures like use of killer robots in war etc., **Machine Ethics** (otherwise called as machine morality or computational ethics) deals with moral behaviour of artificial intelligence. In course of this Blog, discussion would be concentrating on Machine Ethics primarily.

## How the World taking up Ethical AI Journey

Technological Giants in AI space Google, IBM, Microsoft, Amazon and Facebook have set up an Industry led Non-Profit Consortium called **Partnership on AI** in 2016 to come up with **Ethical Standards** around AI work across industries using the help from Policy and Ethics specialists. Accenture and McKinsey joined in that drive in 2017 strengthening the entire approach. Now this initiative has 100+ Partners from 13 countries with more than 50% from Non-Profit Farms including tech conglomerates like Intel, SAP, SFDC et al from different corners of the world.

This consortium works with focus on six (6) areas around AI to promote **Ethical AI** across the AI Implementations –

- Safety-critical AI
- Fair, Transparent and Accountable AI
- AI, Labour and the Economy
- Collaboration between People and AI Systems
- Social and Societal Influence of AI
- AI and Social Good

This consortium believes artificial intelligence technologies hold great promise for raising the quality of people's lives and can be leveraged to help humanity address important global challenges such as climate change, food, inequality, health, and education. They believe in and endeavor to uphold the following tenets –

- AI Benefits would ensure **Empowerment** of as many people as possible
- Active **Engagement of Public** through a both way info sharing and feedback
- Open research towards **Ethical**, **Legal**, **Social** and **Economic** implications of AI
- Engagement through **Accountability** model amongst AI Stakeholders
- Thorough **Involvement of Business SMES** across Domains
- **Privacy** & **Security** of Individuals and Groups would be maintained
- Opposing AI advancement which violates **international conventions** or **human rights**
- AI Systems and Operations should be **understandable** by designated Stakeholders
- Rolling out culture of **Trust**, **Cooperation** and **Openness**

Beyond the Technology Giants, initiatives have been found to be taken by different bodies across Nations as well in search of Ethics in forthcoming AI Implementations.

- **European Economic and Social Committee** called for a code of ethics to cover the development, deployment and use of AI, to ensure AI remains compatible with principles of human dignity, integrity, freedom, cultural & gender diversity and human rights
- **Bureau of Indian Standards** formed a new committee for standardization in AI to focus on projects which revolve around Social, Legal, Ethical and Cybersecurity issues
- **UK's House of Lords** released a report 'AI in the UK: Ready, Willing and Able?' calling for creation of an AI Council targeting a 'common framework for ethical development and deployment of AI'
- In Canada,
  - ✓ **Montreal Declaration on Responsible AI** is trying to stimulate discussion on ethical guidelines
  - ✓ **Treasury Board Secretariat** is looking at responsible use of AI in government
  - ✓ **Global Affairs Team** is leading a multi-university collaboration on AI and human rights
- **New York City** launched a task force to become a global leader on governing Automated Decision Making

## Rudimentary Steps towards Ethical AI

The logical approach can commence with a set of Thumb Rules which all AI implementations must follow.

Rule #1 – Introducing **TRANSPARENCY** by choosing **RIGHT** Algorithm

When ML Algorithm chosen is based on complicated Neural Network or Genetic Logic produced by Directed Evolution, it is nearly impossible for implementer to understand why or even how the algorithm is converging to the decisions or predictions. However, using a path of Decision Tree or Bayesian Network based solution would be much more transparent for programmer's inspection or auditor's discovery.

Rule #2 – Being **ROBUST** against **MANIPULATION** by applying **ADDITIONAL SECURITY MEASURES**

AI / ML solutions handles large amount of data and hence a little manipulation in data can cause in hacked AI output. Additional Security measures like cryptography, multi-factor access verification, biometric validation etc. can warrant no manipulation in AI.

Rule #3 – Precise **ACCOUNTABILITY** as **FALLBACK** of **AI** by imposing Robust **AI GOVERNANCE**

AI having high potential to fail as it has various dependency on algorithm, data, configuration factors like threshold, it is very much susceptible to failure on which often it appears difficult to find an owner of that failure or related fall-back option. Business or / and Technical accountability of every failure (out of service, reduced accuracy, false positive etc.) of AI system should remain on a designated stakeholder as part of the AI Governance defined by the enterprise.

Rule #4 – Attaining **HIGH ACCURACY** through appropriate **ACCURACY GOVERNANCE**

Accuracy in Decision making is utmost critical for success of the AI solution. Design of Brain, Distribution of Use Case and its actions across Brains, in-Brain configuration for Use Cases, deciding right Threshold Value in a coopetitive ecosystem are such areas where fine-tuned optimization is required to ensure High Conversion and Accuracy in AI solutions.

But, even after imposing these Thumb Rules, still there would be some open areas like –

- How to take Moral Decision? E.g. Should an automatic car kill its passengers or the pedestrian who is unexpectedly crossing the road in front of the car?
- How to make Digital Duplication ethical ? E.g. Voice profile of individual can be cloned to generate abusive statement in one's voice to take him to legal action.
- How to restrain AI from Infringement of Privacy? E.g. Should an intelligent Surveillance System stop taking photos when somebody is undressing?

Top AI Providers like Microsoft, Google, Accenture are working on some AI Ecosystem Innovation which might play a role in ensuring Ethics in AI implementation directly or indirectly.

- Microsoft has taken an initiative in March 2018 by forming **AETHER, AI and Ethics in Engineering and Research**, a board of executives drawn from across every division of Microsoft including technical, business as well as legal SMEs. The objective was to create a model to spot issues and potential abuses of AI solutions even before they start which later others can follow . This program is well known a **General AI**. While current systems like Microsoft Cortana (and its competition like Amazon Alexa, Apple Siri, IBM Lucy et al) seem intelligent with ability to say only what they've been programmed to say. a GENERAL AI solution would think and reason like a human including its **Ethical Values** under guidance of AETHER.
- **Google** has started **Ethical AI drive** by dropping AI contract with US Military

- **Accenture** published its Ethical Framework called **Responsible AI & Robotics**

## How General AI Works for Ethical AI

Supervised Learning creates an illusion of Intelligence. While at its core, it is just a Mathematical Optimization and possess an ability to make decisions and classify datasets, it is very narrow in the way it works. Limitation of Supervised Learning are like –

- Thinking is always limited, to a specific domain
- Intelligence depends on the training dataset used i.e. within a controlled boundary. control
- Works with very low accuracy in environments that dynamically change
- Can be used only for either classification and regression

On the other side, **Artificial General Intelligence (AGI)** is the Real intelligent system having ability to

- Think generally to take decisions irrespective of any previous training
- Make decisions based on what they've learnt on their own
- Feel soft senses like pain or suffering like a human being

AGI includes Reinforcement Learning mechanism helping addition of below traits on top of basic AI traits.

- Moral Qualities
  - ✓ Sentience – the capacity to experience sensations like pain and suffering
  - ✓ Sapience – set of capacities associated with higher intelligence, such as self-awareness and responding to reason
- Neural Qualities
  - ✓ Ability of Self Learning from own Mistake
  - ✓ Ability of Auto Optimization as evolving
  - ✓ Environment guided Action

It can be really difficult to design such systems as technology of today is somewhat limited, only **'Partial AGI'** is possible even using very advanced Deep Learning techniques. **Dynamic Programming** and **Control Theory** can be used to implement AGI in any AI implementation. This approach includes –

- Executes Dynamic Programming using Bellman Equation
- Markov Decision Rules (SARSA algorithm) is also used in same purpose
- Deep Q-Learning algorithm is used for Control implementation

# Paradigm of Ethical AI

## Private AI as a Stepping Stone towards General AI

Advances in AI allows to use a trail of data and information to make complex predictions on critical elements like people health, behavior and more. This creates risk of losing individual privacy and uncertainty of unwillingness of people sharing personal data resulting in reduced AI efficiency. **Private AI** is to add privacy feature to AI and ML Processes protecting individuals and saving interest of AI itself, a critical aspect of **Ethical AI** in journey towards General AI Paradigm. Some quick examples of Private AI approach could be –

- In Training Process, building Models on Distributed Data without Sharing the same to Model Builders
- In Prediction Process, making Private Prediction ensuring no or minimal exposure of Prediction Data to Model Owner, Training Data Leakage
- In Statistical Analysis Process, Computing Privacy aggregated Statistics to reduce the risk of exposure of Data during vast aggregation

Private AI uses Advanced Cryptography Technologies and Private AI tools to guaranty Privacy features to Users. Most prominent Techniques of implementing Private AI are – **Homomorphic Encryption**, **Secured Multi-Party Computation** and **Differential Privacy**.

**Approach #1 – Homomorphic Encryption**

Homomorphic Encryption is the most straight forward way to get Private AI implemented. This approach includes the below steps

- Input Data Encryption
- Functional Evaluation of Encrypted Data
- Result in Encrypted Form
- Result Decryption by Data Owner

Most Popular reusable for Homomorphic Encryption is **SEAL** (Simple Encrypted Arithmetic Library) available C++ with .NET wrappers since 2015. Microsoft has implemented this rigorously in Healthcare and Manufacturing industries.

**Approach #2 – Secured Multi-Party Computation**

Multi-Party Computation is the next popular way to implement Private AI. This approach follows the below steps

- Abstracted Computation Fragments executed by individual Modeller
- Dataset for Training and Prediction are distributed
- Aggregation of Fragmented Computation in Secured Channel
- Publishing Result unencrypted

Most Popular reusable for Multi-Party Computation is **SMILY** library having high expense due to communication bottleneck across parties. **ANOVA** is another popular model which uses Neural Network, Boosted Tree, Random Forests to randomize. Open source MPC libraries like **JIFF**, **UMBRAL**, **OPRF** etc. are available in GitHub for usage towards implementing Private AI using Secured Multi-Party Computation.

**Approach #3 – Differential Privacy**

Differential Privacy is a less popular yet very effective approach to get Private AI in place. This approach follows the below steps

- Extract Sensitive Training Data from ML Model
- Inject NOISE in Training Data to introduce ambiguity
- Leak proof model

## Ethical AI is integral part for Future AI

While the space of Ethical AI and enhanced AI Privacy is a very complex area and taking time to get matured in global space, it is eventual that in future, with fast moving AI popularity and demand, any AI implementation would have to include Ethical AI elements within its design to make it a successful implementation. So this is the right time to spend effort in maturing General AI approach both from Technology as well as AI governance perspective.