
RESPONSIBLE AI: HOW FACEBOOK BUILDS AI WITH PRIVACY AND ETHICS BY DESIGN

17TH SEPTEMBER 2019

FACEBOOK, 7TH FLOOR PARSAVNATH CAPITAL TOWERS, BHAI VEER SINGH MARG, NEW DELHI 110001

This is an interactive session where participants will be taken on a journey of how Facebook builds AI in our products from ideation to deployment, in a privacy and ethically responsible manner

9.30 am	Welcome note by Shivnath Thukral
9.50 am	Ethics's Councils: India's vision for ethics in AI- by Ms Anna Roy
10.00 am	What is AI and how are we using it?
10.30 am	Ethical and trustworthy use of AI- by Mr. Gopalakrishnan
10.40 am	Building simplified AI models
11.30 am	Break
11.40 am	How FB builds AI in a fair an unbiased way: FB's approach to AI ethics
12.30 pm	Designing for AI Explainability
1.15 pm	AI integrity
1.25 pm	Thank you
1.30 pm	Lunch

ETHICS IN AI RESEARCH AWARDS- INDIA: PEER REVIEW SESSION

2.30 pm	Introduction
2.40 pm	Patient-Centric Frameworks for the Evaluation of AI-Enabled Medical Tests PI: Amit Sethi, IIT Bombay
3.00 pm	Targeted Bias in Indian Media Outlets PI: Animesh Mukherjee, IIT Kharagpur
3.20 pm	Ethical implications of delegating decision-making journey to AI systems PI: Dr. Rahul De', IIM Bangalore
3.40 pm	Break
3.50 pm	A 'Public Law of Information' for India PI: Sudhir Krishnaswamy, Centre for Law and Policy Research
4.10 pm	Mitigating Bias in Face Recognition for Vast Regional Diversity in India PI: Richa Singh, IIIT - Delhi
4.30 pm	Regulatory Impact Assessment of the National AI Market Place of India PI: Varadharajan Sridhar, IIIT - Bangalore
4.50 pm	Thank you
5.00 pm	Tea

OPERATIONALIZING ETHICS / EXPLAINABILITY / FAIRNESS

PATIENT-CENTRIC FRAMEWORKS FOR THE EVALUATION OF AI-ENABLED MEDICAL TESTS

PI: AMIT SETHI, INDIAN INSTITUTE OF TECHNOLOGY BOMBAY (IIT BOMBAY)
COLLABORATORS: SWAPNIL RANE AND ZAKIA KHAN, TATA MEMORIAL CENTRE

Several ethical concerns and tradeoffs arise when AI-enabled medical tests are being developed and deployed in various jurisdictions, including concerns that generally arise in the AI life cycle and specific concerns related to its application on a local population. Existing laws and ethical guidelines have not sufficiently addressed these concerns that arise in various stages of the AI life cycle. In this study, we propose to use two specific AI problems in medical imaging using data sets from different sources to probe ethical considerations for deploying on a locality. We will develop a patient-centric framework for identifying ethical concerns and prioritizing these concerns. In addition, we will evaluate how high-profile AI diagnostic models can be redesigned for application on a local population using the proposed patient-centric framework.

Specifically, we will highlight weaknesses and pitfalls of popular AI paradigms, and use a patient-centric framework to characterize these weaknesses. We will empirically determine ranges within the proposed framework for proposed models for two specific problems in cancer diagnosis. We will engage experts from legal, medical, AI, and civil rights domains to weigh in on ethical concerns, proposed framework, and findings. We will brainstorm and evaluate additional measures for balancing the tradeoff between patient health outcomes and economics of healthcare delivery. We will synthesize the findings as recommendations for various professional groups involved in design, evaluation and implementation of AI-enabled medical tests.

TARGETED BIAS IN INDIAN MEDIA OUTLETS

PI: ANIMESH MUKHERJEE, INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR (IIT KHARAGPUR)
COLLABORATORS: PAWAN GOYAL AND SOUVIC CHAKRABORTY, IIT KHARAGPUR

There has been little research to identify bias automatically in news media apart from manual studies done by independent journalists. As India crosses the half a billion smartphone users mark, it is important more than ever to characterize information available online through automatic algorithms and auto-updating crowd-sourced knowledge bases to restrict the spread of falsehood thus keeping the fourth pillar of democracy strong with the help of advanced technologies. The scope of the problem lies in leveraging available information on online published media to predict fake news or identify bias in the news media articles.

GOVERNANCE

ETHICAL IMPLICATIONS OF DELEGATING DECISION-MAKING JOURNEY TO AI SYSTEMS

PI: DR. RAHUL DE', INDIAN INSTITUTE OF MANAGEMENT (IIM) BANGALORE
COLLABORATOR: SAI DATTATHRANI, IIM BANGALORE

The objective of the study is to examine the ethical implications of delegating the journey of decision making to autonomous (AI) systems. Much of current research on the ethical concerns that comes with delegation of decision-making to AI systems have focused on the outcome (analogous to the trolley problem) - whether the system should be designed to save a human life vs avoid material loss, whether the decisions made by the systems would be fair and unbiased and such. However, there are ethical concerns irrespective of the outcome such as - is it ethical to be changing the purpose and nature of human decision-making?

A preliminary research by the authors of this proposal reveals that humans value the journey of decision-making - the experience of making these decisions based on their prior insights and experience(s), living through the consequence(s) of making these decisions. This was valued more than the accuracy and efficiency of the outcome of the decisions. With delegating decision-making to AI systems, we would be depriving humans of this valuable experience. Would it be ethical to be modifying this experience?

The study will examine these personal effects and the ethical implications thereof, of delegating this ,journey, using phenomenological analysis of the consciousness of this experience. The output of the study will guide the design of ethical autonomous decision-making systems. This is beyond the trolley problem or human rights. This study examines the ethical implications from the perspective of the purpose of human life.

A 'PUBLIC LAW OF INFORMATION' FOR INDIA

PI: SUDHIR KRISHNASWAMY, CENTRE FOR LAW AND POLICY RESEARCH

The regulation of information and data in India may be classified into three phases. The first phase began in the late 1980s with the right to information advocacy to bring accountable governance to India. While it focused on legally compelling public authorities to enhance transparency and accountability, there was inadequate attention on privacy concerns. Two decades later, we enter the second phase with the inception of the Aadhaar project, which draws on the biometric data of a billion-plus population to ostensibly improve governance. Despite opening the doors for a concerted effort at developing a privacy law for India, it has not resulted in a rigorous data protection regime. In the third phase, we confront the extensive use of data to automate decision making in governance. The interaction between the law, algorithms, and governance calls for innovative ways of addressing privacy concerns and the control of information and is the focus of this proposal.

Under this award, we will explore the prospect of a public law of information in the India that will regulate and direct the use of data in governance systems, by adopting a comparative approach while paying careful attention to the Indian context. We propose to identify the core principles that will govern the collection, storage and use of personal data by public and private entities. We will publish an edited volume of research papers authored by six researchers on regulating algorithmic governance in India, which will also be hosted on the CLPR website and allow open-access.

CULTURAL DIVERSITY

MITIGATING BIAS IN FACE RECOGNITION FOR VAST REGIONAL DIVERSITY IN INDIA

PI: RICHA SINGH, INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY (IIIT) - DELHI
COLLABORATOR: MAYANK VATSA, IIIT DELHI

How can pre-trained deep learning, based face recognition systems be leveraged for unbiased face recognition to handle vast cultural and regional diversity in Indian context? We propose to learn a bias mitigation technique that can be used to extract features from well-trained deep learning models to develop fairer face recognition systems.

REGULATORY IMPACT ASSESSMENT OF THE NATIONAL AI MARKET PLACE OF INDIA

PI: VARADHARAJAN SRIDHAR, INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY (IIIT) - BANGALORE
COLLABORATOR: SHRISHA RAO, IIIT – BANGALORE

We intend to provide a Regulatory Impact Assessment toolkit using Agent Based Model (ABM) that captures the ethical and business behavior of the various stakeholders, namely content creators, annotators, model builders and users, of the proposed National AI Marketplace of India. The tool will enable regulators and policy makers to analyze the transient and stationary behavior of various stakeholders of NAIM, and take informed decisions regarding modes and process of regulating the ethical and business behavior of the participants. Evolution of NAIM will also be analyzed under (i) command-and-control (ii) co-regulation, and (iii) self-regulation models with associated rules and controls.