Ethics and AI in India

Elonnai Hickok

August 2018
Centre for Internet and Society

As used in Takshashila's Online Cogitatum on Future of AI in India

Overview

1. Introduction
2. Defining AI and Ethics
3. Three ethical orientations
4. Exploring Questions related to AI and Ethics
5. Overview of developments at the international level around AI and ethics
6. Key aspects of Ethics and AI
7. Examples from Case studies: healthcare and finance
8. Overview of developments in India around AI and Ethics
9. Further Resources

Introduction

- Why Ethics and AI?
  - As intelligent systems interact with humans either directly or indirectly, machines are increasingly becoming an integral part of the functioning of society.
  - Because of the potential of AI to make intelligent and autonomous decisions there is the question of personhood for machines and the potential that machines will evolve into members of the society.
  - Either as integral to the functioning of society or as a member of society - AI is becoming the underlying layer to our everyday functioning. Because of this it is important to extend the application of ethics from a focus on human to human interaction to a focus on direct and indirect human to machine interaction and interface.
  - It is important that ethical considerations are designed and built into AI solutions from the outset. Multi-disciplinary approaches and input from a range stakeholders is key to ensuring holistic consideration of ethics in AI. This includes working with software engineers and developers, the legal community, civil society, social sciences and humanities, the tech community - as well as domain and sectoral experts.

Defining Ethics

- There are multiple definitions of ethics. A few being:
  - “Moral principles that govern a person's behaviour or the conducting of an activity, the branch of knowledge that deals with moral.” - Oxford Dictionary
  - “Ethics is based on well-founded standards of right and wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, fairness, or specific virtues.” - Markkula Center for Applied Choices - Santa Clara University
  - “Ethics on the other hand, proceeds in a dialectic manner, that is, it uses rigorous analysis to reveal the flaws of logic and the contradictions of the reasoning and seeks to go beyond them. It also deals with what we should or should not do, but it does so by applying reasoning, for or against, in order to decide on the conduct to be taken when faced with a moral problem.” Margot Phaneuf, Ethics some Definitions

Defining AI

There are multiple definitions and categorizations of AI. There are also multiple technologies that make up AI. Below are a few:

- For CIS's research, we took a broad understanding of AI as a dynamic learning system that can be used in decision making and actioning. (CIS Case study compendium)
- Strong AI vs Weak AI: Build systems that think exactly like humans do ("strong AI") vs. Just get systems to work without figuring out how human reasoning works ("weak AI"). Philosophical Arguments Against Strong AI, Univeristy of Texas
- Narrow vs. General vs. Super Intelligence: Narrow AI can perform tasks but works within a parameters and a data set vs. General AI which can perform any task of a human vs. Super AI which can surpass human intelligence. (Medium)
- A number of technologies make up AI including computer vision, audio processing , natural language processing, knowledge representation, machine learning, and expert systems. (Break down as in the NITI Aayog National Strategy for AI)

Three Ethical Orientations

Deontological

Derives from Kant and asks the core question - what is my duty? This can be understood in the form of laws.

Utilitarianism

Seeks to answer the question - what is the greatest possible good for the greater number?

Virtue Ethics

Grounded in Aristotle, virtue ethics seeks to answer the question “who should I be?”. It is organized around helping a person achieve his or her goals, and, to flourish as an individual.

The above three orientations have been suggested as lenses to approach AI in Ethical Considerations in Artificial Intelligence Courses

Exploring Questions around AI and Ethics

The application of ethics to AI in India and more broadly raises a number of different and interesting questions. For example:

- How do you think AI should behave in India? Would this be different in another context? What societal norms are specific to India that you think AI should reflect?
- Should AI be allowed to take autonomous decisions? Are there situations where AI should never take autonomous decisions? Should there always be a kill switch? Should a human always be in the loop?
- Should AI be given personhood?
- Who would push the ‘red’ button fist? An AI or a human?
- Is there a difference between a corporate using AI? Government? A solution?
- Should children be allowed to use AI?
- Who should be held responsible if harm arises from a decision taken or augmented by an AI?
- Can we trust AI? Is there a way for programmers to develop AI to demonstrate trust?

Principles and Frameworks at the International Level

Principles, guidelines, and legal provisions have emerged at the international level as ethical frameworks for guiding the development and use of AI. These have emerged from civil society and academia, industry, standards setting bodies, and are beginning to emerge from governments. Some of these include:

- Academia/Civil Society
  - Asilomar AI Principles
  - Global Union - ten principles of Ethical AI
- Standards Setting Bodies
  - IEEE Global Initiative on Ethics and Autonomous and Intelligent Systems and Ethically Alined Design
  - British Standards Institute BS 8611 - Ethical Design and application of robotics and robotics systems
- Government
  - UK House of Lords Select Committee on AI
  - GDPR Article 22
- Companies
  - Google
  - Microsoft
  - Accenture

Principles and Frameworks at the International Level

Common themes running through these principles, frameworks, standards etc include:

- AI for common and public good
- Placing humans first and well being
- Transparency and Explainability
- Safety and Security
- Responsibility and Liability
- Accountability and Oversight
- Privacy
- Fairness, bias, and discrimination

Examples of Ethics Use Cases at the International Level

- Google and Project Maven: More than 3,100 employees opposed the involvement of Google in offering AI driven solutions to analyse drone footage. Questions were raised of whether or not Google 'should be in the business of war' and urged Google to commit to not building warfare technology. (See: The Business of War: Google Employees Protest Work for the Pentagon)
- Microsoft and I.C.E: Employees at Microsoft voiced concern in Microsoft offering processing data and to ICE. In a letter to the New YOrk Times signatories to a letter noted that they refused to be complicit and demanded that Microsoft take an ethical stand. In response - Microsoft published a statement in which, among other things, they noted the importance of the government regulation in this space and raised a series of questions such as:
    - Should law enforcement use of facial recognition be subject to human oversight and controls, including restrictions on the use of unaided facial recognition technology as evidence of an individual's guilt or innocence of a crime?
    - Similarly, should we ensure there is civilian oversight and accountability for the use of facial recognition as part of governmental national security technology practices?

- What types of legal measures can prevent use of facial recognition for racial profiling and other violations of rights while still permitting the beneficial uses of the technology?
- Should use of facial recognition by public authorities or others be subject to minimum performance levels on accuracy?
- Should the law require that retailers post visible notice of their use of facial recognition technology in public spaces?
- Should the law require that companies obtain prior consent before collecting individuals' images for facial recognition? If so, in what situations and places should this apply? And what is the appropriate way to ask for and obtain such consent?
- Should we ensure that individuals have the right to know what photos have been collected and stored that have been identified with their names and faces?
- Should we create processes that afford legal rights to individuals who believe they have been misidentified by a facial recognition system? (see: Facial recognition technology: The need for public regulation and corporate responsibility)

Learnings from Case Studies: Finance - Uses of AI

- **Customer Interface**: chatbots and concierge apps
- **Customer Insights and Personalization**: YayPay uses previous payment habits and behaviours to predict customer behaviour. Accenture and Grameen Foundation have worked together to develop an application that leverages AI technologies - the Emotional Analytics for Social Enterprises to gain better insights into the emotional and cognitive state of clients by drawing on video and audio inputs to help microfinance advisors better understand when to engage and not to engage.
- **Business Strategy Insights**: Real time insights into internal operations and external market dynamics.
- **Credit Scoring and Loan Decisions**: Loan frame uses AI and machine learning to examine a borrower profile and evaluate their creditworthiness.
- **Fraud Detection and Risk Management**: Bombay Stock exchange has been using AI assisted solutions for rumour detection as a way of reducing information asymmetry.
- **Algorithmic Trading**: Trade Rays provides user friendly algorithmic trading services
- **Transactions**: Niki.ai working with HDFC to offer a conversational interface to streamline the transaction process.

  (See: Center for Internet and Society, AI in the Banking and Finance Industry in India)

Finance Sector - Ethical Considerations

- **Privacy**: AI having access to sensitive personal data and additional new data points.
- **Security**: Creation of honeypots and poor security standards in AI solutions
- **Accountability in algorithmic decision making**: This is addressed only in a limited sense through the SEBI guidelines which are concerned with the processes that work with predefined algorithms and analytics for automated trading systems and robo advisors.
- **Liability**: Who will be held liable for financial harm based on a decision taken or augmented by AI? Currently there are provisions in the consumer protection Act but when services are provided by an entity separate from the one involved in creating the AI system - the issue of liability may be more complicated. As an indication the working group on fintech and digital banking that the onus of consumer protection likes with the fintech companies.
- **Bias and discrimination**: This is a concern particularly in the context of credit scoring - an individual may be denied a loan. India currently does not have non-discrimination provisions that pertain to the financial sector such as in credit scoring.
- **Profiling and nudging**: Particularly in the use for personalization and raises questions about free will.
- **Human AI interaction**: Chatbots are a predominant technology used in the finance sector. Yet, how are the AI solutions accounting the diversity in India including in languages, digital skills, and cultural norms.

  (See: Center for Internet and Society, AI in the Banking and Finance Industry in India)

Learnings from Case Studies: Healthcare - Uses of AI

- **Hospitals**: IBM's Watson for Oncology is being implemented in Manipal. As a note - Watson has come under fire at a global level for being a human driven engine masquerading as an artificial intelligence. Concerns that have been voiced include the lack of an independent study, lack of follow up to understand if recommendations help patients, and that the solution is trained on data that do not reflect the diversity of cancer patients across the world.
- **Pharmaceuticals**: AI is being used in drug discovery to scan through available lit on a particular molecule for a drug as well as to automate the pharmaceutical supply chain management.
- **Diagnostics**: Google, IBM and a host of startups use AI in offering analytical or diagnostic services including mental health. For example, Wysa uses chatbots that provide mental health support where a person can chat anonymously with an AI -enabled system. The chat bots do not provide diagnosis - instead they refer an individual to a doctor.
- **Equipment and Supplies**: Ten3T has created a wireless patch that can be work by heart patients. This patch continuously monitors vitals and transmits this data via the cloud and can be tracked by doctors in real time.
- **Medical Insurance**: Machine Learning is able to automate claims of management by analysing vast amounts of data in less time. For example ICICI Lombard uses chatbot platform MyRA to sell insurance policies.
- **Telemedicine**: Microsoft has teamed up with the Government of Telangana to use cloud based analytics for the Rashtriya Bal Swasthy Karyakram program by adoption MINE to reduce avoidable blindness in children.

  (See: Centre for Internet and Society: Artificial Intelligence in the Healthcare Industry in India)

Healthcare Sector - Ethical Considerations

- **Cultural acceptance and trust**: Startups in India have found cultural acceptance and trust of AI solutions to be a barrier. An aspect of this will be to define acceptable behaviour and norms around informed consent and standards for AI driven medical acceptance.
- **Data security and privacy**: Ensuring the security and privacy of health information that is collected and processed by AI. Informed consent and child consent are particularly important. The Supreme Court in Samira Khli v. Dr. Prahbha Machanda recognized that many patients in India fall below the poverty line without access to ready medical care and thus have no choice but to accept any treatment without question. The enthusiasm to work with new technologies can lead to questions around ethics - for example Google's DeepMind work with the Royal Free hospital in london led to criticism that users were not properly informed about how and what data would be shared with Google.
- **Data Integrity**: There are specific cultural biases in India such as caste and sexuality that can be carried forward in health data sets.
- **Explainability**: The explainability of a decision taken by AI is particularly important in the healthcare sector where diagnosis and treatment need to be backed by a solid chain of reasoning to earn patient trust. Recommendations on how to start addressing explainability in healthcare from Accenture include:
    - What is the range of factors that AI bases its decisions on?
    - What are the desired outcome and how are needs prioritised as the AI makes its decisions?
    - How is the acceptable level of responsibility and liability for the stakeholders in AI determined?
    - Is the logic of a decision taken or recommended by an AI system clear and inline with best practice? Can this be understood by a doctor taking or impacted by the decision?

  (See: Centre for Internet and Society: Artificial Intelligence in the Healthcare Industry in India)

Healthcare Sector - Relevant Regulation

- **Medical Council Professional conduct, etiquette and ethics regulations 2002**: MCI lays down professional ethical standards of interaction of doctors with patients which include patient confidentiality and disclosure of prognosis.
- **Electronic Health Records Standards 2016**: Touch on data ownership, data access, changes to data, disclosure of health information, access to records by courts and government authorities, and places the responsibility of healthcare providers to be responsible for storage of patient information and ensuring removal of personal identifiers
- **Medical Devices Rules 2017**: Drafted with the intention of distinguishing between medical devices for the purpose of regulation - distinguishes between three classes - low, low moderate, moderate high, and high.
- **Draft Digital Information Security in Healthcare Act**: Provides for electronic health data privacy, confidentiality, security, and standardization and provides for establishment of a National Digital Health Authority and health information exchange.

  (See: Centre for Internet and Society: Artificial Intelligence in the Healthcare Industry in India)

Potential Distinguishing Principles to Guide Ethical Deterimations in AI

- **Modeling Human Behaviour**: An AI solution trying to model human behaviour, as in the case of judicial decision-making or predictive policing may need to be more regulated, adhere to stricter standards, and need more oversight than an algorithm that is trying to predict 'non-human' phenomenon such as traffic congestion or weather patterns.
- **Human Impact**: an AI solution which could cause greater harm if applied erroneously-such as a robot soldier that mistakenly targets a civilian requires a different level and framework of regulation than an AI solution designed to create a learning path for a student in the education sector and errs in making an appropriate assessment. Furthermore, situations where human impact is involved can take more complex strategic decision making that an AI may not necessarily be able to carry out. For example, in a war scenario, though all indicators may point to 'pushing the red button' a human may not take this decision.
- **Public Interest**: AI solutions whose primary users are state agents attempting to discharge duties in the public interest such as policemen, should be approached with more caution than those used by individuals such as farmers getting weather alerts.

  (See: Centre for Internet and Society: Artificial Intelligence in the Governance Sector in India)

Developments in India

- Niti Aayog National AI Strategy
- Report of Task Force on Artificial Intelligence
- India to chair the UN GGE on lethal Autonomous Weapons Systems
- Srikrishna Committee Draft Data Protection Bill
- Department of Defense Production set up a Committee on AI to look into the use of artificial intelligence in defence
- BIS Committee for standardization in AI

NITI Aayog National AI Strategy and Ethics

The National AI Strategy for AI noted the following in the context of ethics and AI:

- Privacy concerns in AI are centered around inappropriate use of data for personal discrimination. Misuse and mass collection of data can create an unfair competitive advantage. Some suggestions include establishing a data protection framework with legal backing, establishing sectoral regulatory frameworks, benchmarking national data protection and privacy laws with international standards, encouraging self regulation, investing and collaborating in privacy preserving AI research, and spreading awareness.
- Security and accountability of AI are further concerns that need to be addressed.
- AI systems can reduce bias. Along these lines - proposed that a solution would be to identify inbuilt biases and assess their impact and find ways to reduce the bias This reactive approach, use case based, may suffice until neutrality can be ensured.
- More research is needed into transparency and AI. There is the risk that too much disclosure may induce companies to change their behaviour and game the system.
- Negligence tests for damages caused by AI software are prefered as opposed to strict liability . This includes conducting damage impact assessments. Safe harbors need to be formulated to insulate or limit liability
- India can follow the UK and invest in a new Centre for Data Ethics and Innovation Centre.

  (See: Niti Aayog National AI Strategy)

Report of the Task Force on Artificial Intelligence

The Report of the Task Force on Artificial Intelligence said the following on ethics and AI:

- AI systems need explainable behavior, demonstrable either explicitly or statistically
- AI systems need to be engineered for safety and security
- AI systems need to be audited rigorously to ensure non-contamination by human bias
- Legal provisions that are applicable to human users of AI systems should continue to apply to autonomous machines
- Specific liability provisions may have to be worked out for certain categories of machines
- AI developers must ensure that relevant and applicable legal provisions are respected during the development of AI systems
- The government needs to start thinking about rights and responsibilities of autonomous entities.
- AI systems must be transparent and their learning must be verifiable and auditable. All relevant test and evaluation data must be shared with the users. AI researchers should ensure that an independent social ethics panel screens research proposals
- Data on which AI systems rely must continue to be protected at least to the same level as the original database
- New industrial standards need to be created for robots in India
- Complete autonomy cannot be given to weaponized platforms due to their potential unpredictability, and inability to detect friend or foe. There might also be issues related to the law of armed conflict.
- All aspects of human / robot interaction need to be well thoughts out. The government need to urgently foster interdisciplinary research on AI/human interaction

  (See: Report of Task Force on Artificial Intelligence)

Srikrishna Draft Data Protection Bill 2018

- Specifically addresses companies using emerging technologies including AI through the principle of privacy by design and data protection impact assessments.
- The definition of harm encompasses a number of harms that can result from AI including: financial loss, unemployment, discrimination, and denial of service.
- The principle of data quality can be interpreted as a mechanism towards ensuring that data does not contain or result in bias.
- The restrictions on cross border transfer of personal data and localization requirements could impact companies leveraging AI and storing the data in a cloud.
- In contrast to the GDPR, the Srikrishna Bill does not contain a requirement on data controllers to provide the logic and consequence of an automated decision and does not provide the right of individuals to request to not be subject to a decision based solely on automated decision making.

  (See: The Srikrishna Committee Data Protection Bill and Artificial Intelligence in India)

Further Resources

- The Ethics of Artificial Intelligence - Machine Intelligence Research Institute
- Ethical Artificial Intelligence - An Open Question - Alice Pavaloui and UTku Kose
- The Ethics and Governance of Artificial Intelligence MIT Media Lab
- Ethics of Algorithms Centre For Internet and Human Rights
- Robots: ethical by design - Gordana Dodig Crnkovic and Baran Çürüklü
- Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency - Zeynep Tufekci
- Research priorities for robust and beneficial artificial intelligence - Russell, Stuart, Daniel Dewey, and Max Tegmark
- The nature, importance, and difficulty of machine ethics - James H.Moor
- Infinite Ethics - Nick Bostrom