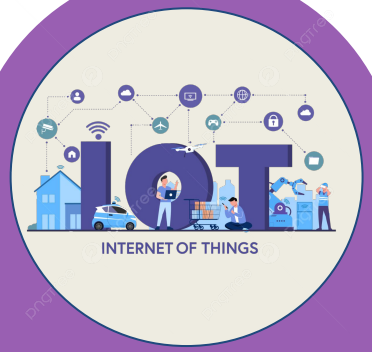
Ethical uses of Data-driven Technologies

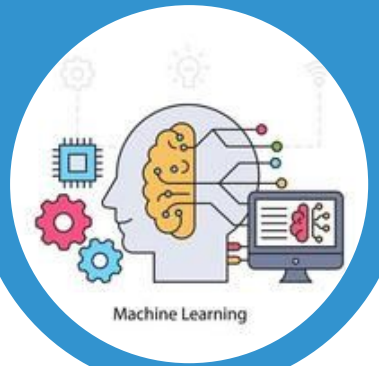
Lecture 7

Technologies suitable for ethical practices



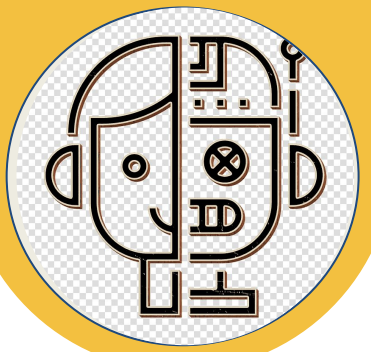
Internet of Things

Machine Learning



Big Data

Artificial Intelligence



Necessity of ethics in emerging technologies

Generate Fake Information

Emerging technologies are much more capable to figure things out, their ability to generate plausible yet fake information.

Repopulate biases

Emerging technologies have a potential to re-perpetuate biases or errors due to their lack of ethical ideologies.

Ethical Considerations for Data Science

01 **What are some ethical issues that should be considered by leaders at organizations relying on data science?**

Ensure that technologies are transparent, accountable, and minimally biased. Data about us should not be incorrect, biased, or taken from us without permission.

02 **How can data be collected or used in a way that doesn't respect our agency?**

For example, data might be collected from sources we didn't intend, such as an app gathering up data copied in the clipboard. Or, data might be used in a way that we didn't authorize or foresee, i.e., data on our purchases being sold to a credit scoring agency.

03 **What kinds of ethical considerations for data science should public be focused on?**

04 **What can organizations do to make sure they collect and use data in a transparent, accountable, and minimally biased way?**

- Processing of data (aggregation, or clean-up of erroneous data) should be handled with care.
- Collected data should not be massaged, or manipulated in any way, or the interpretation of that data altered from the truth, as best as that can be verified.

Benefits of Ethical Data Science

01 Better ways of understanding our societies and inequities

02 Empower effective leadership with more precise and sophisticated information and decisions

03 Mutual data sharing can be helpful for all parties

04 Better predict tradeoffs of proposed policies - **Cost benefit analysis**

05 Both macro and micro views of situations can be seen

06 Future developments in technology can be facilitated

How does a Data Scientist keep ethics a priority

01 Should understand the **provenance** of data. Was it collected in a fair way, with appropriate consent? Has it been sufficiently anonymized to protect privacy? Could data be reverse engineered or uncovered through cross-correlation? What risks to individuals might arise if this process fails?

02 Should be mindful of the **biases** in their datasets. Has care been taken with sampling? Are any pre-trained models potentially bringing their own biases into one's ecosystem?

03 It's easy to lie using statistics. Data can be 'tortured' to tell us almost anything we want to hear from it. Therefore, need to be careful against any such requests that **attempt to massage data into a format that does not reflect what it would tell us at face value**.

04 Should consider if the **intended purpose** of the data is ethical, or if it might potentially be misused in an unethical manner.

Data Science Ethics In Practice

[Ten Data Science Ethics Questions]

[An Ethical Checklist for Data Science]

Problems with Machine Learning

01 These models are created by someone who provided the data, selected the algorithms, and supervised the learning.

02 Many AI systems are used to profile people or to perform other tasks that may have a direct impact on people's lives.

03 Sometimes the creators of these algorithms often may have no idea how or on what basis a decision has been made.

04 Due to the inherent randomness in the statistical models, the same input may not necessarily give the same output each time.

Problems with Machine Learning

05 As they are dependent upon data, any update to the data as well as any update to the model can change the function or the results generated.

06 Due to the nature of the models it can be very challenging to understand which specific piece of data within a dataset contributes towards an output.

Concerns about these issues have led to legislation, for example, data protection regulations provide users the right to demand access to all data held about them.

Ethical Considerations for AI

Why do we need to consider ethics when we're talking about artificial intelligence?

We increasingly hand off a great many of our decisions to machine intelligences, often without thinking about it, or even realising that we are doing so. For example: Google Map.

Do ethical AI professionals have the same or different worries than the public about artificial intelligence?

Ethical AI professionals should have a better grasp of the less obvious ways that AI can be used in undesirable ways. For example, it may re-perpetuate existing biases, either by accident, or by design. It can potentially be used as a tool for silent exclusion, or even targeted harassment.

Benefits of Ethical AI

What kinds of benefits can AI bring to society?

Ethics should be built into technology as **part of a holistic process,** not as an afterthought. It's very important that ethical considerations are applied right **from the inception of a product or service.** Trying to re-engineer things after the fact to incorporate ethical constraints can be very challenging and expensive, and may permanently weaken the ethical integrity of a project, and the team behind it.

How does an AI engineer keep ethics a priority

An AI engineer must bear in mind many different ethical aspects of the role -

01 What is the **intended purpose** of the technology? Is that ethical, and could it potentially serve dual uses?

02 Who is the **beneficial owner** of such technology? Are they reliable and trustworthy organizations?

03 What **safeguards** are in place to help to provide the greatest possible transparency, explainability, accountability, and management of bias, **whilst balancing needs** for end user and organizational privacy?

04 How can one take steps to **minimise bias, protect privacy, provide transparency, and preserve accountability** on this particular scenario?

What kinds of questions does an ethical AI Engineer have to ask about the data they are using?

01 Was data **gathered** in an ethical way? This is especially important to ask if it might have been scraped from other data sources.

02 If the data comes from another organization, it's important to do some **due diligence** as to the terms under which it was collected, and from whom. Was some of it potentially collected from vulnerable individuals or minors?

03 How does the system work **across jurisdictions**? Consider a situation where the data came from one place, the processing is done in a second place, the servers are in a third place, the end user is in a fourth place, and they are interacting with someone in a fifth place?

04 Where is data being t**ransferred from and to**? Who holds responsibility for that data and its security, and are they trustworthy? Who or what owns or controls that entity – are they a subsidiary of another organization (or several in a hierarchy)?