Ethics in Data Science: Navigating the Crossroads of Power and Responsibility

Ritambhara Jha[1] and Ananya Jha[2]

[1]Ritambhara Jha, Master Student, MS in Data Science, Rochester Institute of Technology, NY, USA

[2]Ananya Jha, Partner Lead at Google, USA

A B S T R A C T

Data science has emerged as a game changing force, influencing everything from scientific discoveries to commercial choices and daily interactions. Great power, however, comes with great responsibility, and the ethical implications of data science activities must be carefully evaluated. This research paper digs into the ethical landscape of data science, examining important obstacles and potential answers to the confluence of technical growth and social responsibility. This will lead to building a strong ethical foundation for data science practice, accountability and openness.

Keywords: Ethics, Data Science, Transparency, Privacy, Bias Mitigation, Governance, Regulatory Compliance, Responsible AI

1. Introduction

The explosion of data science skills has culminated in ethical concerns and complications that require attention and proactive resolution.

So, how should businesses begin to consider ethical data management? What safeguards can they put in place to guarantee that consumer, patient, HR, facility, and other types of data are used responsibly across the value chain from collection to analytics to insights?

Data Ethics in simple terms is a branch of study that evaluates the ethical issues that surface during the direct or indirect usage or consumption of the data.

It consists of three dimensions:

1. Consequentialist ethics holds that an agent is ethical if and only if it analyzes the implications of each option and selects the one with the greatest moral outcomes. It is also known as utilitarian ethics since the decisions that arise frequently strive to achieve the best aggregate effects.

2. Deontological ethics: a participant is ethical if and only if it upholds obligations, duties, and rights in specific contexts.

Agents who practice deontological ethics (also known as duty ethics or obligation ethics) follow established societal standards.

3. Virtue ethics holds that an agent is ethical if and only if it behaves and thinks in accordance with certain moral principles (for example, courage, fairness, and so on). Agents who practice virtue ethics should have an inner desire to be seen favorably by others.

Ethical dilemmas are circumstances in which any possible option will violate some recognized ethical ideal, but a decision must be made[1]. As Data Scientists, we commonly interact with large amounts of data generated by people. Thus, it is our responsibility to keep private data secure and utilize it appropriately. To effectively integrate human values such as justice and equality into data-driven technology, we must first understand the underlying human and societal systems.

2. Related Work

Given that data science is intrinsically tied with computers and that computing has a longer history than data science, it is worthwhile to quickly explore computing ethics[2].

While there are examples of high-level attention to the link between computers and ethics dating back to the 1950s, a larger debate only emerged in the 1980s and 1990s. During this time, computer ethics evolved into an applied ethics area[3].

Ethics, according to[4], is a normative practical philosophical discipline that governs how one should treat others[4]. The authors[5], laid out a collection of representation schemas for structuring AI ethical talks[5]. It consists of the following elements:

1. Features: denoting the presence or absence of factors (e.g., harm, benefit) with integer values

2. Duties: denoting an agent's responsibility to minimize/ maximize a given feature as an integer tuple

3. Actions: denoting whether an action satisfies or violates certain duties as an integer tuple

4. Cases: used to compare pairs of actions on their collective ethical impact; and

5. Principles: a tuple of integer tuples representing the ethical choice among various activities.

3. Implementation

Government agencies are also working on projects, such as the White House Report on the Future of Artificial Intelligence (Executive Office of the President 2016) and a draft report on robotics and law by the European Union's Committee on Legal Affairs (Directorate-General for Internal Policies 2016). Such government work typically focuses on broader social concerns such as employment, funding, and economy, all of which have ethical implications[6].

Professional organizations such as the IEEE's Standards Association Global Initiative for Ethical Considerations in the Design of Autonomous Systems have a vast number of research strands prompted by their byline, 'Values By Design.' This is an ongoing initiative to provide industry standards as well as discussion materials on a variety of issues (http://standards.ieee. org/develop/indconn/ec/autonomous_systems.html)[6].

Large and small corporations have initiatives as well, such as the Partnership on Artificial Intelligence to Benefit People and Society, which includes Amazon, DeepMind, Facebook, Google, IBM, and Microsoft (https://www. partnershiponai.org/); and the non-profit initiative AI Austin, which includes a university, City Council, and business (https://www.ai-austin.org/).

Individuals in the professions are also working on projects such as research into prejudice in AI recruiting and the creation of applications to study bias in algorithms (Clark 2016)[6].

The Data Ethics Checklist is as follows[7]:

- Have we discussed how this technology can be used or abused? [SECURITY]

- Have we verified that our training data is fair and representative? [FAIRNESS]

- Have we investigated and comprehended potential causes of bias in our data?[FAIRNESS]

- Is our staff diverse in terms of perspectives, backgrounds, and types of thought?[FAIRNESS]

- What type of user consent do we need to gather in order to utilize the data?[PRIVACY/TRANSPARENCY]

- Do we have a system in place to obtain user consent? [TRANSPARENCY]

- Have we made it obvious to users what they are agreeing to? [TRANSPARENCY]

- Is there a restitution mechanism in place if persons are injured as a result of the results? [TRANSPARENCY]

- Can we disable this program in production if it is misbehaving?

- Have we tested for fairness in relation to various user groups? [FAIRNESS]

- Have we evaluated for differences in mistake rates between user groups? [FAIRNESS]

- Do we test and monitor for model drift in order to guarantee that our software remains consistent over time? [FAIRNESS]

- Do we have a strategy in place to protect and secure user data? [SECURITY]

- Answering the questions from the above checklist, before starting any data related project, is a proposed approach to secure the dataset and ensure its ethical consumption.

4. Conclusion and Future Work

As data science continues to impact our society, the ethical issues it raises deserve our immediate attention. We can leverage the enormous potential of data science for good by adopting an ethical perspective, ensuring that technology innovation serves the best interests of society as a whole. The road toward a more ethical data science is ongoing, and continued commitment to responsible methods is critical to unlocking a future in which technology empowers everybody without sacrificing human values. Data, when handled ethically, will help us make choices and generate significant change in your business and throughout the world.

In the future, we should focus the research on how this research implementation could be integrated into a data science program at a college level. These ethics concepts should be integrated into existing classes by developing key questions (based on our ethical considerations), thereby providing upcoming data analysts with a basic toolkit to help them think about these challenges within the context of a data science class project.

By navigating the complexities of ethics in data science with care and commitment, we can ensure that this powerful technology contributes to a more just, equitable, and sustainable future for all.

5. References

1. Kirkpatrick K. The moral challenges of driverless cars. Commun ACM,2015;58:19-20.

2. Saltz JS, Dewar N. Data science ethical considerations: A systematic literature review and proposed project framework. Ethics Inf Technol, 2019;21: 197-208.

3. Stahl BC, Timmermans J, Mittelstadt BD. The ethics of computing: A survey of the computing-oriented literature. ACM Computing Surveys, 2016;48: 55.

4. Cointe N, Bonnet G, Boissier O. Ethical judgment of agents' behaviors in multi-agent systems. AAMAS, 2016; 1106-1114,

5. Anderson M, Anderson SL. GenEth: A general ethical dilemma analyzer. AAAI, 2018;9: 337-357.

6. Boddington P. Towards a code of ethics for artificial intelligence. Springer, 2017

7. Loukides M, Mason H, Patil DJ. Ethics and data science. O'Reilly Media, 2018.