

The Ethical Dilemma of AI

Catherine Visscher, Nisarg Jhaveri, Aaron Zhao, Zach Gooding, William Yu

07-31-2023

Contents

1	Introduction	7
2	About Us	9
2.1	Nisarg Jhaveri	9
2.2	Catherine Visscher	9
2.3	Zach Gooding	10
2.4	Aaron Zhao	10
2.5	William Yu	11
3	Hardware	13
3.1	Hardware Requirements for NLP Developing	13
3.2	Case Analysis	13
4	Energy	15
4.1	Facts on Current Generative AI Energy Usage	15
4.2	Why Training Generative AI Uses So Much Energy	15
4.3	How Companies Can Make AI Models Greener	15
5	Regulations	17
5.1	US Federal Regulations	17
5.2	Regulation Abroad	17
5.3	Challenges of Regulation	17
5.4	Benefits of Regulation	17

6	Censorship	19
6.1	“Pausing AI” / moratorium	19
6.2	International arms race	19
6.3	Adapting to human values	19
7	Data	21
7.1	Diversity/Biases of Data	21
7.2	Synthetic vs Real Data	21
7.3	Ethical Dilemma of Web Scraping	21
8	Bias	23
8.1	In LLMS	23
8.2	Addressing the Issue	23
8.3	MIT ChatGPT Bias Study	23
9	Alignment	25
9.1	What values is AI supposed to align to?	25
9.2	How can we generalize values despite coming from different training data?	25
9.3	Long-term values	25
9.4	Algorithmic Improvement	25
10	Alignment	27
10.1	Scaling Laws for Neural Language Models	27
10.2	Privacy Enhancing Technologies - Mithril Securities BlindAI . . .	27
10.3	Auto ML	27
11	Personalized Large Language Models (LLMs)	29
11.1	What is hugging face, and why do we need to know about it? . .	29
11.2	Benefits of Personalized LLMs	29

<i>CONTENTS</i>	5
12 Open Source Software (OSS) in AI	31
12.1 What is Open Source Software?	32
12.2 Benefits of Open Source LLMs (https://huggingface.co/blog/os-llms)	32
12.3 Examples of Open-Source Projects for LLMs (Cut some out) . .	32
13 Conclusion - Challenges and Future Directions	33
13.1 Data Diversity (delete)	33
13.2 Privacy and Security	33
13.3 Ethical Frameworks	33
14 Reference	35

Chapter 1

Introduction

This website is a learning tool for the Ross community, including students, alumni, faculty, and staff, on the ethical issues of how we produce and generate artificial intelligence. The information book will first begin with team introductions then discuss a general introduction into the current production/generation of artificial intelligence.

After these introductions, we will transition into the main section of the book, discussing several areas of ethical concern as it relates to artificial intelligence. The first chapter will focus on the significance of data in AI and address concerns on how AI tools are generating/gathering data to feed to users without explicit permission. After discussing data gathering, the following chapter will focus on the limitations of hardware in AI and large language model development. Next, privacy concerns related to AI and language models will be discussed and we will talk about the need for AI companies to handle personal data responsibly. After that, bias issues will be addressed, explaining how data sources and training processes can lead to biased AI models.

The next chapters will talk about the significant energy requirements for training and running these large language models as well as the challenges and benefits of regulations currently in place (and future regulations) for AI and large language model development. We will also talk through the concerns related to censorship in AI models and discuss changes in ChatGPT due to these censorship measures.

The last three chapters will talk about the ethical concerns related to AI in alignment, personalized large language models, and open-source software. Alignment will talk about how models have the potential to be politically or ideologically aligned. Personalized LLM and open-source software will touch on the benefits and challenges of customization and explore the benefits and risks of open-sourcing models like ChatGPT.

We recognize the research and resources our team has provided as surface-level discussions of these topics, but we hope this tool can be a baseline for future

students of the Ross community as we move into a world adapting to rising artificial intelligence.

Chapter 2

About Us

Hi there, we are the Insight Titans! We are the Masters of Business Analytics Class of 2024 and we are here to

2.1 Nisarg Jhaveri

Hey-o! I am Nisarg Jhaveri, an enthusiast from Boston, MA, and an auditor by profession. I completed my undergrad at the University of Massachusetts Lowell in Finance and Accounting with a minor in Psychology. I did five internships during that time frame to explore different industries. I graduated back in 2022, after which I worked at EY for a year with their Wealth and Asset Management team, but I kept itching to go back to school to learn more about programming, a topic that I couldn't fully explore. I decided to join the Ross School of Business because of this and hope to make the most out of it to truly leverage data and create value while working on a few personal projects. P.S. I am a very social person so feel free to reach out, I would love to chat/help :)

2.2 Catherine Visscher

Hello! I am Catherine Visscher, a Michigan native from Livonia, Michigan. I graduated from the University of Michigan College of Engineering Winter 2023 with a B.S.E in Chemical Engineering and am currently a Master's of Business Analytics student at the Stephen M. Ross School of Business. One thing my undergraduate degree has taught is how to break large, complex processes into more manageable components. I have experience as an engineer intern within the utility sector working at Consumers Energy, with work focused in root cause analysis and process optimization.

I am entering my fifth year as a member of the Women's Track and Field team at Michigan as a pole vaulter. Within the last year, we have been fortunate enough to win back-to-back Big Ten Conference Titles and I cannot wait for another season with this amazing team! In my somewhat limited free time, I enjoy reading and spending time with friends and family.

I hope the information we've gathered on artificial intelligence and how it is being generated is helpful to the current and future Ross community!

2.3 Zach Gooding

Hello everyone! My name is Zach Gooding, and I am originally from Westmont, Illinois, a Western Suburb of Chicago. I graduated from Augustana College in the Spring of 2023 with majors in Data Analytics and Business Administration with concentrations in Marketing, Management, and MIS. I was also a four year member of the Varsity Baseball team at Augustana and was a part of back to back Conference Championship teams. My undergraduate experience made me realize I was just beginning to scratch the surface in the field of Data Analytics, so I am thrilled to be taking classes at Michigan Ross to further my education.

My prior work experience is as a Marketing and Data Analytics Intern for the NTT INDYCAR Series and Canes Baseball in their Illinois Division. In my free time, you can find me watching the Cubs game, listening to a lot of music, and experiencing everything Ann Arbor has to offer. I hope you find our information useful on Large Language Models as we enter this new age of Data.

2.4 Aaron Zhao

Hey there, I'm Aaron! I studied in San Diego, California for my undergraduate degree, and since then, I've taken four epic trips from the West Coast to the East Coast, exploring new places and cultures along the way. When I'm not studying for my degree, I love working with wood and 3D printing to make cool and useful stuff that brings joy to myself and others. My creative pursuits extend to the kitchen, where I enjoy experimenting to come up with delicious and creative meals. In my free time, I cherish simple pleasures like indulging in cookies and creme ice cream or savoring the delightful sashimi plate. Psychology has always fascinated me because understanding why people do what they do feels like solving an exciting puzzle. Alongside my academic and creative passions, I have three adorable and playful chinchillas, which inspire me to craft special things for them to enjoy. Moreover, sharing knowledge is something I'm truly passionate about. During my undergraduate years, I loved being a teaching assistant, helping others learn and grow. It's an incredibly rewarding experience to see others flourish through my help.

2.5 William Yu

Hello everyone, my name is William Yu and I am from China. I graduated from Indiana University Bloomington in 2022 majoring in Information Systems and Business Analytics. During my undergraduate studies, I had four internships with some consulting and technology firms such as IQVIA and TikTok. I also worked as a research assistant with a professor at the Kelley School of Business, Indiana University, focusing on the impact of dismantling coal power plants on the surrounding environment, economy and housing prices. These experiences gave me valuable skills with data analysis helped me practice my knowledge with some programming and data analysis tools, such as Python, SQL and R. After graduation, I want to find a quantitative analyst position in some financial firms because I want to be at the forefront of leveraging cutting edge quantitative techniques and mathematical models to solve complex financial problems,

During my free time, I am enthusiastic about cycling and triathlon. As a two time Ironman 70.3 finisher, I enjoy pushing myself beyond what I thought possible and the sense of accomplishment after every long-distance endurance workouts.

Chapter 3

Hardware

3.1 Hardware Requirements for NLP Developing

3.2 Case Analysis

Chapter 4

Energy

- 4.1 Facts on Current Generative AI Energy Usage
- 4.2 Why Training Generative AI Uses So Much Energy
- 4.3 How Companies Can Make AI Models Greener

Chapter 5

Regulations

5.1 US Federal Regulations

5.2 Regulation Abroad

5.3 Challenges of Regulation

5.4 Benefits of Regulation

Chapter 6

Censorship

6.1 “Pausing AI” / moratorium

6.2 International arms race

6.3 Adapting to human values

Chapter 7

Data

7.1 Diversity/Biases of Data

7.2 Synthetic vs Real Data

7.3 Ethical Dilemma of Web Scraping

Chapter 8

Bias

8.1 In LLMS

8.2 Addressing the Issue

8.3 MIT ChatGPT Bias Study

Chapter 9

Alignment

- 9.1 What values is AI supposed to align to?
- 9.2 How can we generalize values despite coming from different training data?
- 9.3 Long-term values
- 9.4 Algorithmic Improvement

Chapter 10

Alignment

10.1 Scaling Laws for Neural Language Models

10.2 Privacy Enhancing Technologies - Mithril
Securities BlindAI

10.3 Auto ML

Chapter 11

Personalized Large Language Models (LLMs)

11.1 What is hugging face, and why do we need to know about it?

11.1.1 Democratizing Access to Advanced AI Models

11.1.2 Accelerating AI Development

11.1.3 Collaboration and Community

11.1.4 Open-Source Philosophy

11.2 Benefits of Personalized LLMs

11.2.1 Enhanced User Experience

11.2.2 Improved Task Performance

11.2.3 Industry Applications

Chapter 12

Open Source Software (OSS) in AI

12.1 What is Open Source Software?

12.2 Benefits of Open Source LLMs (<https://huggingface.co/blog/os-llms>)

12.2.1 Collaboration and Innovation

12.2.2 Transparency and Accountability

12.2.3 Collective Advancement

12.3 Examples of Open-Source Projects for LLMs (Cut some out)

12.3.1 GPT4ALL

12.3.2 Stability-AI

12.3.3 Free Dolly

12.3.4 Open Assistant

12.3.5 RedPajama

Chapter 13

Conclusion - Challenges and Future Directions

13.1 Data Diversity (delete)

13.2 Privacy and Security

13.3 Ethical Frameworks

Chapter 14

Reference