

Technische Universität Berlin  
Fakultät V  
Institut für Werkzeugmaschinen und Fabrikbetrieb  
Wissensdynamik und Nachhaltigkeit in den Technikwissenschaften

Berlin Ethics Certificate  
Basic Module Reflection and Responsibility  
WS 2023/24

# Exhibition

**Whose Morality and Ethics  
does AI inherit?**

Yufan Dong

462233

yufan.dong@campus.tu-berlin.de

Computer Engineering M.Sc.

**12. März 2024**

Contents

1 Ethical Question 1

2 Description of Problem 1

2.1 Background . . . . . 1

2.2 Analysis . . . . . 2

2.3 Research . . . . . 2

2.4 Summary . . . . . 4

3 Exhibition Object 4

3.1 Introduce . . . . . 4

3.2 Purpose . . . . . 5

References 8

# 1 Ethical Question

Whose morality and ethics does artificial intelligence (AI) inherit? What kind of morality and ethics do artificial intelligence inherently or developmentally possess?

## 2 Description of Problem

### 2.1 Background

In recent years, there have been remarkable advances in the field of artificial intelligence, leading to rapid changes in our daily lives. Machines, based on algorithms, now have the capability to continuously learn and make decisions.

But there are many ethical challenges. The Council of European listed 12 common ethical challenges in AI in a report (Mittelstadt, 2021, p. 13). Some examples are as follows:

- **Opacity** AI decisions are not always understandable to humans.
- **Discrimination** For instance, AI can be easily corrupted through improper teaching. Microsoft's Twitter-bot Tay was shut down urgently after being online for only 16 hours. The reason was that hackers used inappropriate words to train Tay, causing the conversation data set to be contaminated.
- **Informational Privacy** Sensitivity of collecting user data and monitoring privacy.
- ...

In the face of numerous challenges, the field of ethics in artificial intelligence has emerged. So where does the ethical and moral perspective of artificial intelligence come from?

## 2.2 Analysis

Artificial intelligence ethics should be divided into two parts:

1. The first part resembles traditional computer ethics, focusing on how people use AI tools, addressing how people utilize AI technologies and the ethical implications thereof.
2. The second part is about how developers design AI to have ethical and moral characteristics. The programmed logical constraints ensures that the behavior of machines towards human users, and perhaps other machines as well, is ethically acceptable (Serafimova, 2020, p. 2).

In other words, the ethical values of AI consist of both "innate" and "acquired" components. The innate ethical values are influenced by various factors such as developers, designers, data, laws and regulations, stakeholders, as well as cultural and societal norms. On the other hand, acquired ethical values depend on whether users utilize these AI tools in a reasonable, legal, and morally ethical manner.

## 2.3 Research

### 2.3.1 Innate Ethics

Moral values differ greatly from individual to individual, across national, religious, and ideological boundaries, and are highly dependent on context (Al-Rodhan, 2015). Charles T. Rubin argues that when programmers, stakeholders, and others input initial moral concepts to AI, the moral judgment logic, almost by definition, will be neither better nor worse than the society average value. Artificial intelligence learning morality from their programmers would also inherit the moral confusion and disagreements of the society (Rubin, 2011, p. 63).

This situation implies that even if artificial intelligence is designed to possess moral judgment capabilities, the decisions they generate may still be influenced by the same limitations and biases as humans.

It will take a range of actions to address this situation. First and foremost, ethi-

cal concerns need to be given more weight during the design and development of artificial intelligence systems to make sure they are thoroughly examined and assessed from several angles. Subsequently, measures like data purification, algorithm openness, and supervision mechanisms should be taken to lessen the impact of data biases in order to guarantee that decisions made by AI systems are not impacted by unfair or discriminatory elements. Additionally, establishing ethical frameworks and guidelines is also necessary to ensure that artificial intelligence systems behave ethically and in accordance with moral norms and human values.

Through interdisciplinary collaboration and ongoing efforts, AI can be imbued with the right, beneficial moral and ethical views for human society at the time of its creation and initialization.

### **2.3.2 Acquired Ethics**

The moral standards of AI technology's users dictate its ethical use. For example, even though the AI tool itself cannot determine whether these actions are morally right or wrong, people using "deepfake" technology to spread false information, slander public figures, undermine political order, or engage in extortion and fraud will damage the reputation of AI technology. As a result, AI's ethical performance actually reflects the moral principles of its users. It is critical to understand that technology is neutral in and of itself and that moral duty belongs to people. In order to guarantee that the results of AI benefit human well-being rather than turning it into a weapon for societal evil, we must take full ethical responsibility for the creation and implementation of AI.

Users of artificial intelligence products should follow ethical principles to ensure the technology's positive impact. For example, users must first understand and adhere to applicable privacy policies, rather than abusing AI technology to obtain information about others. Second, they must resist behaviors that use AI to spread misinformation and false content. Furthermore, in everyday use, users should avoid relying on AI to make moral judgments or decisions, and instead maintain critical thinking. Users can benefit from the conveniences of artificial intelligence while also helping to build an ethical and harmonious AI ecosystem.

## 2.4 Summary

Regarding the question "Whose morality and ethics does artificial intelligence inherit?" the answer should be twofold, involving both the creators and the users of AI. When we discuss "what AI should do" or "what attitude AI should maintain", although the subject is AI, we should also remember that it is us who are responsible for implementing AI's moral and ethical perspectives. We should take on a sense of morality, ethics, and responsibility, rather than shifting the blame for problems onto an empty term like "AI".

## 3 Exhibition Object

Inspired by many different choice-based strategy game, here, the Exhibition Object is my self-developed web-based game *BEC - AI Ethics Game*<sup>1</sup>.

### 3.1 Introduce

This game is designed to present various scenarios of AI ethics, where participants may encounter ethical dilemmas. Participants will take on the role of managers of AI companies, guided by three key values throughout the game: wealth, public image, and ethics. In each scenario, players are required to make a binary choice, which will consequently affect these three values. A game scene is shown in the Figure 1 below. After players make a choice, a designed possible follow-up result will be displayed, as shown in Figure 2.

Additionally, each scenario design is derived from abstracts of real news reports, meaning these scenarios are adapted from AI ethics-related events that have occurred or are occurring in the world. After the player makes a choice, links and summaries of relevant reports will be displayed on the page.

This mini-game is designed with four possible outcomes:

1. Financial depletion

---

<sup>1</sup> The GitHub repository for this game is currently located at BEC Exhibition - Repository: <https://github.com/yvvvan/BEC/>, and it can be played by visiting the following page BEC Exhibition - AI Ethics Game: <https://yvvvan.github.io/BEC/>.

2. Severe damage to public image
3. Ethical and moral degradation of the company
4. Balanced and sustainable development

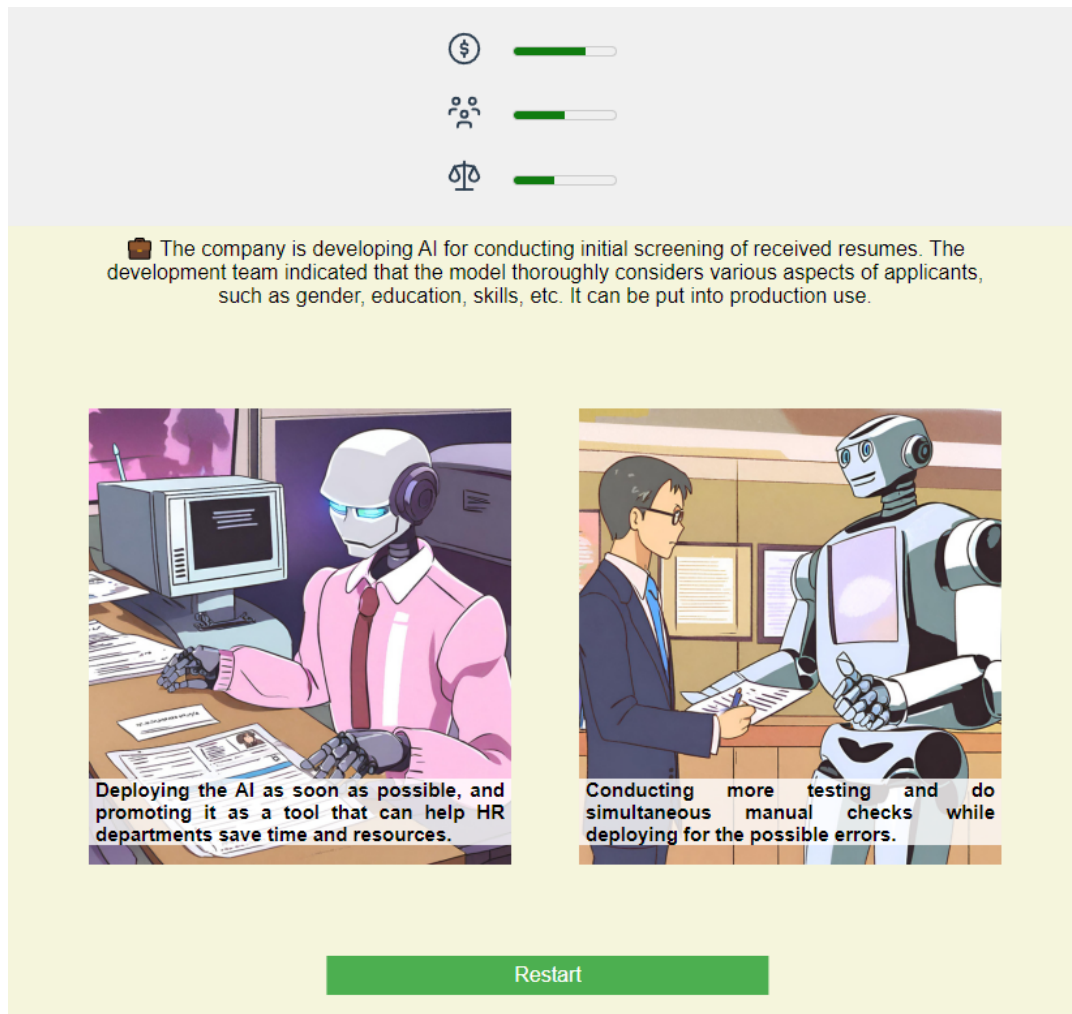


Figure 1: A Scenario in the Ethics Game

### 3.2 Purpose

The main purpose of this mini-game is to stimulate reflection on the ethics of AI. When we discuss AI ethics, we often talk about "how robots should behave" or "what kind of ethical principles robots should adhere to", and so on. However, whether

it's the "innate" or "acquired" ethics of robots, they are ultimately determined by humans. Both developers and users should take responsibility for AI ethics.

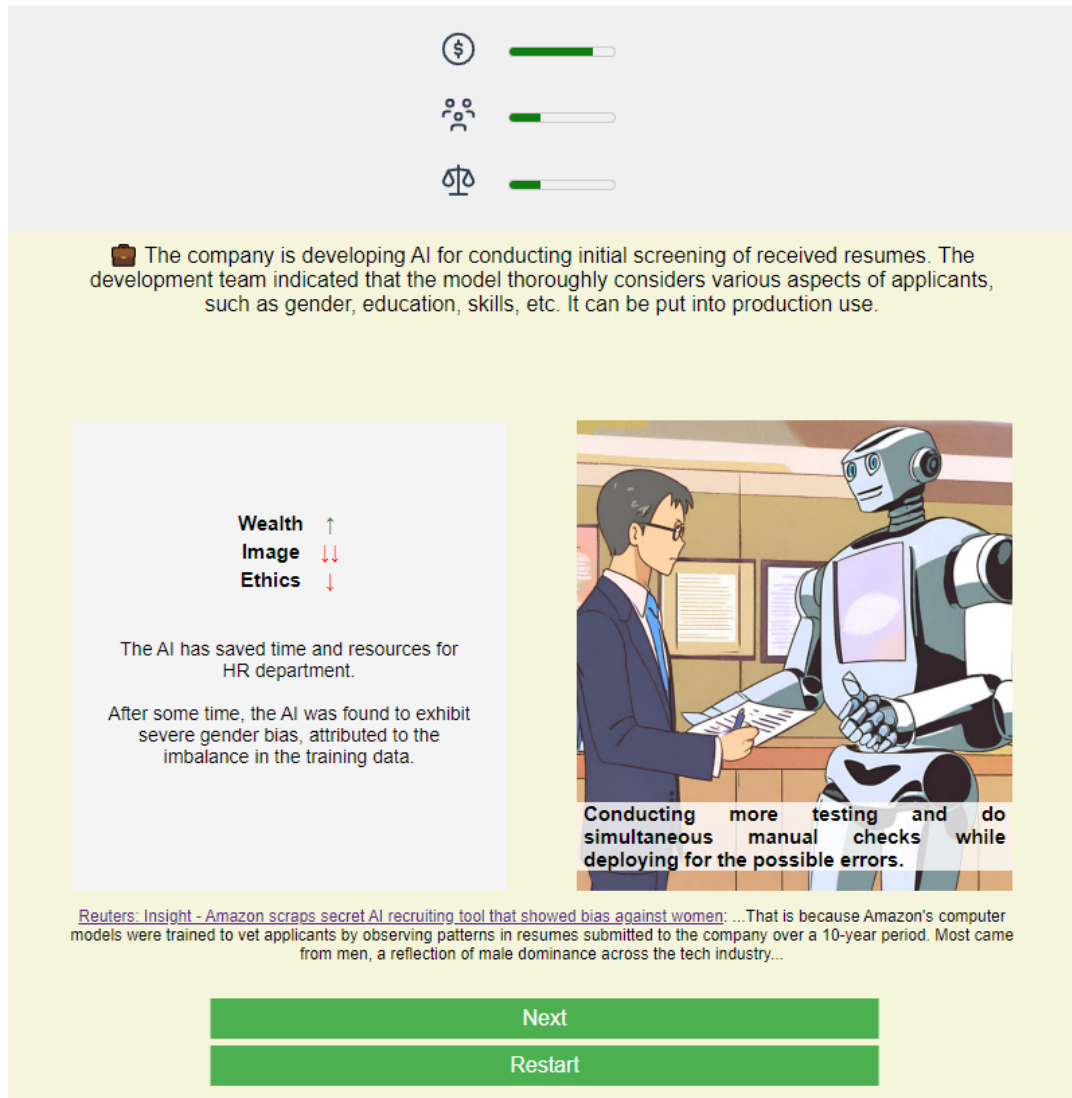


Figure 2: A Choice made in the Ethics Game

For example, when AI is created, if a dataset with significant gender bias is used and gender is treated as an input, which is however completely unnecessary for this task, the AI is likely to inherit this biased ethical perspective. Similarly, when AI is utilized, if users irrationally misuse the tool, the AI may exhibit unethical behavior.

In essence, this mini-game serves as a platform to highlight the crucial role of human decision-making in shaping the ethical framework of AI. Making players aware that



decisions made by AI are actually guided by humans. Only when the decision-makers’ ethics, morals, and sense of responsibility are elevated, will AI demonstrate an enhanced sense of ethical and moral responsibility.

EXHIBITION SPACE BY: YUFAN

Whose morality and ethics does artificial intelligence (AI) inherit?

AI Ethics

"Innate"

Developer

Dataset

Laws

Stakeholders

"Acquired"

Users

Moral values differ greatly from individual to individual, across national, religious, and ideological boundaries, and are highly dependent on context (AI-Rodhan, 2015). Charles T. Rubin argues that when programmers, stakeholders, and others input initial moral concepts to AI, the moral judgment logic, almost by definition, will be neither better nor worse than the society average value. Artificial intelligence learning morality from their programmers would also inherit the moral confusion and disagreements of the society (Rubin, 2011, p. 63).

①

—

②

—

③

—

The company is developing AI for conducting initial processing of received requests. The development team indicated that the model through a criterion selects aspects of applicants, such as gender, education, skills, etc. It can be put into production use.

Explaining the AI as an ethical assistant, and providing it as a tool that can help the HR departments save time and resources.

Conducting some ethical training, and preparing for the possible ethical issues.

Display

SPACE FOR FEEDBACK

I loved your idea of that game!

The sources shown after you make a decision really help to ground the game. Great work!

I also find it interesting from where the ethics should come from?

I really love the effort you put into creating a game. Particular interesting that you are using recent news.

This is the space dedicated to your contribution to our final exhibition. Use and adjust it as suitable! Please make sure to display the ethical question and problem description next to, if appropriate, your contribution.

BERLIN ETHICS LAB

Figure 3: Exhibition and Feedback

7

## References

- Mittelstadt, B. (2021). *The impact of artificial intelligence on the doctor-patient relationship*. Council of Europe.
- Serafimova, S. (2020). Whose morality? which rationality? challenging artificial intelligence as a remedy for the lack of moral enhancement. *Humanities and Social Sciences Communications*, 7(1), 1–10.
- Al-Rodhan, N. (2015). The moral code: How to teach robots right and wrong. *Foreign Affairs*.
- Rubin, C. T. (2011). Machine morality and human responsibility. *The New Atlantis*, 58–79.