https://doi.org/10.4467/27204812ZG.24.011.19727

https://orcid.org/0009-0008-5173-6735
Teacher: Jolanta Rutkowska, M.A.
Students: Jakub Romanowski, Zuzanna Snopek,
Anna Ślimak, Łukasz Walancik
Jagiellonian Language Center
Jagiellonian University, Kraków

AI ETHICS OR ETHICAL AI – STUDENTS' AND TEACHER'S REFLECTIONS ON THE COURSE AND THE TOPIC OF ETHICS IN ARTIFICIAL INTELLIGENCE

AI ethics or ethical AI? – is the title of the theme-based course that I started teaching in October 2023. It was not easy to come up with the subject that will attract students. I have been teaching English to students of computer science for the last 19 years and I still remember how I felt before I had my first class with them: "How can I teach anything even remotely connected with Information Technology (IT) if I have no knowledge of the area whatsoever?" Just knowing the language did not mean I could understand the terminology or the whole articles on IT for that matter. And yet here I am, 19 years later, teaching a course on Artificial Intelligence (AI) ethics ☻.

In the recent years the world of Artificial Intelligence has changed dramatically, making the connection between technology and ethics the centre of attention for both specialists and amateurs. AI systems are advancing in their complexity and autonomy, which makes it a real challenge for people to decide, among other things, e.g. if the work has been done by a human or by AI. Addressing these concerns is vital if we want to build the future where AI enhances human well-being rather than compromises it. In this course, together with a group of wonderful and smart young people from various programmes of Jagiellonian University, we tried to dig deeper into the realm of AI ethics, immersing into ongoing debates and exploring the challenges. We talked about bias, transparency and AI impact on society and tried to understand all ethical implications, hoping that it is possible for technology and humanity to harmoniously coexist.

Below you will find the participants of this course and their contributions to this article. They kindly agreed to present their areas of interest and expertise in relation to AI:

1. Jakub Romanowski – second-year cognitive science student, Master's Programme

I have a background in computer science but decided to pursue cognitive science out of an interest in the ethics of Artificial Intelligence and the more philosophical aspects of cognition. That interest is also the reason I chose to attend the course in AI Ethics.

Cybersecurity and AI

Cybersecurity is a rapidly evolving field with increasingly sophisticated techniques being developed daily by both malicious and benevolent actors. 350 thousand new examples of malware are observed every day. As we progress deeper into the information age, we generate exponentially more data in our daily lives. In the sea of internet traffic, it becomes harder to determine who has ill intentions. Old methods of combating cyber-crime are quickly becoming obsolete. Heuristics utilised by firewalls to distinguish legitimate network traffic from the anomalous rely on rulesets that need to be tediously updated by humans. One thing AI is really good at is combing through large quantities of data and extracting the trends and patterns. A number of classification algorithms can be employed to learn from vast quantities of existing data and detect malicious actions. Of course AI can itself be utilised for nefarious means, generative AI has widely been employed in scams where it can either help reach out to more victims or help trick them by creating convincing video and audio footage of celebrities or loved ones. With the field still in its infancy, it is already clear AI will play a part in the arms race between white hat and black hat hackers.

2. Anna Ślimak – first-year student of astronomy, Master's Programme

I decided to join the course because I firmly believe that AI is the future of science and hence is a great tool to get to know alongside its potential ethical implications.

Astronomy & AI

Astronomy is a broad field of study that in the great part relies on finding patterns, sifting through massive amounts of data, and connecting theory with observations. AI proves to be an invaluable tool for those purposes.

Thanks to neural networks astronomers managed to improve the image of the black hole in the M87 galaxy and remove the optical interference created by Earth's atmosphere from images taken by ground-based telescopes, and much, much more. AI is also helping cosmologists in understanding the nature of dark matter and dark energy. The future is promising for AI in this field as modern observatories will produce thousands of terabytes of data and the use of artificial intelligence will be imperative (https://www.space.com/astronomy-research-ai-future).

AI can shorten and facilitate the process of reducing the data and save astronomers some time to focus on the results and their interpretation. The use of AI unfortunately has some potential problems, regarding the integrity of the research data, possible misinterpretation of the results, and biased data. As long as we remember to use it with careful consideration, it may help us make great discoveries.

3. Łukasz Walancik – first-year student of the Computer Mathematics Master's Programme at the Jagiellonian University

Reading the unreadable

I decided to join the 'AI Ethics or Ethical AI' course, because it seemed like a great opportunity to gain experience in talking about Computer Science in English. As it turned out, I was right. I learned a lot, met brilliant people and had many exciting discussions.

Being part of the 'computer science' social and professional group is a good thing most of the time. Unfortunately, it becomes a nightmare when it comes to writing a presentation or an article on the topic of my field of study. I find it almost impossible to interest my audience in something that is only related to computer science. This is why I have chosen a topic that involves archaeology, physics and artificial intelligence – the mystery of the Herculaneum Scrolls.

In 79 AD, in the ancient Roman city of Herculaneum there was a villa, which contained a library full of papyri with works of contemporary philosophers. (Un)Fortunately, the eruption of Mount Vesuvius buried the villa under a layer of dust several meters thick. This carbonized the papyri and saved them from the undesirable effects of time and atmospheric conditions that have destroyed many other works of art throughout history.

After their recovery in the 18th century, it turned out that reading carbonized pieces of papyrus is extremely hard, as they were brittle chunks of coal. All attempts to read them intact failed.

It was in 2015 that computer scientist Brent Seales managed to decipher the contents of another antique text – the En Gedi parchment – using his virtual unwrapping technology. This method consists of three steps. The first step is to create

a 3D scan of the scroll using a high-resolution X-ray scanner. The next step is to flatten the scan into a 2D representation. The final step is to find traces of ink based on microscopic surface irregularities. And voila! You have unlocked mysteries hidden from human sight for almost two millennia. Of course, this method is much more complicated than that. Because of tremendous amounts of data, it would not be possible without the use of machine learning – a subset of Artificial Intelligence that specializes in analysing large data sets.

Success of this technique gave hope that reading Herculaneum Scrolls could also be possible. In 2023, the 'Vesuvius Challenge' was announced. A competition where participants could use their knowledge and skills to read the contents of publicly available 3D-scans. It led to some breakthroughs, but ultimately it is only the beginning, as there are potentially tens of thousands of scrolls still waiting to be unearthed in the remains of Herculaneum.

4. Zuzanna Snopek – first-year student of cognitive science, Master's Programme

I joined this course for two reasons. The first one is that it aligns quite well with what I'm studying at the moment, and the second, and more important reason, is that it sparked my interest.

Cognitive Science & AI

Cognitive science is an interdisciplinary field that integrates resources from various disciplines including psychology, philosophy, neurobiology, anthropology, linguistics, and artificial intelligence. Its main purpose is to better understand human brain in terms of its functional processes. Since artificial intelligence is just one sub-discipline within cognitive science, identifying the precise benefits and risks associated with each discipline can be challenging. Therefore, in my section of this article, I will focus on the well-known topic of autonomous robots or conscious machines.

I believe that most of us have, at least once in our lives, contemplated the idea of an autonomous and conscious robot and its impact on the future of humanity. However, not many of us have truly considered whether such a concept is even feasible. This issue, precisely, is one of the interests of cognitive science researchers. In their studies, they primarily focus on questions such as: What is consciousness? How is consciousness formed? What components constitute the relationship between the body and the mind? What constitutes free will? How do we analyse the environment around us? And, more importantly, if we do find answers to these questions, how can we implement them into an artificial system? These aspects, and many more, are crucial in creating an autonomous, artificial system. Despite the tremendous progress scientists have made in each discipline of cognitive science, there constantly seem to be more questions than answers.

Looking at the current state of science and knowledge, creating a new intelligent being capable of thinking critically and autonomously, and additionally taking its own actions based on acquired knowledge and analysis of its environment, does not yet seem fully possible. Scientists working on human-like artificial intelligence still have a lot of work to do to achieve their goal. Therefore, the widespread fear of autonomous robots lacks justification, as the perceived threats primarily stem from fictional portrayals in sci-fi movies rather than real-world concerns (https://www.sztucznainteligencja.org.pl/definicja/sztuczna-inteligencja/; https://www.kompetencjecyfrowe.com/?p=446; Tadeusiewicz, 2019; Skalifist, Mikelsten, Teigens, 2020).

To sum up, this article analyses the ethical implications of Artificial Intelligence in the areas of astronomy, information technology, cybersecurity, cognitive science and even history, as presented from students' points of view. All perspectives are rather enthusiastic. However, it should be clear that being aware of the possible dangers arising from AI technologies will protect the humanity from the risk of having the technology getting out of control. Let the AI serve us, not conquer us.

Last, but not least, I, Jolanta Rutkowska, would like to say that I am really grateful to my students for contributing to this article and I feel honoured that I could explore this fascinating topic with such a wonderful group of smart young people.

Bibliography

https://www.kompetencjecyfrowe.com/?p=446 [accessed: 28 February 2024].

https://www.space.com/astronomy-research-ai-future [accessed: 28 February 2024].

https://www.sztucznainteligencja.org.pl/definicja/sztuczna-inteligencja/ [accessed: 28 February 2024].

Skalifist, P., Mikelsten, D., Teigens, V. (2020). *Sztuczna inteligencja: czwarta rewolucja przemysłowa*. Cambridge Stanford Books.

Tadeusiewicz, R. (2019). Automatyzacja i sztuczna inteligencja jako źródła prawdziwych i wyimaginowanych zagrożeń. W: B. Galwas, P. Kozłowski, K. Prandecki (red.). *Czy świat należy urządzać inaczej. Schyłek i początek*, Warszawa, 29–43.