

Principles for Working in Artificial Intelligence

Our interconnected world allows people of many different opinions to work on the same projects. While in these diverse environments, it is vital to understand how your value system interacts with the technology you are developing. Throughout this piece, I will delve into seven principles that are helpful in guiding professionals in the artificial intelligence field. I will talk about the principles in groups of two, with the exception of the first principle.

The most important principle for the majority of people, especially Christians, is centered around the idea that humanity and artificial intelligence have different purposes in this world. Drawing from scripture, this can be easily seen in how God talks about created humans compared to the rest of creation. In Genesis 1:26-28 God states that he made man in his image and gave them the responsibility to rule over creation. He didn't give any other part of the creation this role. It was given to humanity, and we must take this responsibility seriously. We should use artificial intelligence as a tool to better people's lives, but it will never replace people on earth or in the eyes of God.

The next two principles I would like to discuss are transparency and safety. They are closely related and work well together. Transparency means that both developers of AI are honest about their AI, and we understand why AI is making certain decisions. Microsoft calls that later part of this idea interpretability or the ability to explain the reasoning behind actions. This principle opens up the door for many of the following principles. The idea behind safety is simple. We don't want AI to be harming people physically, mentally, or in any other way. According to Jessica Fjeld and her team, this "generally refers to the proper internal functioning

of an AI system.” This requires careful planning and checks on the part of the people making the AI.

Another set of two principles is accountability and security. Accountability means that people in charge of designing and implementing artificial intelligence take responsibility for their work. In Luke 17:3, we see the value in holding each other accountable for our actions and repenting if we have done something wrong. We should recognize the weight of the responsibility we undertake while we are building an AI. One of the rights we are responsible for in an AI is maintaining the security of people’s private property. The World Economic Forum states, “Individuals should have the right to manage their data that is used to train and run AI solutions.” We should respect people’s wishes with their private information and do our best to protect it from getting stolen.

The final two principles are equality and the promotion of human rights and values. Equality comes into play when we attempt to minimize a system's bias and to treat everyone fairly. We are all the same in the eyes of God, according to Galatians 3:28, and we should strive to align our AI systems with the goal of treating everyone equally. The promotion of human rights and values consists of three areas. Jessica Fjeld and her team describe these as human values and flourishing, access to technology, and leveraged to benefit society. Overall, AI should be used to uphold and protect key rights given to us by God.

Our values system should be a guiding force in the work we do or don’t do in this world, and it is important to think about how that values system applies to artificial intelligence. The seven principles I discussed include recognizing the differences between artificial intelligence and humanity, transparency, safety, accountability, security, equality, and promoting human rights

and values. If we follow principles similar to the ones listed above, I think we are less likely to cause harm to ourselves and others in the future.

Bibliography

Bennett, Matthew. “Artificial Intelligence: An Evangelical Statement of Principles.” *ERLC*, 7 Nov. 2022, erlc.com/resource-library/statements/artificial-intelligence-an-evangelical-statement-of-principles/.

Fjeld, Jessica, et al. “Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI.” *SSRN Electronic Journal*, 2020, <https://doi.org/10.2139/ssrn.3518482>.

Our Approach to Responsible AI at Microsoft, www.microsoft.com/en-us/ai/our-approach?activetab=pivot1%3Aprimaryr5. Accessed 20 May 2023.

“9 Ethical AI Principles for Organizations to Follow.” *World Economic Forum*, www.weforum.org/agenda/2021/06/ethical-principles-for-ai/. Accessed 20 May 2023.