

The ethics and politics of artificial intelligence

*Philosophers often argue that knowledge is neutral, but the use we make of it can bring good or bad consequences. Seeing technology as knowledge, **Thomas Ferretti** considers ethical and political issues raised by the ongoing revolution in artificial intelligence (AI) and machine learning (ML).*

Technology is knowledge

We often think that acquiring more knowledge is always good. If I acquire the knowledge of calculus, I can now do some things that I could not do before.

Yet, what philosophers call a “consequentialist” perspective might propose instead that knowledge itself is neutral: we must know how I am going to use my knowledge in practice before judging whether it has good or bad consequences. If I use calculus to launch a satellite to monitor climate change, the effects are good. But the same knowledge can be used to calculate missile trajectories in an unjust war.

Because technology like artificial intelligence (AI) and machine learning (ML) can be understood as the knowledge of specific techniques, skills, and know-how, this perspective has led many to conclude that the technology itself is neutral: only the way we decide to use it in society determines whether it has good or bad effects.

On the positive side, AI and ML have already been [used in science](#) to discover new exoplanets by analysing data from telescopes, Google uses these techniques to reduce energy consumption and CO2 emissions in its data centres, and image recognition is used in healthcare to improve diabetic retinopathy detection. But AI could also be used in [autonomous weapon systems](#) that could select and attack a human target without being instructed to.

Therefore, whether or not you subscribe to the view that technology is neutral, and setting aside technical problems in [value-alignment](#), the question remains: How can we make sure that technologies like AI and ML are used as a force for good and how to avoid harmful effects?

Knowledge is power

In many people’s minds, AI raises the threat of work automation. Like other technologies, AI will lead to creative destruction and disrupt existing modes of production to replace them with more productive ones. There is, however, some [unwarranted alarmism](#) in this domain.

Despite more than a century of unprecedented innovation, we are nowhere near a jobless society. This is because automation has spillover effects on the whole economy: falling prices in sectors where work is automated allow customers to spend more in other sectors, and increasing demand for new goods and services keeps us working.

Yet, creative destruction always involves winners and losers. How we choose to manage these economic transformations will make a difference in the [distributive impact of AI](#). To mitigate potentially unfair inequalities, governments should implement policies such as adequate unemployment benefits, retraining programs, and public investments to improve labour markets.

Besides automation, the increasing adoption of AI gives rise to a range of ethical issues calling for careful consideration. For example: AI coupled with [facial recognition](#) can enable widespread surveillance; AI can also threaten privacy by facilitating the analysis and exploitation of large-scale, sensitive datasets such as [medical records](#); ML can lead to algorithmic bias and discrimination by reproducing [historic injustice](#); and algorithmic decision-making is often opaque, sometimes even to engineers who cannot explain how their AI system is able to produce relevant results, which is a problem for [transparency and accountability](#) in decision-making.

More generally, when investigating the potentially harmful effects of AI, it is important to remember that knowledge is power. If knowledge enables me to do things that I could not do before, this means that it increases my power. Therefore, if some people have access to new technology and others do not, and if people who are already in powerful positions can use it to further increase their power, this can exacerbate already existing [power imbalances in society](#).

Power requires legitimacy

We need principles of AI ethics and institutional safeguards to make sure that these principles are implemented in practice. This is important to protect us all from harmful effects and potential abuses of power by the ones wielding this new technology.

Yet, in pluralistic societies, we must expect moral disagreements about the proper use of AI. Reaching a compromise between citizens requires legitimate consensus-building strategies:

AI ethics principles and institutional safeguards should be selected through legitimate decision procedures that everyone could agree on. Examples include consultations led by [UNESCO](#), the [Montreal Declaration](#), and government initiatives (e.g. [UK](#), [Canada](#), [Europe](#)) bringing together all stakeholders to deliberate on the responsible use of AI.

The use of AI in public administration and business should also be subjected to public scrutiny to make sure that agreed principles are adequately interpreted and implemented in practice and to hold decision-makers accountable when they are not.

Navigating the opportunities and complex ethical challenges of the fourth industrial revolution thus requires the collaboration of governments, businesses, and all of us.



Notes:

- This blog post is based on the **Ethics of AI**, the online course designed by Thomas Ferretti together with Kate Vredenburg, both from LSE's Department of Philosophy, Logic and Scientific Method. If you're interested, you can learn more about the three-week masterclass [here](#).
- The post expresses the views of its author(s), not the position of the Bank of Italy, LSE Business Review or the London School of Economics.
- Featured [image](#) by [geralt](#), under a [Pixabay](#) licence
- When you leave a comment, you're agreeing to our [Comment Policy](#)