



Awareness and Skills Against Automation Bias

by **Andrea Volpe**

When I first had the opportunity to use a satellite navigator while driving my car in the early 2000s, it seemed like magic—even though it was a basic version that merely pointed in the right direction with an arrow and spoke in a metallic voice. At the time, the challenge was to outsmart the technology, question it, and find an alternative route that was faster. Years later, with navigators that are undoubtedly more powerful, precise, and much more affordable, I must admit that I now obediently follow the technology, simply setting the destination and then meticulously following its instructions.

I still get a certain satisfaction from complaining about the device when I end up on a deserted road or stuck in traffic, blaming it and convincing myself that, with my knowledge of the area and my sense of direction, I might have done better... a small consolation!

*The truth is that over time, I have placed total and excessive trust in the machine, to the point of switching off my own critical thinking when choosing which road to take. This is what is known as *Automation Bias*, a cognitive mechanism that is especially relevant today. It describes how people tend to place too much trust in decisions, suggestions, or recommendations from automated systems, even when these may be erroneous or incomplete.*

When it comes to picking the best route, the risk is relatively low. However, automation bias can lead to critical errors in fields such as healthcare (incorrect diagnoses based on automated systems), justice (inaccurate assessments and judgments), and even autonomous driving or trading. The same risks are present in business (data



interpretation, strategic or managerial decisions, hiring, performance evaluation, etc.), where human error, amplified by an excessive reliance on automated systems, can have serious and irreversible consequences. It is essential in these cases to maintain a balance between automation and human intervention, ensuring that technology serves as a support tool rather than a substitute for critical thinking and human responsibility.

At the root of this bias is the excessive belief that an automated system is more accurate and reliable than a human, as well as an evident underestimation of human capability and, consequently, a reduction or suspension of control: one relies on automation, reducing their level of supervision and critical thinking, accepting results without thoroughly verifying the correctness of the information.

*As a society, we tend to look favorably on technology (often through forms of anthropomorphism that make it seem like a friendly entity) and to delegate our most complex and stressful decisions to it. This can fuel automation bias, even in areas like consumer choices and, more delicately, political opinion formation. There is a serious risk of manipulation—not by the technology itself, but by those who design, build, and train it. As Federico Cabitza points out, it would be wrong to view the so-called “algo-crazy” as a dictatorship **by** algorithms; rather, it is a form of power exercised **through** algorithms. Thus, it is a dictatorship **with** algorithms, not **of** algorithms. Ethics and, therefore, responsibility must remain with the people who commission and create algorithms, not with the algorithms themselves, which are deterministic systems that execute what they are designed for (the principle of **ethics by design**).*

“AI is inspired by people, created by people, and, most importantly, affects people. Therefore, it is a tool that requires responsibility” Fei-Fei Li

*We must therefore guard against the tendency to project identity and subjectivity onto machines and to recognize agency (the capacity to act with intention and motivation) in computers, applications, and software as social actors. This is a human tendency that dates back to other eras, such as when our ancestors saw fire as a deity or an animated being. Philosopher Luciano Floridi fears that humans may become, gradually and unknowingly, part of a mechanism he calls the “infosphere”, where not only is there a separation between **agere** (acting) and **intelligere** (understanding), but also an inversion of the relationship, with intelligence ending up in the service of action rather than guiding it. Umberto Galimberti, using the term **téchne** (τέχνη), expressed the idea of technology as a dominant force guiding civilization, no longer subordinated to ethics or philosophy, marking the supremacy of the “how to do” over the “why to do.”*



In summary, both computer scientists and philosophers agree on the importance of maintaining a critical spirit, doubt, and supervision to fully utilize and harness the potential of technology, particularly Artificial Intelligence, acknowledging its numerous benefits when applied to reproductive objectives, but equally recognizing its inadequacy for generative goals, which must remain within the domain of human beings.

Just as an unused muscle atrophies, alongside technical skills and knowledge of how AI works and its possibilities, we must continue to strengthen the distinctive and unique capabilities of human beings. We must build a virtuous relationship with machines that fosters productive, ethical, and value-driven interaction.

An example of an agenda, non-exhaustive but orientative, for the development of human capacities for the effective use of AI includes: effective communication with AI and through AI, problem-setting and problem-solving (from defining objectives and context to critically evaluating results, and eliminating errors, biases, or misleading information), creativity, intellectual curiosity, openness to innovation, decision-making skills, and assumption of responsibility, as well as a strong ethical dimension, empathy, and emotional intelligence.

To start or continue a critical and responsible path of understanding AI's potential, I recommend reading the book "Artificial Intelligence" by L. Floridi and F. Cabitza, drawing on the invitation of Cardinal Carlo Maria Martini, the authors explore the philosophical, ethical, and technical implications of artificial intelligence from different professional perspectives and areas of expertise.

#automationbias #ethics #generativeAi #ai #critique #algocracy #responsibility #agency