

The Future of AI: Stanislaw Lem's Visions for Cyber-societies and Humanity

Roman Krzanowski*, Pawel Polak

Despite its rational and mathematical foundations, AI represents a big unknown for the future of humanity. AI researchers pay only marginal attention to the potential impact AI technology on society. The technologists never realize what they are creating, much like in the tale of the sorcerer's apprentice. They never realize that behind the design there is a sphere of unpredictability; the more complex the technology is, the more opaque the sphere becomes. We are still at the beginnings of AI revolution, and we simply do not know what may be coming. The case of AI technology and its impact of humanity is not a known known or even a known unknown but rather an unknown unknown. So, how do we lift this veil and get a glimpse of a future landscape with an AI-permeated society? One way to anticipate what may be coming is to look at the images of the future found in the visionary writings of some sci-fi writers. In this study, we explore Stanislaw Lem's visions of a robotic and AI society. We explore Lem's prose to see what challenges our technological societies may face when we entrust our lives to AI technology; what questions we should ask and what questions we forget to ask when developing AI systems and allow these systems to control our lives; and what problems we may face, many of which we may not think about yet, being blinded by the illusory magic of technology and our typical shortsightedness.

Lem's stories, which were written in the mid-20th century, may have seemed implausible and fantastic at the time. While Lem's language may seem bizarre now, when we match our newly coined terms to Lem's ideas, Lem's stories represent penetrating philosophical explorations of General AI, Universal AI, supra-intelligence, genetic engineering, biological enhancement, artificial consciousness, genetic eugenics, post-humanism, trans-humanism, autonomous systems well beyond level 5. Lem depicts societies of the future in which software and hardware constructs control every aspect of private and public life. These societies are entirely computerized and automated, and smart machines are everywhere making decisions for people. Lem explores how these societies function, what their values are, and what drives them. Reading Lem is not easy, and it is much easier to imagine oneself in the world of *Blade Runner* than Lem's *Cyberiad* worlds.

However, for all their fantastic aspects, as Lem himself indicates, the stories should not stop the reader from perceiving their deeper meaning. As Lem is unbounded by local technological and rational considerations, at least in terms of his time and place, his landscape of possible variations of societal forces and mechanisms reaches beyond what the visionaries and prophets of technology can conceive or imagine. While Lem's fantastic worlds are societies of AI automata, at the bottom they are deeply human with people's suffering, desires, pleasures, values, sins, problems, and challenges. That is why we may regard Lem's robotic societies as potential human societies in a digital world of the future. Lem points out that technology is always unpredictable and fallible, and it always has consequences that we cannot foresee. There are also always errors in such systems. Technology, as our creation, is never perfect, so it is bound to fail sooner or later in some unpredictable way at an unknown place and time. Technology has a hidden dark side, and it always enslaves us in some way. We design technology to help us, but technology itself changes us, and with each technological enhancement, we lose a part of ourselves. But Lem asks: What do we gain? What would happen if we lose agency, and when would we know that we had lost it? To avoid completely irreversible disasters, whether foreseen or unforeseen, Lem says we need a red button and someone to control it. If we do not have one, we may be heading for a disaster on a cosmic or at least planetary scale. The technologically trained minds of our technocrats are too limited, too ossified, or simply unwilling to guide us into the future. The future they foresee is more of the same in a different shade. To really explore what the future may be, and what it may bring to us, we need the mind unhinged, like the one of Lem. Lem's visions of future societies oriented around

AI and robotics explore AI technology in a deeper and more nuanced way than the current technological visions that our leading technological prophets offer. What is coming, based on Lem's visions, may not be an AI-driven nirvana.

* the corresponding author; rmkrzan@gmail.com

Dr. Roman Krzanowski teaches the philosophy of information and computing at The Pontifical University of John Paul II. His interests in philosophy include the philosophy of information and informatics, ontology and the metaphysics of nature, and ethics and ethical problems created in an information society.

Prof. Paweł Polak is a professor of philosophy at The Pontifical University of John Paul II in Krakow. His interests in philosophy include the history and philosophy of computing/informatics, the history of Polish philosophy, the ethics of autonomous robots, and philosophy in silico.