Editorial

In July 2023, the leaders of seven American companies currently driving innovation in artificial intelligence (AI) announced that they accept an obligation to ensure their technology is safe before releasing it to the public. The backdrop to this agreement is the astonishing progress in the abilities of AI to perform complex tasks. No longer confined to performing specialised tasks determined by human programmers, AI is increasingly able to carry out more general and non-predetermined functions. Drawing on the categorisation of risks introduced in IGJR 1/2022, it is clear that AI-related risks are anthropogenic in origin. And they are largely unknown – which brings them at the centre of IGJR 2/2022. 'Known risks' are defined here as those whose consequences are already manifest, or for which we have a detailed understanding as to their potential causes and consequences. The notion of 'unknown risks', on the other hand, may refer to risks that we have reason to believe are an actual possibility already but we are as unable to fully grasp them. AI falls into this category. To illustrate such 'unknown what-risks', think of a ship's crew that steer their vessel into the Bermuda triangle. They have reason to believe that this is risky on the grounds that other ships have vanished there, but no one really knows why.

There is a second category of unknown risks: the 'unknown where-risks' or 'unknown when-risks'. Take climate change as an example. At the moment, the CO_2 concentration in the atmosphere is already higher than 350 ppm (the level deemed to be safe) and continues to increase at accelerating speeds. In the 1970s, the annual average increase was 0.7 ppm/year. In the 1990s, the rate of increase was 2.2 ppm/year. Currently the rate is at 2.6 ppm/year. This finding is deeply discouraging. A good 30 years after the publication of the first IPCC report and 27 climate conferences later, humanity has failed to reverse or even slow down this dangerous trend. In climate science, about 15 climate 'tipping points' have been identified. A tipping point is a critical threshold that, when crossed, leads to irreversible changes within the climate system and severe impacts on human society. For instance, if the melt of Greenland's and West Antarctica's iceshield surpassed a certain point, the meltdown could not be stopped even if global temperatures were to revert to their pre-industrial level again. As of yet, it is unknown when this (or other) tipping points might be reached. We might compare the climate crisis to a ship driving towards a cliff in a dense fog: We have an exact idea of what awaits us after the crash. But the exact location of the cliff in the fog remains – quite literally – unclear.

For most of human history, people primarily feared natural, well-known risks. In the Anthropocene, this focus has now shifted. As a species, we must come to terms with our unprecedented power and learn a new prudence if we wish to avoid civilisational collapse. Though our cognitive and technological abilities have brought us many benefits, they may also cause our downfall; indeed, there are no peaks without abysses.

In the novel *Gulliver's Travels*, published by Jonathan Swift in 1726, the protagonist (whom the Lilliputians call the Man-Mountain) has to come to grips with a new environment which only seemingly resembles his own. He becomes aware very quickly that there is much he doesn't know and that due to his height, every misstep (literally speaking) can have disastrous consequences for his environment. This is certainly a good metaphor for our unintentional disturbance of the earth's and our own societal boundaries.

The distinction between known and unknown risks forms the conceptual framework for the first article of this edition. Augustine Akah takes AI as his main example for elaborating on its practical implications. He details some possible ways in which AI might cause a civilisational collapse, demanding that more public funding be put into planning for and raising awareness about unknown risks associated with scientific innovations.

In the second article, Christoph Herrler shifts the focus to our moral responsibilities towards future generations, suggesting that we use the language of human rights as a framework for discussing existential risks for them. Herrler takes climate change as his prime example, arguing that we have a moral obligation to ensure that future generations be able to exercise their human rights to the fullest extent possible. These rights include having adequate access to basic goods such as food, water, and safe living environments as a minimum standard of living to which all people – now and in the future – are entitled.

The third article, by Dominik Koesling and Claudia Bozzaro, deals with an often neglected issue within risk research: antibiotic resistance. Though such a problem is unlikely to cause human extinction, it could lead to the deaths of millions of people, which the authors see as an intergenerationally unjust (re)imposition of vulnerability onto future generations and healthcare systems. They examine the danger posed by a post-antibiotic era in which the efficacy of antibiotics is either completely or drastically reduced, a process that unfortunately is already underway.

There follows the book review section. First, Grace Clover compares Roman Krznaric's *The Good Ancestor: How to think long term in a short-term world* (2020) with Richard Fisher's *The Long View: Why we need to transform how the world sees time* (2023) in a double review, considering proposals for long-term mindsets and structural changes.

Kritika Maheshwari then reviews Thomas Moynihan's *X-Risk: How humanity discovered its own extinction* (2020), a study which frames the history of humanity's preoccupation with its own extinction within the context of Kantian philosophy.

Jörg Tremmel, Editor
Grace Clover, Co-Editor
Markus Rutsche, Co-Editor