

Why Existential Risks Matter in International Relations

Written by Brian Wong

This PDF is auto-generated for reference only. As such, it may contain some conversion errors and/or missing information. For all formal use please refer to the official version on the website, as linked below.

Why Existential Risks Matter in International Relations

<https://www.e-ir.info/2022/10/17/why-existential-risks-matter-in-international-relations/>

BRIAN WONG, OCT 17 2022

An existential risk can be defined as a “risk that threatens the destruction of humanity’s long-term potential”. To put it bluntly, it is a risk that can credibly lead to human extinction, with irreversible damage to the ability of human civilisation to repair itself. The ‘terminal impacts’ to existential risks – e.g. their challenges to our existence – need not manifest in the short-term; and this is why they are oft-neglected. Existing communities of researchers focusing on existential risks (x-risks), remain divided over the exact boundaries and constituents of the set of x-risks – though most agree that the following count as ‘core’ examples: a global nuclear winter (arising from the deployment of nuclear weapons or other sources of fallout), a (engineered) pandemic that infects the entire Earth’s population, or Artificial Intelligence (AI) that destroys humanity. Much as these risks are often dubbed to be excessively speculative or exaggerated in nature, they deserve our concern not necessarily because of the probabilities with which they occur, but the absolute scale and intensity of devastation that they would wreak upon humanity.

Many could well spell the end to humanity. Despite so, most discussions on x-risks tend to remain within the domains of moral and applied philosophy – more notably, the Effective Altruism and Long-termism Movements have been instrumental in spearheading the popularisation of the concept. Yet it remains the case that not enough attention is paid to the subject in the international relations (IR) community, with a notable exception being the joint research project by Jordan Schneider and Pradyumna Prasad, which pointed to the risks arising from potential war between the US and China, two sizeable nuclear powers with precipitously tense relations. Indeed, long-termism/x-risk and international relations communities have remained, by far and large, fundamentally disjointed. The following seeks to sketch out a few conceptually rooted arguments concerning why the field of IR and IR scholars must take seriously the possibility of existential risks, to grapple fully with the stakes and challenges confronting us today.

Picture this: a series of explosions storm the world in quick succession, incinerating vast swathes of the Earth’s population, and killing many more through the smoke emissions and environmental damages that immediately follow. The radioactive traces of the detonations and bombings permeate the thickest walls of overground buildings, affecting the billions left behind. The gargantuan volume of particles emitted by the detonations fill the skies with fog and smoke so dense that it would take years, if not decades, before the skies clear. Darkness prevails.

The above picture is one of a global nuclear winter. As Coupe et al. note in a 2019 paper, a potential nuclear winter following on from a hot war between the US and Russia could give rise to a “10°C reduction in global mean surface temperatures and extreme changes in precipitation”. At first glance, there exist sufficient fail-safe mechanisms to render this worst-case scenario improbable: military commanders that are cognizant of the risks of escalation; the existence of bunkers in which individuals can seek refuge; the fact of the mutually assured destruction imposing sufficient deterrence upon key decision-makers.

Yet, this possibility cannot be so simply dismissed. It has been over two hundred days since the Russian army invaded Ukraine. Recent setbacks on the battlefield and growing dissatisfaction amongst the Russian population have precipitously heightened the likelihood that Putin would contemplate deploying a tactical nuclear weapon on the battlefield. Without wading into specific quantitative estimates (though examples of these can be found here), the underlying explanations are relatively straightforward: in seeking an increasingly unlikely victory over nominally Russia-claimed territories in Ukraine, preserve his domestic credibility and political standing, and to force the hands of NATO and Ukraine to come to the negotiation table, Putin might feel that he is running out of viable options.

Why Existential Risks Matter in International Relations

Written by Brian Wong

The nuclear option is most certainly undesirable even to Putin given the potential repercussions, but could be seen as preferable to perceived capitulation and the eventuation of overthrow by actual opposition – for which there is currently relatively limited chance of success. Indeed, categorically, the full-blown military conflicts between any two nuclear powers – Russia, China, or the US; Pakistan and India – could escalate, through the security dilemma – into inadvertently precipitating a nuclear confrontation between such powers.

Nuclear winters are by no means the only x-risks. Take the much-touted AI ‘arms race’ for instance – as AI progresses precipitously towards greater speeds, greater accuracy, and cultivates a deeper capacity to adjust and course-correct through self-driven calibration and imitation, it is apparent that it, too, could equip countries with substantially greater capacities to do harm. Whilst the research and development itself would generate relatively innocuous outputs – such as programmes capable of tracking and monitoring individuals’ behaviors and speech patterns, or AIs guiding lethal autonomous weapons in choosing their targets, it is the yearning for competition and victory that poses a fundamental threat to global safety.

We have seen leading powers such as China and the US seek to out-maneuver one another through punitive and preemptive measures pertaining to chips and semiconductors. Correspondingly, the level of coordination and communication across sensitive issues – such as AI and deployment of drones – has declined considerably, reflective of the broader attitudes of mistrust and skepticism that underpin the bilateral relationship. In the dearth of clearly agreed-upon frameworks for regulation and expectation alignment, it would be of no surprise if the AI race between the two largest economies in the world culminated at a vicious race to a particular bottom: a bottom in human welfare as AI is wielded by antagonistic powers to achieve geopolitical objectives, and, in the process of so doing, causes substantial disruptions and irrevocable destruction to our digital and data infrastructure.

Setting aside the prospective dangers of clashing world powers, there exists a further, positive case for genuine international cooperation. Existential risks require coordination in resources, strategies, and broader governance frameworks in order to be properly addressed. The risks arising from a non-aligned, strong artificial intelligence – that is, a self-conscious, improving, and truly autonomous AI whose preferences diverge from those of human (interests), could well culminate at human extinction. Such risks require careful management and installation of both guardrails and responsive programmes that could mitigate against prospective non-alignment, and/or the premature arrival of strong AI. In theory, countries that lead technology and innovation should be allocating substantial resources to devising a shared and transparent framework of AI regulation, as well as foresight-driven research aimed at planning for various scenarios and possible trajectories adopted by AI. In practice, governmental cynicism and strategic importance attached to accelerating domestic-national developments in AI have rendered such long-term-oriented initiatives incredibly difficult. Even European legislation on AI – arguably the most advanced amongst its counterparts – remains vulnerable to internal discrepancies and nonalignment. More interlocution between continents and geopolitical alliances is thus vital in enabling the devising of regulations, laws, and decision-making principles that can tell for what to do in face of AI risks.

An alternative concern looms, concerning the stability of food supply in the face of extreme weather and other geopolitical disruptions. Consider the ongoing global food crisis, which has arisen from a combination of the ongoing war in Ukraine and regional droughts and floods resulting from a prolonged La Niña (attributed by some to climate change). A key to resolution of such crises requires both targeted and comprehensive agreements over production and distribution of food, as well as a fundamental structural push for more rapid green transition. Short of global coordination, much of this would be hugely difficult – food supply chains cannot be optimised if trade barriers and border skirmishes continually disrupt cross-border flows. Attempts to curb carbon emissions and advance a shift away from non-renewables, would require countries to see value in their committing and adhering to stringent yet much needed pledges concerning shrinking their carbon footprints. Short of genuine distribution of labour and collaboration – over the production of solar panels and renewable energy, for one – we would be trending dangerously down a path of no return.

There are those who argue that climate change is not, in fact, an existential risk; that its effects are unevenly distributed throughout the world and could be overcome through adaptive technologies. Yet this underestimates the extent to which disruptions to food production and supply can cause or exacerbate preexisting geopolitical and

Why Existential Risks Matter in International Relations

Written by Brian Wong

cultural tensions, thereby precipitating conflicts that could eventually escalate into total or nuclear war. The probability may be objectively low, but the harms are sufficiently weighty as to merit serious attention.

The academic community would thus benefit from taking seriously the quantum of impacts that international conflict and collaboration possesses in relation to existential challenges to mankind. There remains much to be explored in the intersection of long-termism and IR – quantification and mechanisation of causal processes, the devising and evaluation of prospective solutions. Fundamentally, it is imperative that IR theory can account for not just the probabilistically likely and proximate – but also structural threats that could undermine the continuity and survival of the human species.

About the author:

Brian Wong is a DPhil candidate in Politics at Balliol College, Oxford. Brian's primary research examines the responsibilities borne by citizens residing under authoritarian regimes for their states' injustices. Brian has taught modules in politics to undergraduate students at Oxford and Stanford Universities, has presented on Sino-American relations at a range of universities, has advised leading MNCs on macro risks in Asia, and is the Chief Tutor of the Apollo Fellowships.