## NC

Priortizing environmental protection entails AI development.

**Cortes et al 2k** writes[[1]](#footnote-1)

As pointed out by several authors, the environment is a complex and dynamic system where different aspects can lead to the same impact (e.g., the emission of global warming gases), while other actions might combine synergistically to create an impact which is much greater than that which would be predicted using a reductionist approach (e.g., nitrogen dioxide and hydrocarbons can react to produce tropospheric ozone). Thus, further analysis is required to assess the likely impacts that the significant aspects will have [4]. 78 Cort´es et al. The problem of global change is complex in nature and may be represented by various interactions that operate on different spatio-temporal scales. Addressing these issues demands an integrated consideration of relevant interactions between humans and the environment [5]. Information technologies have played an increasing and central role in the planning, prediction, supervision and control of environmental processes at many different scales and within various time spans. At the same time, organisations, industries (e.g., the ISO14001, the European standard EMAS, etc.) and **governments** (pioneered by the 1969/70 National Environmental Policy Act (NEPA) of the United States) **have started to take on a** more **proactive relationship with the environment by introducing** appropriate **legislation** calling for the explicit consideration of environmental impact in the planning and decision-making process for large projects (e.g., Kyoto Summit, Agenda 21 or, the Rio do Janeiro Declaration) [6]. During the last two decades, the fast developments in information technologies and the **rapid development of** new and faster **hardware made** the establishment of interdisciplinary research **links between environmental and computer scientists** possible and very **fruitful**. A new discipline, known as **Environmental Informatics** [7], which **combines** research fields such as **A**rtificial **I**ntelligence**,** Geographical Information Systems (**GIS**)**, Modelling and Simulation**, User Interfaces, **etc**., is emerging. Animportant and difficult task for this new area is to serve as a catalyst for the integration of data, information, and knowledge from various sources in the environmental sector [8, 9]. In this paper, we will explore the Artificial Intelligence contributions to the field. 1.1. Organisation of this Paper The goal of this paper is to show how **A**rtificial **I**ntelligence in particular and Information Technologies, in general **have succeeded in developing adequate tools for** modelling, **design,** simulation, prediction, **planning and decision-support systems for environmental** management and **protection. Many environmental problems**, such as damage to the biosphere, local air pollution, the spread of harmful substances in the water, and global climatic changes, **cannot be studied by experimentation. Hence,** mathematical models and **computer simulations are being used** as the appropriate means **to get more insight**. In the review, we will not study the field of Environmental Modelling as very little has been developed using AI techniques and tools. A pioneering work in this field is by Robertson et al. [10] who tried to give a Eco-Logic approach to the problem. For more information about this field, thework by P. Zannetti [11] is recommended.

AI development risks an intelligence explosion that spills over to military use – causes extinction, outweighs nuclear war, and comes first under any moral theory.

**Shulman and Armstrong 11** write[[2]](#footnote-2)

II. **An AI arms race may be “winner-take-all”** The threat of an AI arms race does not appear to be primarily about the direct application of AI to warfare. While automated combat systems such as drone aircraft have taken on greatly increased roles in recent years (Singer, 2009; Arkin, 2009), they do not greatly disrupt the balance of power between leading militaries: slightly lagging states can use older weapons, including nuclear weapons, to deter or defend against an edge in drone warfare. Instead, the military impact of an intelligence explosion would seem to lie primarily in the extreme acceleration in the development of new capabilities. **A state might launch an AI Manhattan Project to gain** a few months or **years of sole access** to advanced AI systems**, and then initiate an intelligence explosion** to greatly increase the rate of progress. Even if rivals remain only a few months behind chronologically, they may therefore be left many technological generations behind until their own intelligence explosions. It is much more probable that such a large gap would allow the leading power to safely disarm its nuclear-armed rivals than that any specific technological generation will provide a decisive advantage over the one immediately preceding it. If states do take AI potential seriously, how likely is it that a government's “in-house” systems will reach the the point of an intelligence explosion months or years before competitors? Historically, there were substantial delays between the the first five nuclear powers tested bombs in 1945, 1949. 1952, 1960, and 1964. The Soviet Union's 1949 test benefited from extensive espionage and infiltration of the Manhattan Project, and Britain's 1952 test reflected formal joint participation in the Manhattan Project. If the speedup in progress delivered by an intelligence explosion were large, such gaps would allow the leading power to solidify a monopoly on the technology and military power, at much lower cost in resources and loss of life than would have been required for the United States to maintain its nuclear monopoly of 1945-1949. **To the extent that states distrust their rivals with** such **complete power**, or wish to exploit it themselves, **there would be strong incentives to vigorously push forward AI research**, and to ensure government control over systems capable of producing an intelligence explosion. In this paper we will discuss factors affecting the feasibility of such a localized intelligence explosion, particularly the balance between internal rates of growth and the diffusion of or exchange of technology, and consider historical analogs including the effects of the Industrial Revolution on military power and nuclear weapons. III. Accidental risks and negative externalities A second critical difference between the nuclear and AI cases is in the expected danger of development, as opposed to deployment and use. Manhattan Project scientists did consider the possibility that a nuclear test would unleash a self-sustaining chain reaction in the atmosphere and destroy all human life, conducting informal calculations at the time suggesting that this was extremely improbable. A more formal process conducted after the tests confirmed the earlier analysis (Konopinski, Marvin, & Teller, 1946), although it would not have provided any protection had matters been otherwise. The historical record thus tells us relatively little about the willingness of military and civilian leaders to forsake or delay a decisive military advantage to avert larger risks of global catastrophe. In contrast, **numerous scholars have argued that advanced AI poses a nontrivial risk of** catastrophic outcomes, including **human extinction.** (Bostrom, 2002; Chalmers, 2010; Friedman, 2008; Hall, 2007; Kurzweil, 2005; Moravec, 1999; Posner, 2004; Rees, 2004; Yudkowsky, 2008). Setting aside anthropomorphic presumptions of rebelliousness, a more rigorous argument (Omohundro, 2007) relies on the instrumental value of such behavior for entities with a wide variety of goals that are easier to achieve with more resources and with adequate defense against attack. Many decision algorithms could thus appear benevolent when in weak positions during safety testing, only to cause great harm when in more powerful positions, e.g. after extensive self-improvement. Given abundant time and centralized careful efforts to ensure safety, it seems very probable that these risks could be avoided: development paths that seemed to pose a high risk of catastrophe could be relinquished in favor of safer ones. However, the context of an arms race might not permit such caution. A risk of **accidental AI disaster would threaten all of humanity**, while the benefits of being first to develop AI would be concentrated, creating a collective action problem insofar as tradeoffs between speed and safety existed. A first-pass analysis suggests a number of such tradeoffs. Providing more computing power would allow AIs to either operate at superhumanly fast timescales or to proliferate very numerous copies. Doing so would greatly accelerate progress, but also render it infeasible for humans to engage in detailed supervision of AI activities. To make decisions on such timescales AI systems would require decision algorithms with very general applicability, making it harder to predict and constrain their behavior. Even obviously **risky systems might be embraced for competitive advantage**, and the powers with the most optimistic estimates or cavalier attitudes regarding risk would be more likely to take the lead. IV. Barriers to AI arms control Could an AI arms race be regulated using international agreements similar to those governing nuclear technology? In some ways, there are much stronger reasons for agreement: the stability of **nuclear deterrence, and** the **protection afforded by existing nuclear powers to their allies, mean that** the **increased threat of a new nuclear power is not overwhelming**. No nuclear weapons have been detonated in anger since 1945. **In contrast,** simply **developing AI capable of producing an intelligence explosion puts all states at risk** from the effects of accidental catastrophe, or the military dominance engendered by a localized intelligence explosion. However, AI is a dual-use technology, with incremental advances in the field offering enormous economic and humanitarian gains that far outweigh near-term drawbacks. Restricting these benefits to reduce the risks of a distant, novel, and unpredictable advance would be very politically challenging. Superhumanly intelligent AI promises even greater rewards: advances in technology that could vastly improve human health, wealth, and welfare while addressing other risks such as climate change. Efforts to outright ban or relinquish AI technology would seem to require strong evidence of very high near-term risks. However, agreements might prove highly beneficial if they could avert an arms race and allow for more controlled AI development with more rigorous safety measures, and sharing of the benefits among all powers. Such an agreement would face increased problems of verification and enforcement. Where nuclear weapons require rare radioactive materials, large specialized equipment, and other easily identifiable inputs, AI research can proceed with only skilled researchers and computing hardware. Verification of an agreement would require incredibly intrusive monitoring of scientific personnel and computers throughout the territory of participating states. Further, while violations of nuclear arms control agreements can be punished after the fact, a covert intelligence explosion could allow a treaty violator to withstand later sanctions. These additional challenges might be addressed in light of the increased benefits of agreement, but might also become tractable thanks to early AI systems. If those systems do not themselves cause catastrophe but do provide a decisive advantage to some powers, they might be used to enforce safety regulations thereafter, providing a chance to “go slow” on subsequent steps. V. Game-theoretic model of an AI arms race In the full paper, we present a simple game-theoretic model of a risky AI arms race. In this model, the risk of accidental catastrophe depends on the number of competitors, the magnitude of random noise in development times, the exchange rate between risk and development speed, and the strength of preferences for developing safe AI first. VI. Ethical implications and responses The above analysis highlights two important possible consequences of advanced AI: a disruptive change in international power relations and a risk of inadvertent disaster. From an ethical point of view, the accidental risk deserves special attention since it threatens human extinction, not only killing current people but also denying future generations existence. (Matheny, 2007; Bostrom, 2003). **While AI systems would outlive humanity, AI systems might lack key features contributing to moral value, such as** individual **identities, play, love, and happiness** (Bostrom, 2005; Yudkowsky, 2008). Extinction risk is a distinctive feature of AI risks: **even a catastrophic nuclear war or** engineered **pandemic that killed billions would still likely allow survivors** to eventually rebuild human civilization**, while AIs killing billions would likely not** leave survivors. (Sandberg & Bostrom, 2008). However, a national monopoly on an AI intelligence explosion could also have permanent consequences if it was used to stably establish its position. Permanent totalitarianism is one possibility (Caplan, 2008). We conclude by discussing some possible avenues for reducing these long-term risks.

1. U. Cortes, M. Sanchez-Marre, and L. Ceccaroni (Software Department, Technical University of Catalonia (UPC), Jordi Girona 1-3. E08034 Barcelona, Catalonia, Spain), I. R-Roda and M. Poch (Chemical and Environmental Engineering Laboratory, University of Girona, Campus de Montilivi, E17071 Girona, Catalonia, Spain). “Artificial Intelligence and Environmental Design Support Systems.” Applied Intelligence. 2000. http://www.cin.ufpe.br/~rvf/AI%20and%20Environmenal%20Issues.pdf [↑](#footnote-ref-1)
2. Carl Shulman (Singularity Institute of Artificial Intelligence) and Stuart Armstrong (InhibOx, an organization dedicated to developing and delivering the best services and technologies in computer-aided drug discovery ). “Singularity Hypotheses: A Scientific and Philosophical Assessment.” April 13th, 2011. http://singularityhypothesis.blogspot.com/2011/04/arms-races-and-intelligence-explosions.html [↑](#footnote-ref-2)