# War or Peace Between Humanity and Artificial Intelligence

*Wolfhart Totschnig Universidad Diego Portales (Chile)*

*International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL*

**Keywords:** artificial intelligence, prisoner’s dilemma, war

The thinkers who have reflected on the potential risks of a future artificial general intelligence (AGI) have focused on the possibility that the AGI might carry out its assigned objective in a way that we did not anticipate, with potentially catastrophic effects (Yudkowsky, Bostrom, Omohundro, Yampolskiy, Tegmark, Russell). They have neglected the possibility that the AGI could come to see us as a threat to its existence and, therefore, deliberately try to eliminate us. The aim of the present paper is to show that this neglect is mistaken. I will describe a possible situation where an AGI and humanity find themselves vulnerable vis-à-vis each other, which could lead to an all-out war. I will then argue that, in view of this possibility, the approach of the said thinkers, which is to search for ways to keep an AGI under control, is potentially counterproductive because it might, in the end, bring about the existential catastrophe that it is meant to prevent.