Are we going to be the oppressors of artificial beings?

(a shortened version of this journal paper: https://journals.tdl.org/jvwr/index.php/jvwr/article/view/7369/6456)

Starting from the hypothesis that artificial beings will one day achieve a level of sophistication as to become morally relevant for us, I will discuss whether and under which circumstances it would be ethically viable to include them in virtual environments. The focus will be on virtual environments designed to fulfil specific purposes, such as entertainment, education, training, or persuasion.

To figure out if we are going to be the bad guys, I will begin by introducing a criterion for moral consideration that rely on the notions of 'autonomy' and 'damage'. Adopting this framework, I will tackle the question of whether including morally relevant artificial beings in virtual environments constitutes an immoral action on the part of human creators. The main tools to address this question, will be three conceptual lenses taken from the philosophical branch of ethics:

- the lens commonly used for parenthood and procreation,
- the lens concerning the moral status of animals, and
- the lens of the classical problem of evil.

After having explored the question through those three lenses, I will propose an original, contractualist answer where artificial beings could be presented with epistemic access to the information and characteristics of each virtual environment in which they could take part, as well as details concerning the activities expected to take place within these environments. In this scenario, the artificial intelligences would not be forced into potentially damaging or demeaning situations, but would be allowed to act in virtual environments with levels of knowledge and autonomy comparable to those of human users.