

Prof. Lode Lauwaert


KU Leuven

Can Artificial Intelligence Embody Moral Values?

A solution to the value alignment problem

Date : 02 April 2024 (Tuesday)

Time : 16:30 – 18:00

Venue : WYL111 & Zoom Seminar 

Here is the link to the seminar – <https://lingnan.zoom.us/j/91923328554>

It is often claimed that AI is not neutral, and that this is undesirable. To support this, scholars usually refer to the fact that the output of some smart technologies is biased, or worse, discriminatory. Underlying this is an unrepresentative dataset that the technology has been trained with.

In this presentation I focus on a specific type of AI, namely artificial agents. I defend the thesis that it may be that these agents are not neutral in a desirable way, since they can be laden with moral values such as autonomy and fairness. This results not so much from the data with which the artificial agents are trained, but rather from their ability to rank options and subsequently make choices.

This conclusion has practical relevance, specifically to the question of value alignment. After all, if a technology can be morally laden in a desirable sense, then the likelihood that the system's output will also align with what we find morally important increases.

All are welcome

For enquiries, please call 2616 7445 / 2616 7488 or e-mail to dphil@ln.edu.hk