



Lingnan 嶺南大學
University 香港 Hong Kong

Hong Kong
Catastrophic Risk Centre

Online Research Seminar

What Power-Seeking Theorems Do Not Show

Prof. David THORSTAD

Vanderbilt University

Recent years have seen increasing concern that artificial intelligence may soon pose an existential risk to humanity. One leading ground for concern is that artificial agents may be power-seeking, aiming to acquire power and in the process disempowering humanity. A range of power-seeking theorems seek to give formal articulation to the idea that artificial agents are likely to be power-seeking. I argue that leading theorems face five challenges, then draw lessons from this result.

Venue: Zoom Seminar

Click [HERE](#) or scan the QR code
to join the zoom meeting.



For more information:

<https://www.ln.edu.hk/philoso/hkcrc/>

2616 7445/ hkcrc@ln.edu.hk

**18
FEB
2025**

10:00-11:15 a.m. (HKT)