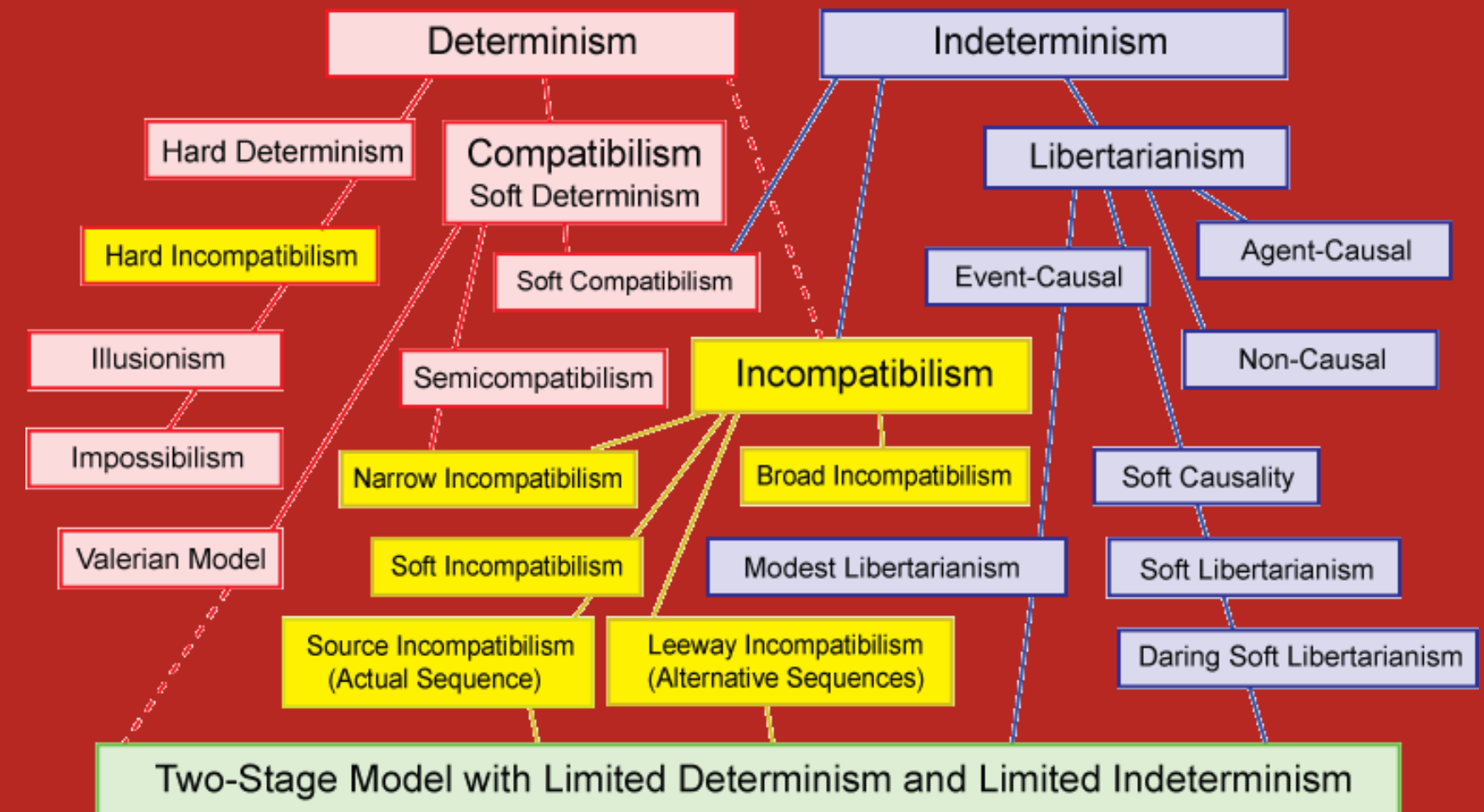# Values and agency

Luca Consoli and Michel Vitale, 02-12-2024

Radboud University

Ethics of AI in (higher) education

https://www.ox.ac.uk/news/2024-11-13-new-ethical-framework-help-navigate-use-ai-academic-research

https://www.unesco.org/en/articles/placing-ethics-center-ie-university-launches-unesco-chair-ai-ethics-and-governance

https://www.upressonline.com/2024/11/ai-revolution-in-higher-education-college-professors-navigate-new-territory/

https://artificialintelligenceact.eu/recital/56/

Radboud University

Outline

- Design of technology

- Regulation of technology

- Agency (Part I, will be continued on December 9)

Radboud University

'Information systems are intentionally or unintentionally informed by moral values of their makers'

'Information technology has become a <u>constitutive technology</u> and partly constitutes the things to which it is applied. It shapes our discourses, practices and institutions in important ways'

https://link.springer.com/chapter/10.1007/978-0-387-72381-5_8

Radboud University

The VSD design process

The value-sensitive design process

Identify
the values and needs most important to the users.

Review
the goals of the project and how the project affects those values.

Design
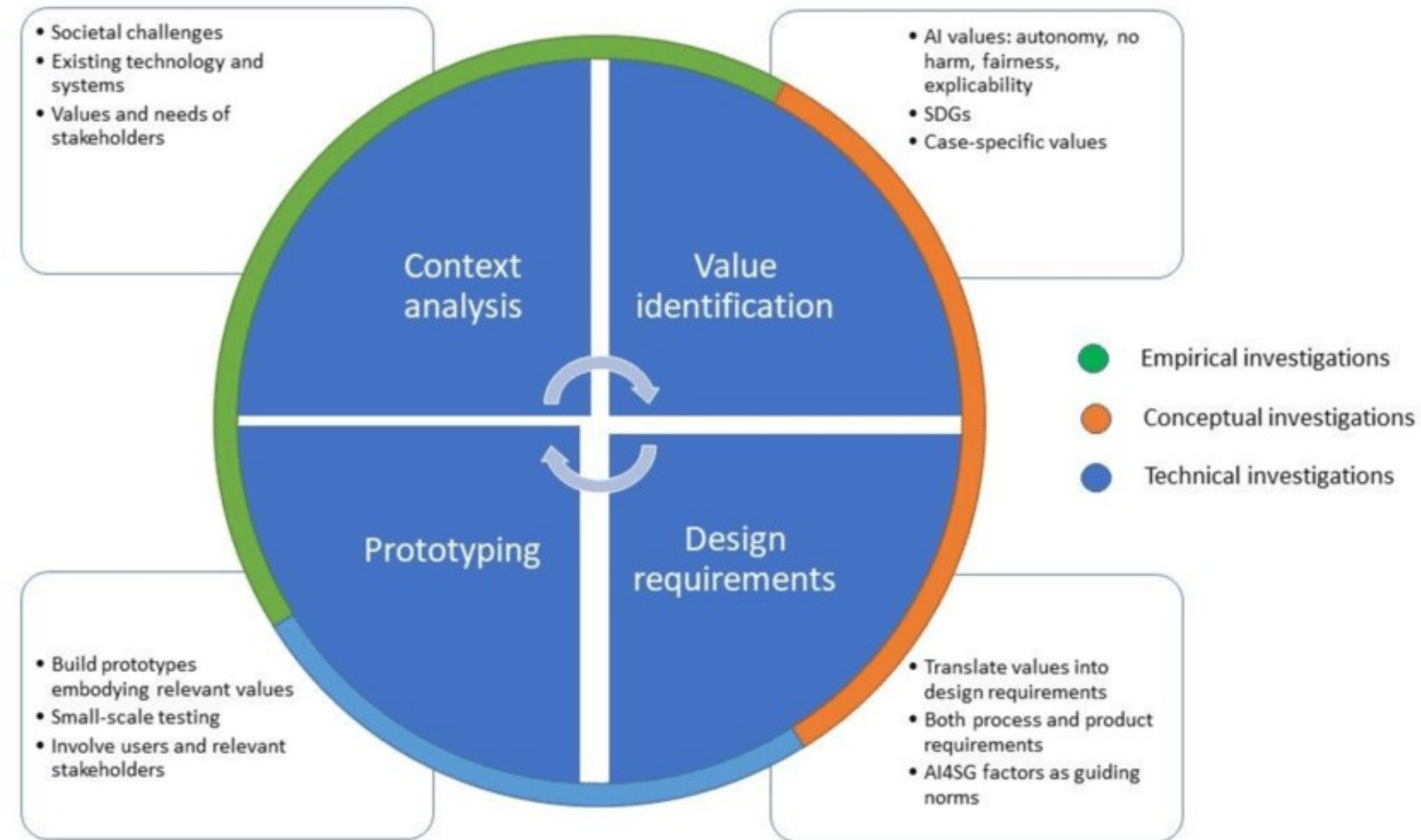the product or service to minimize negative impact on the users' values.

Evaluate
the outcomes of the project to ensure users' needs and values were met.

ICONS: GENESTRO/GETTY IMAGES

©2024 TECHTARGET. ALL RIGHTS RESERVED   TechTarget

https://www.techtarget.com/searchcio/definition/value-sensitive-design-VSD

PECIS, 02-12-2024

The VSD design process applied to AI

https://www.researchgate.net/figure/SD-design-process-for-AI-technologies_fig6_343078979
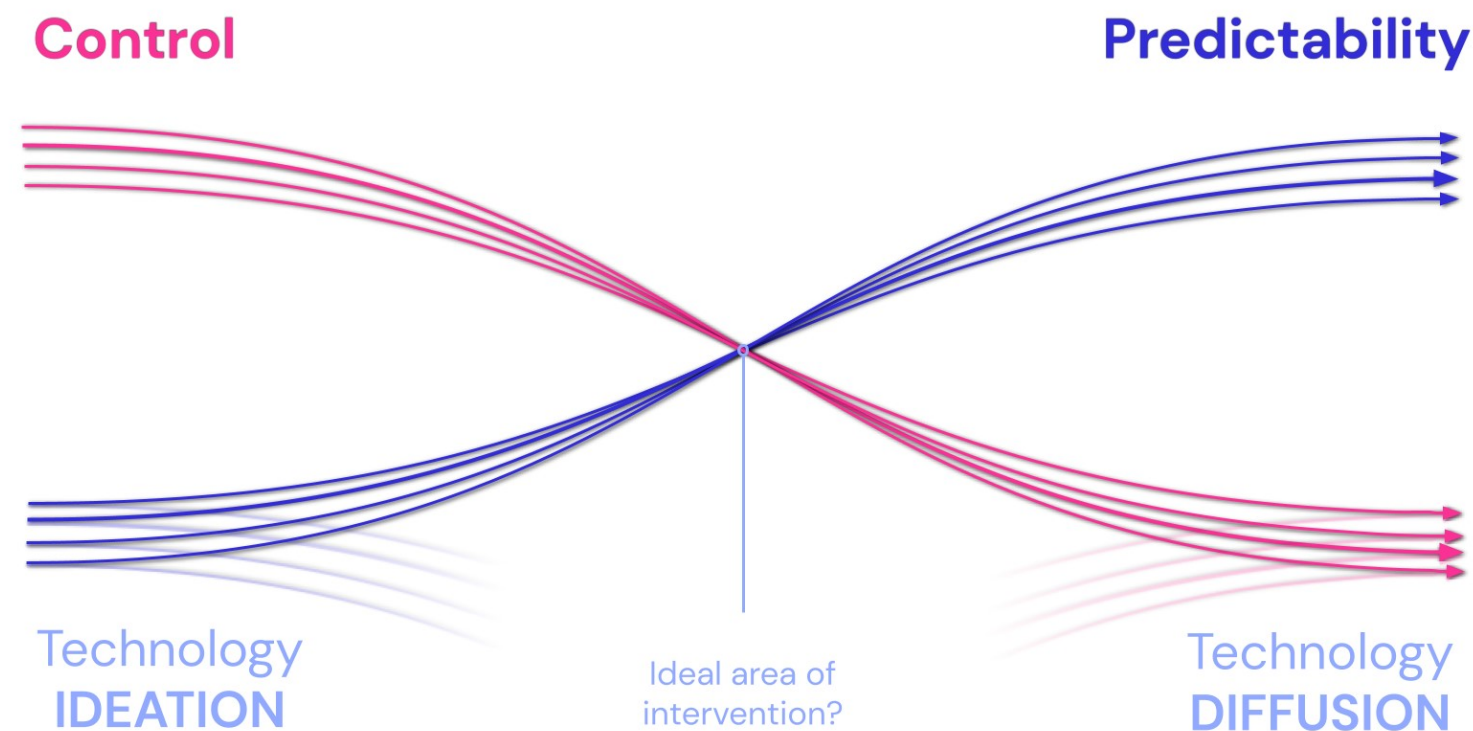
How to do VSD in practice?

Can we design systems that are truly neutral, or do you believe that every design inherently reflects the values of its creators? How might these values influence the outcomes of systems in different domains?

In what ways do you think AI companies could incorporate the VSD into their work?

Radboud University

How can autonomous systems (like self-driving cars) be designed to incorporate crucial societal values? For instance to prioritize human safety, equity, and justice?

Radboud University

The Collingridge dilemma

Control   Predictability

Technology
IDEATION

Ideal area of
intervention?

Technology
DIFFUSION

Adapted from Besti, F. & Samorè, F. (2018). Responsibility driven design for the future self-driving society. Fondazione Giannino Bassetti

https://demoshelsinki.fi/2022/02/15/what-is-the-collingridge-dilemma-tech-policy/
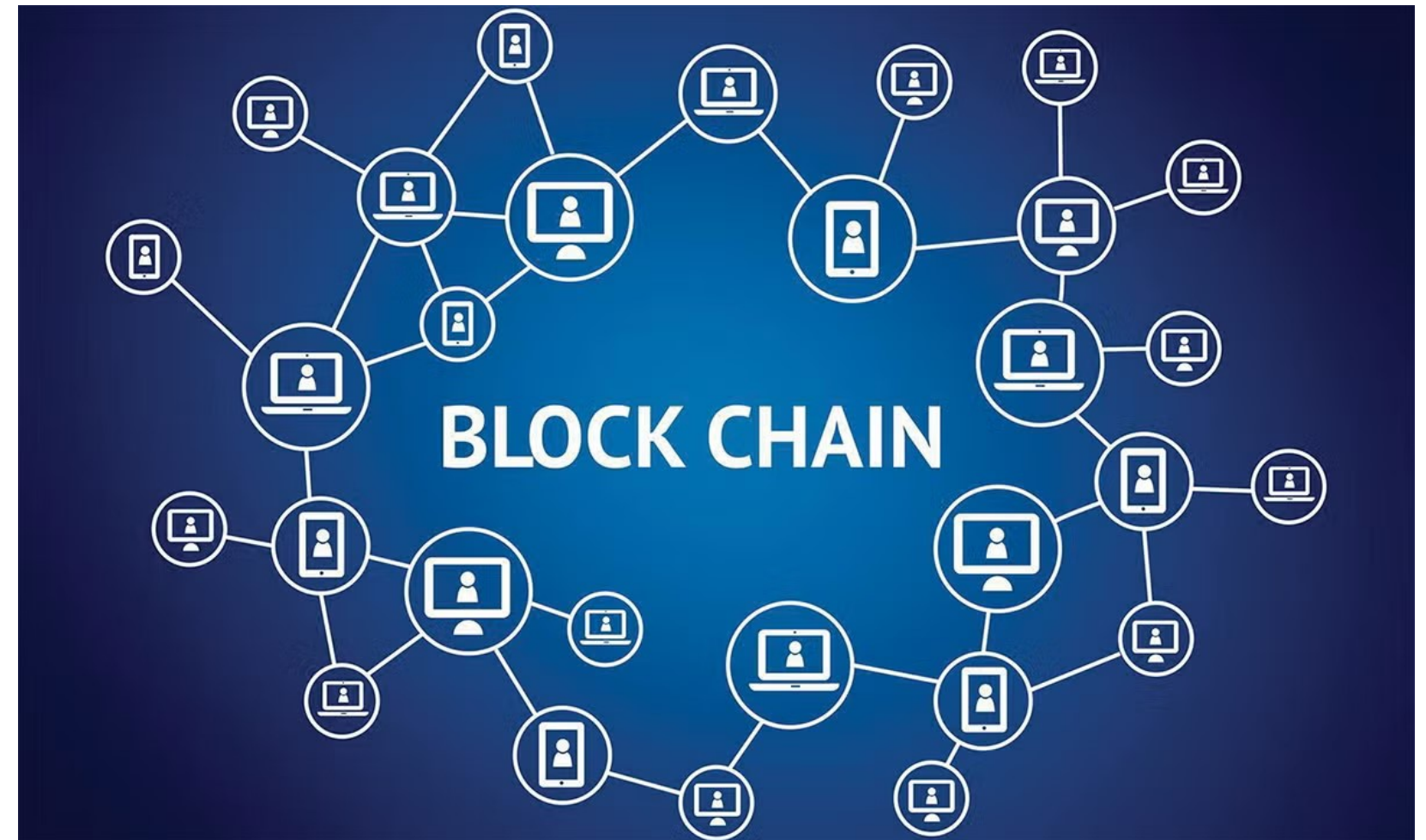
Efforts to influence or control the further development of technology face a double-bind problem:

- An **information problem**: as impacts cannot be easily predicted until the technology is extensively developed and widely used.
- A **power problem**: because control or change over the technology is difficult when it has become entrenched.

Radboud University

Blockchain technology

https://www.binance.com/fr-AF/square/post/16894122544369

BLOCK CHAIN

Given the Collingridge dilemma, do you think we should regulate AI technologies early, despite the uncertainty of their impact, or wait until their consequences are clearer, knowing it might be too late to act by then?
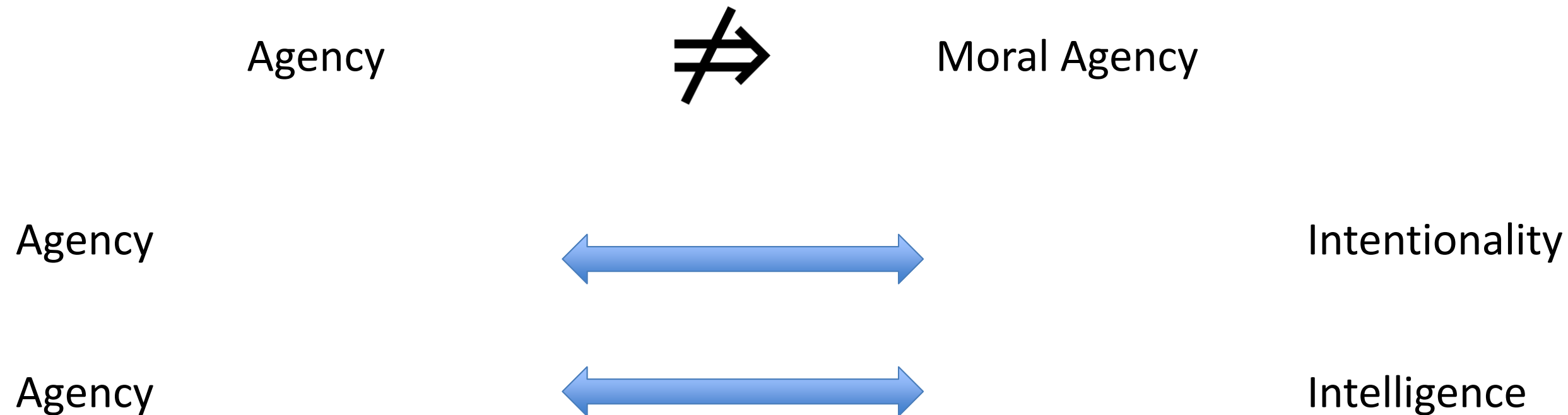
Agency

Agency

"In very general terms, an agent is a **being** with the **capacity** to **act**, and 'agency' denotes the exercise or manifestation of this capacity."

https://plato.stanford.edu/entries/agency/

Agency   ≠   Moral Agency

Agency   ⟷   Intentionality

Agency   ⟷   Intelligence

What does agency entail?

Human vs. Non-human agents

AI and the Future of Human Agency: Are We Outsourcing Decision-Making or Evolving with Machines?

Working Alongside Non-Human Agents

Next week

- **Floridi's theory on AI as a new form of agency (not intelligence!)**

- **Introduction to philosophy of mind**

[AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models](#)