



Between Hammer and Terminator

AI, Law and Philosophy workshop (ALP2023), joint to JURIX 2023
Maastricht University, 18th December 2023

Giovanni Sileno
g.sileno@uva.nl
University of Amsterdam

Tomasz Zurek
t.a.zurek@uva.nl
University of Amsterdam

Context

- Current debates on AI can be summarized on two main axes

Context

- Current debates on AI can be summarized on two main axes



**machines
can NEVER be
like humans**



merely
symbolic
processors

```
Welcome to
EEEEEE LL IIII ZZZZZZZ AAAA
EE LL II ZZ AA AA
EEEEEE LL II ZZZ AAAAAA
EE LL II ZZ AA AA
EEEEEE LLLLLL IIII ZZZZZZZ AA AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
```

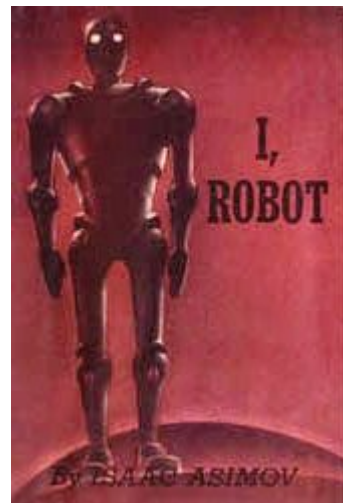
Context

- Current debates on AI can be summarized on two main axes



**machines
can NEVER be
like humans**

**machines
can be
like humans**

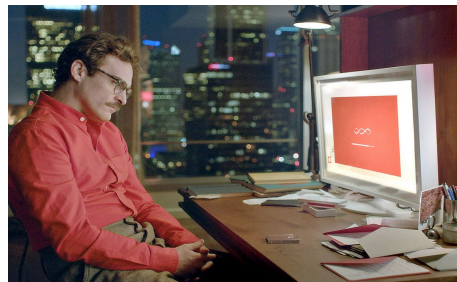


merely
symbolic
processors

```
Welcome to
EEEEEE LL IIII ZZZZZZZ AAAA
EE LL II ZZ AA AA
EEEEEE LL II ZZ AAAAAA
EE LL II ZZ AA AA
EEEEEE LLLLL IIII ZZZZZZZ AA AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
```



responsible,
(possibly) sentient,
(even less possibly)
conscious entities

Context

- Current debates on AI can be summarized on two main axes



**machines
are just
like hammers**

essentially
mechanical
entities



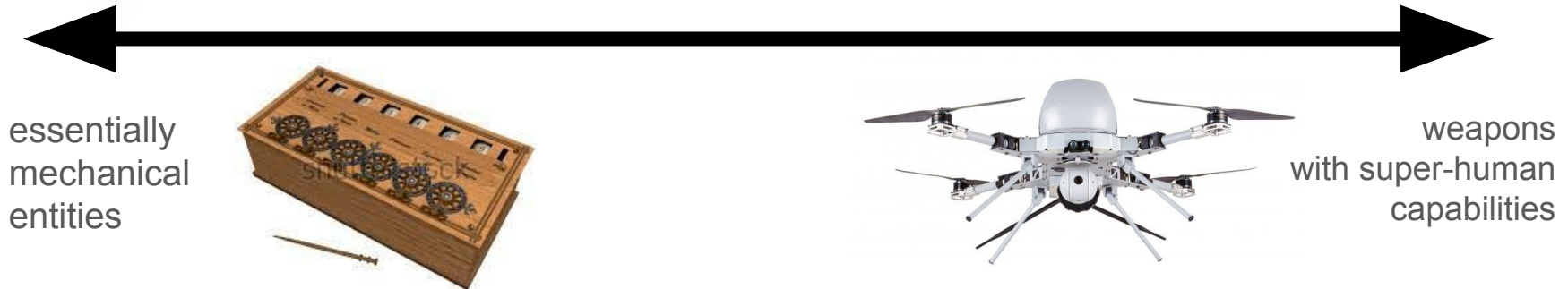
Context

- Current debates on AI can be summarized on two main axes



**machines
are just
like hammers**

**machines
can become
terminators**



Context

- Current debates on AI can be summarized on two main axes



**machines
are just
like hammers**

**machines
can become
terminators***



Context

- Philosophers, technologists, legal experts, ethicists, natural scientists, engineers, CEOs, advisors, journalists, politicians... all take positions!

**machines
can NEVER be
like humans**

**machines
are just
like hammers**

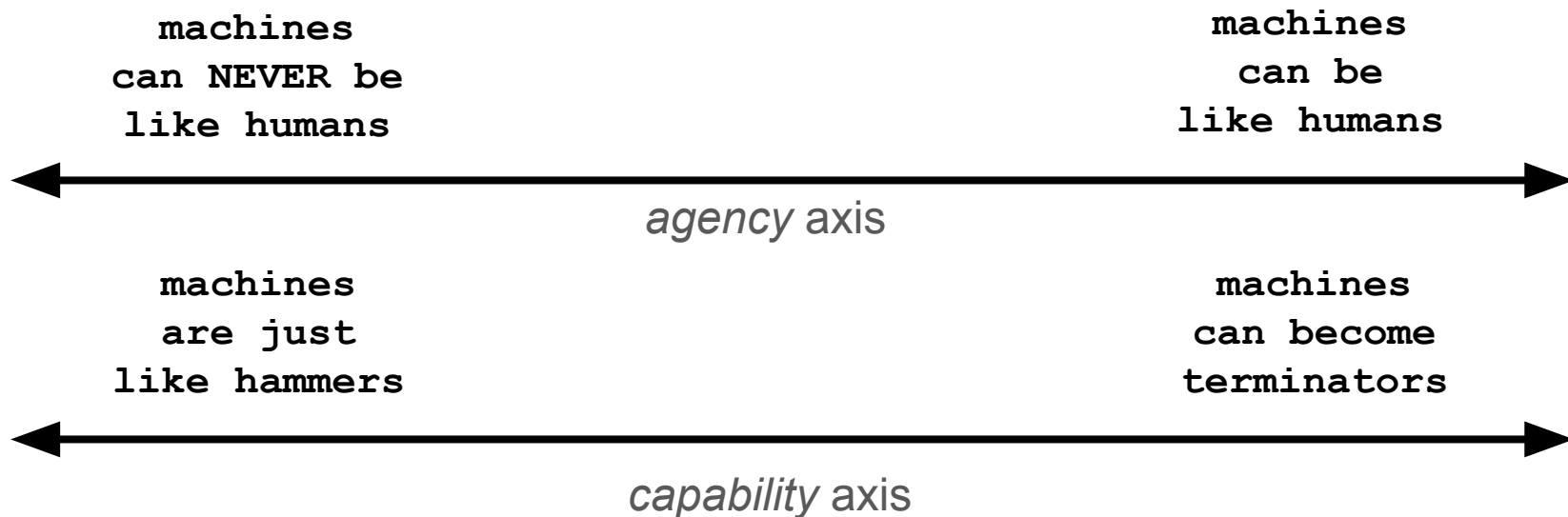


**machines
can be
like humans**

**machines
can become
terminators**

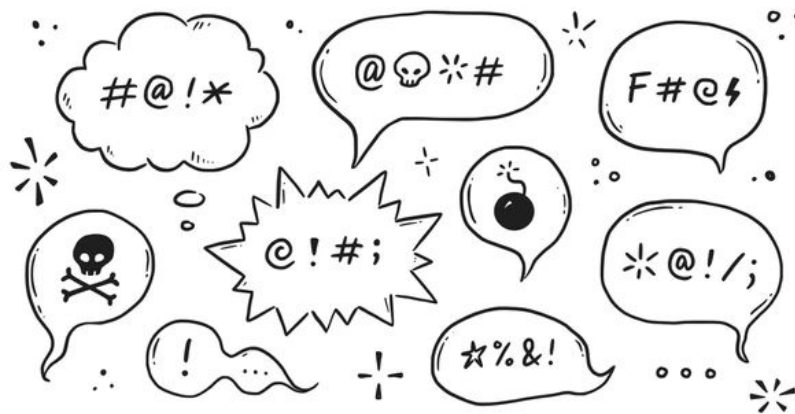
Aim of the paper

- Our paper aims to set up a framework to clarify the concepts at stake, trying to reduce the ambiguity and to unveil assumptions usually left implicit.



Example?

- Consider the word “**responsibility**”: *moral responsibility, legal responsibility, political responsibility, causal responsibility, functional responsibility, (common-sense) responsibility, and so on!*

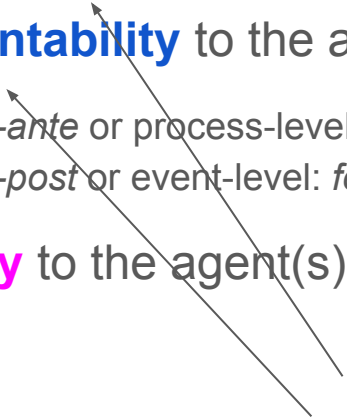


Responsibility per domain

- **responsibility** to the agent(s) determining the action to occur
- **accountability** to the agent(s) justifying the action to occur
 - *ex-ante* or process-level: *auditability, compliance checking*
 - *ex-post* or event-level: *forensics, judiciary activity*
- **liability** to the agent(s) be blamed or praised for the action

Responsibility per domain

- **responsibility** to the agent(s) determining the action to occur
- **accountability** to the agent(s) justifying the action to occur
 - *ex-ante* or process-level: *auditability, compliance checking*
 - *ex-post* or event-level: *forensics, judiciary activity*
- **liability** to the agent(s) be blamed or praised for the action



can be ascribed to artificial entities
(to some extent)

Responsibility per domain

- **responsibility** to the agent(s) determining the action to occur
- **accountability** to the agent(s) justifying the action to occur
 - *ex-ante* or process-level: *auditability, compliance checking*
 - *ex-post* or event-level: *forensics, judiciary activity*
- **liability** to the agent(s) be blamed or praised for the action

can be ascribed to artificial entities
(to some extent)

is always upon humans

Facets of morality (and weaker forms)

- agency
- patiency

Facets of morality (and weaker forms)

- agency —————> requirements to ascribe responsibility of action:
- patiency
 1. **control** on actions
 2. **foreseeability** on outcomes
 3. **evaluative ability** on behaviours

Facets of morality (and weaker forms)

- agency —————> requirements to ascribe responsibility of action:
- patiency
 1. **control** on actions
 2. **foreseeability** on outcomes
 3. **evaluative ability** on behaviours

can be ascribed to artificial entities
(to some extent)

Facets of morality (and weaker forms)

- agency —————> requirements to ascribe responsibility of action:
- patiency

1. **control** on actions
2. **foreseeability** on outcomes
3. **evaluative ability** on behaviours

?

can be ascribed to artificial entities
(to some extent)

Evaluative ability

The process of evaluation can be decomposed into:

- **basic level (*applied morality*)**, specifying:
 - *content*, ie. situations and actions to be evaluated
 - *criteria*, ie. the basis against which to perform the evaluation
 - *process*: ie. how (heuristics, or procedure) to evaluate
 - *acceptance conditions*, ie. when a particular behaviour is acceptable

Evaluative ability

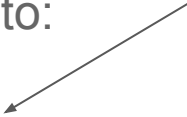
The process of evaluation can be decomposed into:

- **basic level (*applied morality*)**, specifying:
 - *content*, ie. situations and actions to be evaluated
 - *criteria*, ie. the basis against which to perform the evaluation
 - *process*: ie. how (heuristics, or procedure) to evaluate
 - *acceptance conditions*, ie. when a particular behaviour is acceptable
- **meta-level (*volitional morality*)**: how, and on the basis of what we define components at the basic level.

Evaluative ability

The process of evaluation can be decomposed into:

can be ascribed to
artificial entities
(to some extent)

- **basic level** (*applied morality*), specifying: 
 - *content*, ie. situations and actions to be evaluated
 - *criteria*, ie. the basis against which to perform the evaluation
 - *process*: ie. how (heuristics, or procedure) to evaluate
 - *acceptance conditions*, ie. when a particular behaviour is acceptable
- **meta-level** (*volitional morality*): how, and on the basis of what we define components at the basic level.

Evaluative ability

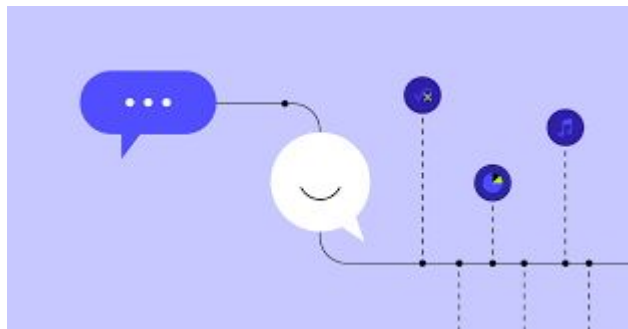
The process of evaluation can be decomposed into:

can be ascribed to
artificial entities
(to some extent)

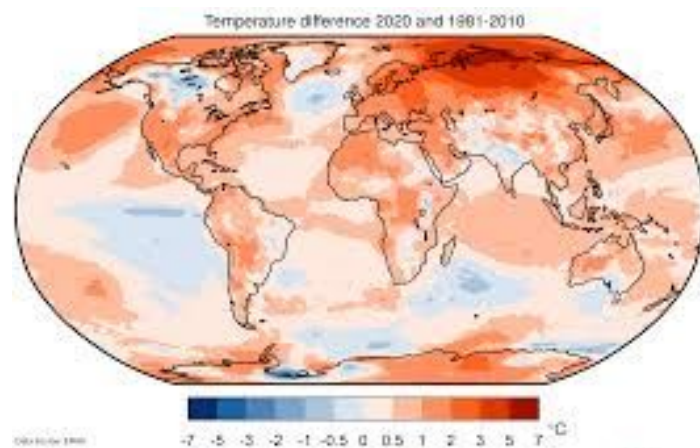
- **basic level** (*applied morality*), specifying:
 - *content*, ie. situations and actions to be evaluated
 - *criteria*, ie. the basis against which to perform the evaluation
 - *process*: ie. how (heuristics, or procedure) to evaluate
 - *acceptance conditions*, ie. when a particular behaviour is acceptable
- **meta-level** (*volitional morality*): how, and on the basis of what we define components at the basic level.

domain of **ethics** (for morality) and **jurisprudence** (for legality)
eventually lies upon humans

Two distinct examples in the paper...



Large Language Models (LLMs),
including ChatGPT



Climate warming

The proposed framework can be applied on both!

A COMPUTER
MORALLY/LEGALLY
CAN NEVER BE HELD ACCOUNTABLE

THEREFORE A COMPUTER MUST NEVER
MAKE A MANAGEMENT DECISION



Between Hammer and Terminator

AI, Law and Philosophy workshop (ALP2023), joint to JURIX 2023
Maastricht University, 18th December 2023

Giovanni Sileno
g.sileno@uva.nl
University of Amsterdam

Tomasz Zurek
t.a.zurek@uva.nl
University of Amsterdam