

PERFORMANCE EXAMPLE 1

What is it?

How is it measured?

In terms of driver? Tesla? AI?



PERFORMANCE EXAMPLE 2

What is it?

How is it measured?

In terms of designer? the AI?

Ethical Issues in Advanced Artificial Intelligence

Nick Bostrom

Oxford University
Philosophy Faculty
10 Merton Street
Oxford OX1 4JJ
United Kingdom

[This is a slightly revised version of a paper

a superintelligence whose top goal is the manufacturing of paperclips, with the consequence that it starts transforming first all of earth and then increasing portions of space into paperclip manufacturing facilities.

research and

As a cognitive pursuit, a superintelligence could also easily surpass humans in the quality of the superintelligence may become unstoppable because of its intellectual power. I survey some of the unique ethical issues in creating superintelligence, and discuss the development of superintelligent machines ought to be accelerated or retarded.

KEYWORDS: Artificial intelligence, ethics, uploading, superintelligence, global security, cost-benefit analysis

1. INTRODUCTION

A *superintelligence* is any intellect that is vastly outperforms the best human brains in practically every field, including scientific creativity, general wisdom, and social skills.^[1] This definition leaves open how the superintelligence is implemented – it could be in a digital computer, an ensemble of networked computers, cultured cortical tissue, or something else.

On the
corporate
level

We need to be careful about what we wish for from a superintelligence, because we might get it.

Several authors have argued that there is a substantial chance that superintelligence may be created within a few decades, perhaps as a result of growing hardware performance and

OUTLINE

What is performance? Experience

What is performance? Definition

Intrinsic value

DEFINITION PERFORMANCE (IN AI)

1. ChatGPT: Describes performance as **mediated** value...



ChatGPT

Measuring the ethical value of the performance of an AI tool involves evaluating several key aspects:

1. **Fairness:** Assess whether the AI treats all user groups equitably without bias.
2. **Transparency:** Determine the extent to which the AI's decision-making processes are understandable and explainable.
3. **Accountability:** Evaluate mechanisms in place for responsibility, especially in cases of error or harm.
4. **Privacy and Security:** Measure how well the AI protects user data and ensures confidentiality.

Seeking performance *in-itself*, intrinsic value: But, what does that mean?

DEFINITION PERFORMANCE (FOCUSED ON THE AI)

Performance

How well the machine serves its purpose

Independent of what the purpose is

Independent of what other values are sacrificed

What are the aims? Where? (In the code? The human designer's head? Elsewhere?)

DEFINITION PERFORMANCE

Hands not on wheel, but car going fine, and paperclips made...

Ethical Issues in Advanced Artificial Intelligence

Nick Bostrom

Oxford University
Philosophy Faculty
10 Merton Street
Oxford OX1 4JJ
United Kingdom

...superintelligence whose top goal is the
...ing of paperclips, with the consequence that
...g first all of earth and then increasing porti
...ip manufacturing facilities.

...s a cognitive pursuit, a superintelligence could also easily
...e the superintelligence may become unstoppable power
...er surveys some of the unique ethical issues in creating s
...ought to give a superintelligence, and introduces some cost-benefit considerations relating to whether the development of superintelligent machines ought to be

...l intelligence, ethics, uploading, superintelligence, global security, cost-benefit analysis

...s any intellect that is vastly outperforms the best human brains in practically every field, including scientific creativity, general wisdo
...en how the superintelligence is implemented – it could be in a digital computer, an ensemble of networked computers, cultured cortic

...need to be careful about what we wish for from
...elligence, because we might get it.

...s have argued that there is a substantial chance that superintelligence may be created within a few decades, perhaps as a result of growing

OUTLINE

What is performance? Experience

What is performance? Definition

Intrinsic value

PERFORMANCE AS INTRINSIC VALUE

What *is* this excellence?

Sensation of purity/transcendence

Imperceptible excellence, idealism (Spiritual, like religious revelation)
(Can the surreal explain the real?)

Innovation (Noun)

Value = the new/original as desirable in itself

Similar to art: What is original *is* good

(Can originality be *itself* the explanation that needs no explanation?)

Innovating (Verb)

Value = the discovering (not what is discovered)

Similar to the traveler who buys a one way ticket (Journey, not destination)...

(Like artist maintaining rights to creation while selling the created....)

PERFORMANCE AS INTRINSIC VALUE

What *is* this excellence?

Efficiency, possibly more generally as visible excellence...

Value = the efficiency, not what is efficiently produced...

Similar to Kant's sublime: art excellent for formal reasons like symmetry...

(Can great art be separated from what it depicts?)

(Can great technology be separated from what it does?)

Gratification/fulfillment (of the AI engineer)

Feeling of accomplishment

(Virtue ethics – the idea of a developed skill...)

OUTLINE

What is performance? Experience

What is performance? Definition

Intrinsic value