

Exercises

1	2	3	4
---	---	---	---

Surname, First name

Filosofie en AI (5082FIAI6Y)

24-03-2025 Tentamen

24 Mar, 2025, 13:30-15:30

1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9
0	0	0	0	0	0	0	0

a	<input checked="" type="radio"/>	c	d	e	f	→ b
<input checked="" type="radio"/>	b	c	<input checked="" type="radio"/>	e	f	→ a
<input checked="" type="radio"/>	b	<input checked="" type="radio"/>	d	e	f	→ c
<input checked="" type="radio"/>	b	<input checked="" type="radio"/>	<input checked="" type="radio"/>	e	f	→ d

Answer multiple-choice questions as shown in the example.

BEFORE YOU START

- Please **wait** until you are instructed to open the booklet.
- Check if your version of the exam is complete.
- Write down **your name** and **student ID number**.
- Your **mobile phone** has to be switched off and in the coat or bag. Your **coat and bag** must be under your table.

PRACTICAL MATTERS

- The first 30 minutes and the last 15 minutes you are not allowed to leave the room, not even to visit the toilet.
- You are obliged to identify yourself at the request of the examiner (or his representative) with a proof of your enrollment or a valid ID.
- During the examination it is not permitted to visit the toilet, unless the invigilator gives permission to do so.
- 15 minutes before the end, you will be warned that the time to hand in is approaching.

EXAM

- Check before handing in that you have answered all questions.
- Try to stay within the boxes when writing your answer.
- Don't be afraid to leave lines in the boxes empty! Many questions can be answered briefly, and shorter is often better, the extra lines are there for large handwriting and/or to make corrections.

Good luck!



Multiple Choice

- 5p **1a** In high-stakes scenarios like medical triage, a purely consequentialist approach can be criticized for:
- ☐ (a) Relying solely on qualitative assessments.
 - ☐ (b) Prioritizing accuracy over human empathy.
 - ☐ (c) Being too rigid to adapt to emergent situations.
 - ☐ (d) Neglecting individual rights by favoring aggregate outcomes.
- 5p **1b** Heidegger's concept of "enframing" suggests that technology does what?
- ☐ (a) Technology allows humans to control nature more effectively.
 - ☐ (b) Technology reveals the world primarily as a resource to be optimized.
 - ☐ (c) Technology is a natural extension of human capabilities.
 - ☐ (d) Technology should be developed with ethical considerations in mind.
- 5p **1c** An AI uses casuistry (i.e., case-based reasoning) to decide ethical actions. What is a significant limitation of this approach?
- ☐ (a) It overly relies on rigid, predefined rules.
 - ☐ (b) It may perpetuate biases from historical data.
 - ☐ (c) It relies on principles which are too vague.
 - ☐ (d) It cannot generalize across different demographics.
- 5p **1d** Kripke's theory of reference differs from descriptivist theories by stating that:
- ☐ (a) The reference of a name is based on the context.
 - ☐ (b) Names refer through a historical chain of events.
 - ☐ (c) The reference of a name is entirely determined by internal mental states.
 - ☐ (d) Reference is established by logical equivalence alone.

5p **1e** What is a key characteristic of “language games” as described by Wittgenstein?

- ☐ a Language games rely solely on formal logic and fixed definitions.
- ☐ b Language games show that meaning is dependent on social practices.
- ☐ c Language games claim that meaning is predetermined and static.
- ☐ d Language games claim that syntax comes before semantics.

5p **1f** What is the key idea behind the institutional theory of art?

- ☐ a Art is defined by its adherence to beauty standards.
- ☐ b Art’s value is determined by its originality.
- ☐ c Art is what the art world recognizes as art.
- ☐ d Art’s value is determined by its technical innovation.

Short questions

- 8p **2a** Describe the difference between top-down approaches and bottom-up approaches in machine ethics. Give one example for each.

- 8p **2b** Describe Frege's distinction between sense and reference for *names* (e.g., Hesperus), and for *sentences* (e.g., Hesperus is bright).





- 8p **2c** What is the Grounding Problem as it pertains to AI systems? Explain also how Reinforcement Learning from Human Feedback (RLHF) might improve grounding in language models.

- 8p **2d** Outline what counts as art according to the Mimetic Theory and the Expressive Theory in aesthetics. Then, for each theory, explore one key implication for AI-generated art (this could be a challenge, an analogy, or a central question that arises when considering AI-created works in relation to the theory).



AI Box Experiment

- 5p **3a** Outline the AI Box Experiment proposed by Eliezer Yudkowsky in the topic of AI containment. Explain what the implications of the experiment are.



4p **3b** Identify and explain one critique of the AI Box Experiment.

4p **3c** Formulate a response for the critique you identified.



- 4p **3d** Discuss whether containment measures can effectively eliminate the risks demonstrated by the AI Box Experiment. What other measures might be necessary to ensure AI systems do not act in ways that could be harmful or unpredictable?

Rawls and Fairness

- 4p **4a** Outline John Rawls' "veil of ignorance" thought experiment (also known as "original position") and explain its intended role in promoting fairness.



[illegible]

5p

[illegible]

4c

[illegible]

4d

- 4p **4e** Finally, discuss how Rawls' framework can be modified or combined with other approaches to mitigate the criticism you discussed in the previous question.



This page is left blank intentionally