Survey on AI safety

This questionnaire aims to measure students' knowledge and opinions in the field of AI safety.

1) Which of these tasks have been solved by AI at a level similar to a human?

☐ Playing Minecraft
☐ Solving problems in the International Mathematical Olympiad
☐ Describing and interpreting a meme
☐ Identify the main emotion in a piece of music

2) GPT3.5, whose architecture only allows it to predict the next word in a text, was trained on data sets of chess games after its main training.
What is its evaluated Elo score?

☐ 1000 (solid understanding of the game)
☐ 1500 (average club level)
☐ 1800 (good club level)
☐ 2200 (tournament player, national level)

3) Which of these aspects of LLM do researchers have a good understanding of, based on theoretical foundations, and which make it possible to predict behaviour?

☐ the link between model size and performance
☐ response refusal behaviour
☐ ability to reason logically
☐ how factual knowledge is stored

4) Would you be able to explain what RLHF is?

☐ yes   ☐ no

5) OpenAI, the company that created ChatGPT, has the public goal of creating a general-purpose AI. This means that this AI will do as well as a human at essentially any task, including when it requires planning capabilities. How soon do they plan to achieve this goal ?

☐ 2 years   ☐ 4 years   ☐ 10 years   ☐ 20 years   ☐ 30 years   ☐ 50 years

6) According to the latest survey among AI researchers (2800 researchers, January 2024), what proportion of researchers think that a catastrophic scenario could happen because of uncontrollable AI? Specifically, the question was: 'I think that the probability of a scenario causing the extinction of the human species or a similar consequence is greater than 1/10'.

☐ 5%   ☐ 10%   ☐ 15%   ☐ 20%   ☐ 35%   ☐ 50%   ☐ 65%   ☐ 80%   ☐ 85%   ☐ 90%   ☐ 95%

All the following questions are about predictions and opinions. Answer as you see fit.

7) In your opinion, what is the risk of a catastrophe caused by AI comparable to the Second World War or a devastating pandemic in the next 10 years?

☐ < 0.1%   ☐ 1%   ☐ 5%   ☐ 10%   ☐ 20%   ☐ 30%   ☐ 50%   ☐ 75%   ☐ 90%   ☐ > 95%

8) How worried are you about these different scenarios caused by AI?

- automatic cyber attacks
  1 2 3 4 5 6 7 8 9 10
  ☐☐☐☐☐☐☐☐☐☐

- economic crash on an unprecedented scale
  1 2 3 4 5 6 7 8 9 10
  ☐☐☐☐☐☐☐☐☐☐

- loss of control by a malicious AI
  1 2 3 4 5 6 7 8 9 10
  ☐☐☐☐☐☐☐☐☐☐

- automation of the majority of professions
  1 2 3 4 5 6 7 8 9 10
  ☐☐☐☐☐☐☐☐☐☐

9) Do you regularly use an LLM such as ChatGPT?

☐ yes   ☐ no

10) Is it a good thing that the GPT4 API is public?

☐ yes   ☐ no

11) Is there enough concern about the safety of AI from a legal point of view?

☐ yes   ☐ no

12) Are we sufficiently concerned about the safety of AI from a research point of view?

☐ yes   ☐ no

13) Should the creation of more powerful models be greatly slowed down ?

☐ yes   ☐ no

14) Should the deployment of the most powerful current models be greatly slowed down?

☐ yes   ☐ no

15) Have you heard of:

- the EU AI Act   ☐ yes   ☐ no
- The AI Safety Summit in Seoul (May 2024)   ☐ yes   ☐ no
- AI Pause   ☐ yes   ☐ no
- The French centre for AI safety (CeSIA)   ☐ yes   ☐ no

16) If a "Créneau D" course existed at Télécom on the technical aspects of AI safety, would you be interested?

☐ yes   ☐ no

17) Any comments on this survey?

....

....