Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

1. Which of the following statements about Reinforcement Learning from Human Feedback (RLHF) are true? **1 / 1 point**

- ○ After applying RLHF, an LLM will reflect a similar degree of bias and toxicity as text on the internet.
- ○ RLHF fully addresses all concerns about AI.
- ◉ RLHF helps to align an LLM to human preferences, and can reduce the bias of an LLM's output.
- ○ RLHF is a common technique for training a small (say 1B parameter) LLM to do as well as a larger (say 10B parameter) one.

Correct
RLHF trains models to produce output that better aligns with human preferences, including honesty, helpfulness, and harmlessness. The process can reduce biases in an LLMs output.

2. **True or False.** Because AI automates tasks, not jobs, absolutely no jobs will disappear because of AI. **1 / 1 point**

- ○ True
- ◉ False

Correct
Even if all of the tasks of a role can't be completely automated, some jobs may be eliminated as efficiency increases and cost savings can be realized. It is important that we support the individuals

3. If we manage to build Artificial General Intelligence (AGI) some day, which tasks should AI be capable of performing? (Check all that apply.) **1 / 1 point**

- [x] Write a software application to let users manage their household spending budgets.

  Correct
  By definition, AGI can carry out any intellectual task that a human can do. Since software applications like this already exist in the world, the AGI should be able to write one from scratch.

- [x] Learn to drive a car in roughly 20 hours of practice.

  Correct
  By definition, AGI can carry out any intellectual task that a human can do. So it should be able to learn to drive a car in roughly 20 hours, just like a human teenager.

- [x] Compose the music for a movie soundtrack.

  Correct
  By definition, AGI can carry out any intellectual task that a human can do. So it should be able to create music for a movie soundtrack.

- [ ] Predict the future (such as make stock market and weather predictions) with perfect accuracy.

4. You are working on a chatbot to serve as a career coach for recent college graduates. Which of the following steps could you take to ensure that your project follows responsible AI? (Check all that apply.) **1 / 1 point**

- [x] Engage diverse recent college graduates and ask them to offer feedback on the output of your chatbot.

  Correct
  Working with diverse stakeholders can help you identify problems your own team may not recognize, and ensure that the behavior of your chatbot takes into account the perspectives of people from diverse backgrounds.

- [x] Organize a brainstorming session to identify problems that could arise for users chatting with the career coach.

  Correct
  Building a culture that encourages discussion and debate of ethical issues can help you identify problems early in the development phase and avoid issues of bias or toxicity later in the process.

- [ ] Allow a single engineer on your team to determine whether the output of the chatbot is helpful, honest, and harmless.

- [x] Engage employers (because they are a key stakeholder group) and ask them to offer feedback on the output of your chatbot.

  Correct
  Working with all stakeholders can reveal points of view that your team may not have realized and can help identify problems or issues that may have been missed.

5. Now that you've made it to the end of the course, which of these statements are true? (Please check all, because all apply!) **1 / 1 point**

- [x] You understand how Generative AI technology works, and what it can and cannot do.

  Correct

- [x] You're well positioned to use Generative AI responsibly to help yourself and others.

  Correct

- [x] You've achieved the significant accomplishment of finishing this course.

  Correct

- [x] Andrew is thrilled at your completing, and sends you his warmest thank you and congratulations!

  Correct