Back **Week 3 Quiz**
Graded Assignment • 1h

coursera

Due Nov 19, 11:59 PM +01

Your grade: 80%

Your latest: **80%** • Your highest: **80%**

To pass you need at least 80%. We keep your highest score.

Graded Assignment

Week 3 Quiz

Review Learning Objectives

1. Which of the following are true in regards to Constitutional AI? Select all that apply. **1 / 1 point**

☑ In Constitutional AI, we train a model to choose between different responses.

✓ **Correct**
This is the role of the preference model, that will learn what responses are preferred following the constitutional principles.

□ For constitutional AI, it is necessary to provide human feedback to guide the revisions.

☑ Red Teaming is the process of eliciting undesirable responses by interacting with a model.

✓ **Correct**
Red Teaming is the process of eliciting undesirable responses, and it is necessary for the first stage of Constitutional AI, as we need to fine-tune the model with those “red team” prompts and revised answers.

☑ To obtain revised answers for possible harmful prompts, we need to go through a Critique and Revision process.

✓ **Correct**
This process is necessary for Constitutional AI, and its done by asking the model to critique and revise the elicited harmful answers.

2. What does the "Proximal" in Proximal Policy Optimization refer to? **1 / 1 point**

Item Navigation

Next item →

Assignment details

Due
Nov 19, 11:59 PM +01

Attempts
Unlimited

Try again

Submitted
Nov 6, 12:47 AM +01

Your grade
To pass you need at least 80%. We keep your highest score.

80%

View submission **See feedback**

Like **Dislike** **Report an issue**