

🏠 Topics

👤 My Posts

⋮ More

▼ RESOURCES

📖 Documentation

🔗 API reference

🔗 Help center

▼ CATEGORIES

Announcements

API

Plugin Development

Prompting

Documentation

☰ All categories

▼ TAGS

🏷️ chatgpt

🏷️ gpt-4

🏷️ api

🏷️ plugin-development

🏷️ codex

☰ All tags

▼ MESSAGES

📧 Inbox

Skip to main content

+

🔍

Welcome to the OpenAI Developer Forum!

✕

What to know before posting a new question:

1. Search the forum for similar topics - the question might have been discussed before.

2. If the question relates account issues (e.g., billing and login issues), please contact us through our [Help Center](#).

3. Please be kind and helpful in conversations!

RLHF after Fine-Tuning Davinci?

vp12

7d

Hi, I fine-tuned a model for code generation and while it is good, it could be better. I have already tried adjusting hyperparameters and data. I want to try using the InstructGPT process (RLHF → PPO) but I am having trouble figuring out if it is possible using the API.

Any help is appreciated, Thanks.

markhennings

6d

How many examples are in your training set for the model? At what scale are you trying to do RLHF?

One simplified way to improve your dataset with human feedback is to prompt your model to generate multiple completions (let's say 3) at a temperature greater than 0.

Then select the best one to save to your dataset.

j

6d

RLHF → fine-tune? Feel like hitting a button 78 thousand times? (or letting GPT-4 first pick by criteria)

file:///C:/Users/int.mingyuan.ma/OneDrive - mails.tsinghua.edu.cn/Research/2306 燧原实习/ArchResearchDevelop/docs/files/RLHF after Fine-Tun... 1/6

We released **WizardCoder-15B-V1.0** (trained with **78k** evolved code instructions), which surpasses **Claude-Plus (+6.8)**, **Bard (+15.3)** and **InstructCodeT5+ (+22.3)** on the [HumanEval Benchmarks](#).

Narrow down single instruction examples like this, on 175b davinci, to a LoRA specialization, and it just might kill gpt-3.5:

“all_code”: "Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction:

Create a Python script for this problem:

```
import math
```

```
def poly(xs: list, x: float):
```

```
    """
```

Evaluates polynomial with coefficients xs at point x.

```
    return xs[0] + xs[1] * x + xs[1] * x^2 + ... xs[n] * x^n
```

```
    """
```

```
    return sum([coeff * math.pow(x, i) for i, coeff in
                enumerate(xs)])
```

```
def find_zero(xs: list):
```

```
    """ xs are coefficients of a polynomial.
```

```
    find_zero find x such that poly(x) = 0.
```

```
    find_zero returns only only zero point, even if there
    are many.
```

```
    Moreover, find_zero only takes list xs having even
    number of coefficients
```

```
    and largest non zero coefficient as it guarantees
    a solution.
```

```
>>> round(find_zero([1, 2]), 2) # f(x) = 1 + 2x
-0.5
```

```
>>> round(find_zero([-6, 11, -6, 1]), 2) # (x - 1) * (x -
2) * (x - 3) = -6 + 11x - 6x^2 + x^3
1.0
```

```
    """
```

Response:The script for the problem is as follows:

```
import math
```

```
def poly(xs: list, x: float):
```

[Skip to main content](#)

```

"""
Evaluates polynomial with coefficients
return xs[0] + xs[1] * x + xs[1] * x^2
"""

return sum([coeff * math.pow(x, i) for i in range(len(xs))])

def find_zero(xs: list):
    """ xs are coefficients of a polynomial
    find_zero find x such that poly(x) = 0
    find_zero returns only one zero point
    Moreover, find_zero only takes list xs
    and largest non zero coefficient as its
    a solution.
    >>> round(find_zero([1, 2]), 2) # f(x)
    -0.5
    >>> round(find_zero([-6, 11, -6, 1]),
    1.0
    """

    if len(xs) % 2 != 0:
        raise ValueError("Length of xs should be even")
    max_coeff = max(abs(coeff) for coeff in xs)
    xs = [coeff / max_coeff for coeff in xs]
    for i in range(len(xs) // 2):
        if xs[i] != 0:
            break
    xs = xs[i:] + xs[:i]
    for i in range(len(xs) // 2):
        if xs[i] != 0:
            break
    xs = xs[i:] + xs[:i]
    a = xs[0]
    b = xs[1]
    c = xs[2]
    d = xs[3]
    e = xs[4]
    f = xs[5]
    g = xs[6]
    h = xs[7]
    if a == 0:
        if b == 0:
            if c == 0:
                if d == 0:
                    if e == 0:
                        if f == 0:
                            if g == 0:
                                if h == 0:
                                    raise ValueError("All coefficients are zero")
                                else:
                                    return -h/g
                            else:
                                return -g/f
                        else:
                            return -f/e
                    else:
                        return -e/d
                else:
                    return -d/c
            else:
                return -c/b
        else:
            return -b/a
    else:
        return -a/b

```

[Skip to main content](#)

```

        else:
            return -g
    else:
        return -f / 4
    else:
        return -e / 6
    else:
        return -d / 8
    else:
        return -c / 10
    else:
        return -b / 12
    else:
        return -a / 14

```

The `poly` function takes a list of coefficients `xs` and a point `x` and returns the value of the polynomial at that point. The `find_zero` function takes a list of coefficients `xs` and returns the value of `x` such that the polynomial `poly(x)` is equal to zero. The function first checks if the length of `xs` is even, and raises a `ValueError` if it is not. It then normalizes the coefficients by dividing them by the maximum absolute value, so that the largest non-zero coefficient is 1. It then rearranges the coefficients so that the first non-zero coefficient is at the beginning of the list. It then uses the coefficients to find the value of `x` using the method of Horner's rule. The function returns the value of `x` as a float."

vp12
6d

Currently I have a fine tuned model with about 500 samples it was trained on. It is somewhat accurate, it generates decent code but it is still not good enough to compile. I want to improve it as much as possible before working on post processing functions.

My question is that I fine-tuned an openai model so I do not have direct access to update the model based on the feedback. Do you believe that adding to the dataset could work? Or maybe retraining with the new data?

[Skip to main content](#)

J

6d

That’s GPT-4 - a coding specialist among its expert models.

Except with ridiculous thousands and thousands of code tunes, possibly synthetic.

The WizardCoder above has opensource training in long answering. The largest percentage is interpreted languages, you could start by classifying the language of them to see what you can use.

You’d have to assure yourself that you can beat GPT-4 by your own training – making the tuned model worth its cost of twice as much per token as GPT-4, and still having half the context.

markhennings

6d

Personally, I would review the examples in the dataset to make sure they compile, as well as add examples. If you are getting decent results then you are on the right track.

Later this year, GPT 3.5 Turbo and GPT-4 will become available for fine-tuning (according to OpenAI [they are working on this](#)), and you can try the same dataset on the improved base model, probably leading to large improvements in accuracy, and continue to iterate on your results.

Related Topics

Topic	Replies	Views	Activity
Fine-Tune Davinci to write programming language codex	2	396	May 3
How to fine tune a model using GPT-3 models?	1	593	Jan 18

[Skip to main content](#)

Topic	Replies	Views	Activity
Fine Tuning text-davinci-003 Models	9	3.4k	Feb 15
Fine tuned model is hallucinating fine-tuning, davinci, hallucinations	3	97	15d
Fine-tuning GPT to learn a new coding language chatgpt, api, plugin-development, codex, fine-tuning	1	275	May 18

OpenAI ©
2015–2023

The OpenAI developer forum is a place to connect with other people building with OpenAI models.

Research
[Overview](#)
[Index](#)

Product
[Overview](#)
[Customer stories](#)
[Safety standards](#)
[Pricing](#)

Safety
[Overview](#)
[Security](#)

Help
[Support](#)

Developers
[Documentation](#)
[Service status](#)
[Examples](#)

Company
[About](#)
[Blog](#)
[Careers](#)
[Charter](#)

[Terms & policies](#) [Privacy policy](#) [Brand guidelines](#)

