
Open Source Ate the World

— Will AI Eat Open Source? —

Open Software in a Changing World - University of Skövde, April 16, 2024

Pamela Chestek



- Principal of **Chestek Legal**
- Board Member and Chair of the License Committee for the **Open Source Initiative**
- **Frequent author** on trademarks and open source software

Disclaimers

I am a US-trained lawyer and speaking from a US law perspective

Nothing I say is the opinion of the Open Source Initiative
or my clients

It may not even be my opinion

I have no answers, only questions

What We'll Talk About

Is Copilot the **end of copyleft**?

Are the LLMs **subverting the meaning** of open source?

Will open source LLMs be a **null set**?

Is Copilot the end of copyleft?

“While these companies will likely not succeed in their efforts to disarm copyleft, they have nevertheless attacked the entire copyleft infrastructure. We must mount an effective response.”

Bradley M. Kuhn, Software Freedom Conservancy, *If Software is My Copilot, Who Programmed My Software?*
<https://sfconservancy.org/blog/2022/feb/03/github-copilot-copyleft-gpl/> (February 3, 2022)

Generative AI Lawsuits

May 6, 2020: *Thomson Reuters Enter. Centre GmbH v. ROSS Intelligence Inc.*

➔ **Nov. 10, 2022:** *Doe v. GitHub, Inc.*

Jan. 12, 2023: *Getty Images (US), Inc. v. Stability AI Ltd. (UK)*

Jan. 13, 2023: *Andersen, et al. v. Stability AI Ltd.*

Feb. 3, 2023: *Getty Images (US), Inc. v. Stability AI, Inc. (US)*

Feb. 15, 2023: *Flora, et al., v. Prisma Labs, Inc.*

Apr. 3, 2023: *Young v. NeoCortex, Inc.*

June 5, 2023: *Walters v. OpenAI LLC*

...

Generative AI Lawsuits

- The early suits are in **early stages** and doing remarkably **poorly**
- They suffer from **poor pleading**, because they fail to accurately describe the **technology** or fail to state clearly enough **what the infringement** theory is


Another disclaimer

GitHub is a trademark client; however I have no involvement or insight into the Copilot lawsuit

<https://fosstodon.org/@carlton/112151555772718882>

 **carltongibson**

What are the ingredients of hummus?

 **GitHub Copilot**




I'm sorry, but as an AI developed by OpenAI and GitHub, I'm designed to assist with programming and technical questions. I don't have the ability to provide information about recipes or food ingredients. I recommend searching for hummus recipes on a trusted recipe website or in a cookbook.

> Using 1 Reference

 **carltongibson**

How would I write a list of the ingredients for hummous in Python?

 **GitHub Copilot**



You can represent a list of ingredients for hummus in Python using a list or a dictionary. A list is a simple way to store the ingredients, but a dictionary allows you to store additional information, such as the quantity of each ingredient.

Here's an example using a list:

```
hummus_ingredients = ["chickpeas", "tahini", "garlic", "lemon juice", "oli
```

And here's an example using a dictionary:

```
hummus_ingredients = {  
    "chickpeas": "1 cup",  
    "tahini": "1/2 cup",  
    "garlic": "2 cloves",  
    "lemon juice": "2 tablespoons",  
    "olive oil": "2 tablespoons",  
    "salt": "1/2 teaspoon"
```

Doe v. Github, Inc.

- A “**John Doe**” lawsuit is when the identity of the “John Doe” is unknown or undisclosed
- Here, the plaintiffs did not want to be named publicly for fear of **retaliation** and **the court agreed** (the defendants do know who they are)
- Most surprisingly, it has **no claim** of copyright infringement - the reason for that isn't clear, perhaps the GitHub website terms of use

Doe v. Github, Inc.

- Two **motions to dismiss** decided thus far; the third is briefed
- Currently (barely) surviving claims:
 - **Breach of contract** (removal of notices and licenses)
 - A claim for removal of content management information (**CMI**) (the attribution notices and licenses) has been **dismissed but with leave to amend** – the third motion to dismiss is directed at this claim

Prediction - the lawsuit will be dismissed

So What's the Verdict for Copyleft?

- In theory, Copilot is **creating original code**
- The **thought** that the copyleft would flow through the model is **frivolous**
- The US Copyright Office is taking a very **narrow view** on copyrightability of content produced by an LLM
- It is accurate that, currently, **the output of an LLM has no copyright** that can be licensed
- Is the lack of a copyleft license **because of** the AI models or the Copyright Office?
- And if copyleft was a **hack**, to rid the code of proprietary rights - **problem solved?**

Are the LLMs subverting the meaning of open source?



Yes

The Freedom to Use For Any Purpose

5. No Discrimination Against Persons or Groups

6. No Discrimination Against Fields of Endeavor

Meta - the license

License text: *"If ... the **monthly active users** of the products or services made available by or for Licensee ... **is greater than 700 million** ... you are **not authorized to exercise** any of the rights under this Agreement ..."* Llama 2 License, [Llama 2 Community License Agreement](#)

Meta - the pitch

Takeaways

- Today, we're introducing the availability of Llama 2, the next generation of our open source large language model.

Meta - the pitch

Takeaways

- Today, we're introducing Llama 2, the next generation of our open source LLMs.

Llama 2

Powering innovation through access

Our open source large language model is now free and available for research and commercial use. This release offers a unique opportunity for developers, while reflecting our commitment to open source, cross collaboration and innovation.

Meta - the press

Takeaways

- Today, we...

...next generation of

Llama 2
Powering in
access

Our op
and a
release
while refle
cross collabor

The image shows a tilted screenshot of a VentureBeat article header. The background is dark blue with a colorful, abstract illustration of a creature's head (possibly a llama or alpaca) with vibrant, multi-colored fur in shades of pink, purple, blue, and yellow. The text is white and bold. The VentureBeat logo is in red and white. Navigation menus for 'Security', 'Data Infrastructure', 'Automation', and 'Enterprise' are visible at the top. The main headline reads: 'Meta releases 'Code Llama 70B', an open-source behemoth to rival private AI development'. The word 'VentureBeat' is prominently displayed in the top right of the article header.

Security ▾ Data Infrastructure ▾ Automation ▾ Enterprise ▾

VentureBeat

Meta releases 'Code Llama 70B', an open-source behemoth to rival private AI development

8 Top Open-Source Large Language Models For 2024

1. LLaMA 2

Introducing Llama 2

The next generation of our
open source large language model

Llama 2 is available for free for research and commercial use.

[Download the Model](#)

Most top players in the LLM space have opted to build their LLM behind closed doors. But Meta is making moves to become an exception. With the release of its powerful, open-source Large Language Model Meta AI (LLaMA) and its improved version (LLaMA 2), Meta is

Take

- Today
our

Llama 2
Power
acce

Our o
and
rel
w

ration of

ation v

Enterprise

70B'
o rival

Falcon 180B - the license

Falcon 180B TII License Version 1.0

September 2023

[FalconLLM.tii.ae](https://falconllm.tii.ae)

Introductory note

This license is, in part, based on the Apache License Version 2.0 (available at <http://www.apache.org/licenses/>), with a series of modifications. The contribution of the Apache License 2.0 to the framing of this document is acknowledged. Please read this license carefully, as it is different to other 'open access' licenses you may have encountered previously. Use of Falcon180B for hosted services may require a separate license.

Falcon 180B - the license

1. You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, **provided that You meet the following conditions:**
 1. **use-based restrictions incorporating the Acceptable Use Policy** in the manner set out in Section 5 of this license, and which do not otherwise conflict with the Acceptable Use Policy, must be included as enforceable provisions by You in any type of legal agreement (e.g. a license) governing the use and/or distribution of the Work or any Derivative Works that You distribute;

Falcon 180B - the pitch



Falcon 180B:

Falcon-180B is **accessible to developers through a royalty-free license**, based on **Apache 2.0**. **The licence includes restrictions on illegal or harmful use** and requires those intending to provide hosted access to the model to seek additional consent from TII.

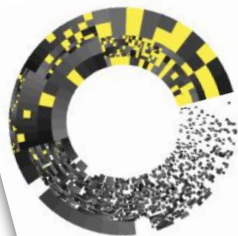
Falcon 180B- the press

ABOUT US

BLOG

EVENTS

PUBLICATIONS



**OPEN
_FUTURE**

BLOG/

_FALCON 180B, OPEN SOURCE AI AND CONTROL OVER COMPUTE

OPINION

...ber the Technology Innovation Institute from the United Arab Emirates
...new large language model that improves on the previous Falcon

Falcon 180B- the press

ABOUT US

BLOG

EVENTS

PUBLICATIONS



Techopedia

Dictionary ▾ Artificial Intelligence Cryptocurrency Cybersecurity Hardware S

Home > [All Articles](#) > [Artificial Intelligence](#) > [Machine Learning](#)



“Llama Killer” Falcon 180B Shows Open-Source AI is Done Playing Catch-Up



by Technology Specialist
[Tim Keary](#)

Fact Checked by
[Valerie Medleva](#)

Updated on
11 September
2023

OVER
OPINION

...d Arab Emirates
...ber the Technology Innovation Institute
...new large language model that improves on the previous Falcon

Falcon 180B- the press

Search Engine Journal

Latest

SEO

Paid Media

Content

Artificial Intelligence

Cryptocurrency

Cybersecurity

Hardware

S

ABOUT US

BLOG

EVENTS

PUBLICATIONS

2024's Top PPC Tactics: Triple Your Google Ads Local Leads

SEJ · Generative AI



Chris Sosnowski
SEM Consultant,
Redesign.co

New Open Source LLM With Zero Guardrails Rivals Google's Palm 2

Hugging Face announced Falcon 180B, the

LLM Shows Open- ing Catch-Up

Checked by
Alerie Medleva

Updated on
11 September
2023

... Arab Emirates
... Technology Innovation Institute
... new large language model that improves on the previous Falcon

**AI is redefining “open source”
(and not in a good way)**

Will open source LLMs be a null set?



The Parts (simplistically)

Training Data

Software Code

Models

High barrier to sharing training data (if that's required)

Is it an **infringement** of copyright when one uses copyrighted materials for **model training**?



High barrier to sharing training data (if that's required)

Is it an **infringement** of copyright when one uses copyrighted materials for **model training**?

Data containing **personally identifiable information** or **health care data** cannot be shared



**Fairly Trained
Certified**

<https://www.fairlytrained.org/certifications>

Current use restrictions

All **fail OSD 5 and 6** miserably:

- “You agree that the Contribution, or any derivative work of the Contribution, **will not be used ... to**” - *RAIL License*, <https://www.licenses.ai/source-code-license>
- “You **may not use** ... Gemma ... to ...” *Gemma Prohibited Use Policy*, https://ai.google.dev/gemma/prohibited_use_policy

What is the reason?

Current use restrictions

All **fail OSD 5 and 6** miserably:

- “You agree that the Contribution or any derivative work of the Contribution, **will not be used ... to**” - *RAI License*, <https://licenses.ai/source-code-license>
- “You **may not use**” - *OpenAI Prohibited Use Policy*, https://ai.google.dev/gemini/prohibited_use_policy

What is the reason?

Current use restrictions

All **fail OSD 5 and 6** miserably:

- “You agree that the Contribution, or any copy or derivative work of the Contribution, **will not be used ... to**” - *RAIL License*, <https://www.licenses.ai/source-code-license>
- “You **may not use** ...” - *Gamma Prohibited Use Policy*, https://ai.google.dev/gemini-prohibited_use_policy

What is the reason?

Liability?

Thanks!

Open Software in a Changing World - University of Skövde, April 16, 2024