Skip to main content

[![Cornell University](/static/browse/0.3.4/images/icons/cu/cornell-reduced-white-SMALL.svg)](https://www.cornell.edu/)

We gratefully acknowledge support from the Simons Foundation, [member institutions](https://info.arxiv.org/about/ourmembers.html), and all contributors. [Donate](https://info.arxiv.org/about/donate.html)

](/IgnoreMe)

[![arxiv logo](/static/browse/0.3.4/images/arxiv-logo-one-color-white.svg)](/)

> [cs](/list/cs/recent) > arXiv:2205.01068v1

[Help](https://info.arxiv.org/help) | [Advanced Search](https://arxiv.org/search/advanced)

All fields Title Author Abstract Comments Journal reference ACM classification MSC classification Report number arXiv identifier DOI ORCID arXiv author ID Help pages Full text

Search

[![arXiv logo](/static/browse/0.3.4/images/arxiv-logomark-small-white.svg)](https://arxiv.org/)

[![Cornell University Logo](/static/browse/0.3.4/images/icons/cu/cornell-reduced-white-SMALL.svg)](https://www.cornell.edu/)

open search

GO

open navigation menu

quick links

  * [Login](https://arxiv.org/login)
  * [Help Pages](https://info.arxiv.org/help)
  * [About](https://info.arxiv.org/about)

Computer Science > Computation and Language

arXiv:2205.01068v1 (cs)

[Submitted on 2 May 2022 (this version), _latest version 21 Jun 2022_

([v4](https://arxiv.org/abs/2205.01068v4))]

Title:OPT: Open Pre-trained Transformer Language Models

Authors:[Susan

Zhang](https://arxiv.org/search/cs?searchtype=author&query=Zhang,+S), [Stephen

Roller](https://arxiv.org/search/cs?searchtype=author&query=Roller,+S), [Naman

Goyal](https://arxiv.org/search/cs?searchtype=author&query=Goyal,+N), [Mikel

Artetxe](https://arxiv.org/search/cs?searchtype=author&query=Artetxe,+M),

[Moya Chen](https://arxiv.org/search/cs?searchtype=author&query=Chen,+M), [Shuohui Chen](https://arxiv.org/search/cs?searchtype=author&query=Chen,+S), [Christopher Dewan](https://arxiv.org/search/cs?searchtype=author&query=Dewan,+C), [Mona Diab](https://arxiv.org/search/cs?searchtype=author&query=Diab,+M), [Xian Li](https://arxiv.org/search/cs?searchtype=author&query=Li,+X), [Xi Victoria Lin](https://arxiv.org/search/cs?searchtype=author&query=Lin,+X+V), [Todor Mihaylov](https://arxiv.org/search/cs?searchtype=author&query=Mihaylov,+T), [Myle Ott](https://arxiv.org/search/cs?searchtype=author&query=Ott,+M), [Sam Shleifer](https://arxiv.org/search/cs?searchtype=author&query=Shleifer,+S), [Kurt Shuster](https://arxiv.org/search/cs?searchtype=author&query=Shuster,+K), [Daniel Simig](https://arxiv.org/search/cs?searchtype=author&query=Simig,+D), [Punit Singh Koura](https://arxiv.org/search/cs?searchtype=author&query=Koura,+P+S), [Anjali Sridhar](https://arxiv.org/search/cs?searchtype=author&query=Anjali), [Tianlu Wang](https://arxiv.org/search/cs?searchtype=author&query=Wang,+T), [Luke Zettlemoyer](https://arxiv.org/search/cs?searchtype=author&query=Zettlemoyer,+L)

View a PDF of the paper titled OPT: Open Pre-trained Transformer Language Models, by Susan Zhang and 18 other authors

[View PDF](/pdf/2205.01068v1)

> Abstract:Large language models, which are often trained for hundreds of
> thousands of compute days, have shown remarkable capabilities for zero- and

> few-shot learning. Given their computational cost, these models are

> difficult to replicate without significant capital. For the few that are

> available through APIs, no access is granted to the full model weights,

> making them difficult to study. We present Open Pre-trained Transformers

> (OPT), a suite of decoder-only pre-trained transformers ranging from 125M to

> 175B parameters, which we aim to fully and responsibly share with interested

> researchers. We show that OPT-175B is comparable to GPT-3, while requiring

> only 1/7th the carbon footprint to develop. We are also releasing our

> logbook detailing the infrastructure challenges we faced, along with code

> for experimenting with all of the released models.

Subjects: | Computation and Language (cs.CL); Machine Learning (cs.LG)

---|---

Cite as: | [arXiv:2205.01068](https://arxiv.org/abs/2205.01068) [cs.CL]

| (or [arXiv:2205.01068v1](https://arxiv.org/abs/2205.01068v1) [cs.CL] for this version)

| <https://doi.org/10.48550/arXiv.2205.01068> Focus to learn more arXiv-issued DOI via DataCite

Submission history

From: Susan Zhang [[view email](/show-email/a2d4df92/2205.01068)]

[v1] Mon, 2 May 2022 17:49:50 UTC (9,196 KB)

[[v2]](/abs/2205.01068v2) Tue, 3 May 2022 15:04:06 UTC (9,190 KB)

[[v3]](/abs/2205.01068v3) Thu, 5 May 2022 11:44:30 UTC (7,822 KB)

[[v4]](/abs/2205.01068v4) Tue, 21 Jun 2022 17:04:40 UTC (7,832 KB)

Full-text links:

Access Paper:

View a PDF of the paper titled OPT: Open Pre-trained Transformer Language Models, by Susan Zhang and 18 other authors

* [View PDF](/pdf/2205.01068v1)
* [Other Formats](/format/2205.01068v1)

[![license icon](https://arxiv.org/icons/licenses/by-4.0.png) view license](http://creativecommons.org/licenses/by/4.0/ "Rights to this article")

Current browse context:

cs.CL

[< prev](/prevnext?id=2205.01068&function=prev&context=cs.CL "previous in cs.CL \(accesskey p\)") | [next >](/prevnext?id=2205.01068&function=next&context=cs.CL "next in cs.CL \(accesskey n\)")

[new](/list/cs.CL/new) | [recent](/list/cs.CL/recent) | [2022-05](/list/cs.CL/2022-05)

Change to browse by:

[cs](/abs/2205.01068?context=cs)
[cs.LG](/abs/2205.01068?context=cs.LG)

References & Citations

* [NASA ADS](https://ui.adsabs.harvard.edu/abs/arXiv:2205.01068)
* [Google Scholar](https://scholar.google.com/scholar_lookup?arxiv_id=2205.01068)
* [Semantic Scholar](https://api.semanticscholar.org/arXiv:2205.01068)

[3 blog links](/tb/2205.01068)

([what is this?](https://info.arxiv.org/help/trackback.html))

[a](/static/browse/0.3.4/css/cite.css) export BibTeX citation Loading...

BibTeX formatted citation

×

loading...

Data provided by:

Bookmark

[![BibSonomy logo](/static/browse/0.3.4/images/icons/social/bibsonomy.png)
](http://www.bibsonomy.org/BibtexHandler?requTask=upload&url=https://arxiv.org/abs/2205.01068&
description=OPT:
Open Pre-trained Transformer Language Models "Bookmark on BibSonomy") [
![Reddit logo](/static/browse/0.3.4/images/icons/social/reddit.png)

](https://reddit.com/submit?url=https://arxiv.org/abs/2205.01068&title=OPT:

Open Pre-trained Transformer Language Models "Bookmark on Reddit")

Bibliographic Tools

Bibliographic and Citation Tools

Bibliographic Explorer Toggle

Bibliographic Explorer _([What is the

Explorer?](https://info.arxiv.org/labs/showcase.html#arxiv-bibliographic-

explorer))_

Connected Papers Toggle

Connected Papers _([What is Connected

Papers?](https://www.connectedpapers.com/about))_

Litmaps Toggle

Litmaps _([What is Litmaps?](https://www.litmaps.co/))_

scite.ai Toggle

scite Smart Citations _([What are Smart Citations?](https://www.scite.ai/))_

Code, Data, Media

Code, Data and Media Associated with this Article

alphaXiv Toggle

alphaXiv _([What is alphaXiv?](https://alphaxiv.org/))_

Links to Code Toggle

CatalyzeX Code Finder for Papers _([What is CatalyzeX?](https://www.catalyzex.com))_

DagsHub Toggle

DagsHub _([What is DagsHub?](https://dagshub.com/))_

GotitPub Toggle

Gotit.pub _([What is GotitPub?](http://gotit.pub/faq))_

Huggingface Toggle

Hugging Face _([What is Huggingface?](https://huggingface.co/huggingface))_

Links to Code Toggle

Papers with Code _([What is Papers with Code?](https://paperswithcode.com/))_

ScienceCast Toggle

ScienceCast _([What is ScienceCast?](https://sciencecast.org/welcome))_

Demos

Demos

Replicate Toggle

Replicate _([What is Replicate?](https://replicate.com/docs/arxiv/about))_

Spaces Toggle

Hugging Face Spaces _([What is

Spaces?](https://huggingface.co/docs/hub/spaces))_

Spaces Toggle

TXYZ.AI _([What is TXYZ.AI?](https://txyz.ai))_

Related Papers

Recommenders and Search Tools

Link to Influence Flower

Influence Flower _([What are Influence

Flowers?](https://influencemap.cmlab.dev/))_

Core recommender toggle

CORE Recommender _([What is CORE?](https://core.ac.uk/services/recommender))_

* Author
* Venue
* Institution
* Topic

About arXivLabs

arXivLabs: experimental projects with community collaborators

arXivLabs is a framework that allows collaborators to develop and share new arXiv features directly on our website.

Both individuals and organizations that work with arXivLabs have embraced and accepted our values of openness, community, excellence, and user data privacy. arXiv is committed to these values and only works with partners that adhere to them.

Have an idea for a project that will add value for arXiv's community? [**Learn more about arXivLabs**](https://info.arxiv.org/labs/index.html).

[Which authors of this paper are endorsers?](/auth/show-endorsers/2205.01068) | [Disable MathJax](javascript:setMathjaxCookie\(\)) ([What is MathJax?](https://info.arxiv.org/help/mathjax.html))

* [About](https://info.arxiv.org/about)
* [Help](https://info.arxiv.org/help)

* contact arXivClick here to contact arXiv [Contact](https://info.arxiv.org/help/contact.html)
* subscribe to arXiv mailingsClick here to subscribe [Subscribe](https://info.arxiv.org/help/subscribe)

* [Copyright](https://info.arxiv.org/help/license/index.html)
* [Privacy Policy](https://info.arxiv.org/help/policies/privacy_policy.html)

* [Web Accessibility Assistance](https://info.arxiv.org/help/web_accessibility.html)
* [arXiv Operational Status](https://status.arxiv.org)

Get status notifications via [email](https://subscribe.sorryapp.com/24846f03/email/new) or [slack](https://subscribe.sorryapp.com/24846f03/slack/new)