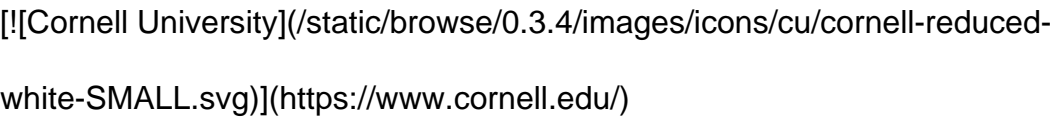


Skip to main content


(<https://www.cornell.edu/>)

In just 3 minutes help us improve arXiv:

[Annual Global Survey](https://cornell.ca1.qualtrics.com/jfe/form/SV_6m22mbqW9GQ3pQO)(https://cornell.ca1.qualtrics.com/jfe/form/SV_6m22mbqW9GQ3pQO)

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

[\[IgnoreMe\]](#)

([/static/browse/0.3.4/images/arxiv-logo-one-color-white.svg](#))([/](#))
> [\[cs\]](#)([/list/cs/recent](#)) > arXiv:2210.07316

[\[Help\]](https://info.arxiv.org/help)(<https://info.arxiv.org/help>) | [\[Advanced Search\]](https://arxiv.org/search/advanced)(<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:2210.07316**** (cs)

[Submitted on 13 Oct 2022 ([v1](https://arxiv.org/abs/2210.07316v1)), last revised 19 Mar 2023 (this version, v3)]

Title:MTEB: Massive Text Embedding Benchmark

Authors:[Niklas

Muennighoff](<https://arxiv.org/search/cs?searchtype=author&query=Muellnighoff,+N>),

[Nouamane Tazi](<https://arxiv.org/search/cs?searchtype=author&query=Tazi,+N>),

[Loïc Magne](<https://arxiv.org/search/cs?searchtype=author&query=Magne,+L>),

[Nils Reimers](<https://arxiv.org/search/cs?searchtype=author&query=Reimers,+N>)

View a PDF of the paper titled MTEB: Massive Text Embedding Benchmark, by

Niklas Muennighoff and 3 other authors

[View PDF](</pdf/2210.07316>)

> Abstract:Text embeddings are commonly evaluated on a small set of datasets
> from a single task not covering their possible applications to other tasks.
> It is unclear whether state-of-the-art embeddings on semantic textual
> similarity (STS) can be equally well applied to other tasks like clustering
> or reranking. This makes progress in the field difficult to track, as
> various models are constantly being proposed without proper evaluation. To
> solve this problem, we introduce the Massive Text Embedding Benchmark
> (MTEB). MTEB spans 8 embedding tasks covering a total of 58 datasets and 112
> languages. Through the benchmarking of 33 models on MTEB, we establish the
> most comprehensive benchmark of text embeddings to date. We find that no
> particular text embedding method dominates across all tasks. This suggests
> that the field has yet to converge on a universal text embedding method and
> scale it up sufficiently to provide state-of-the-art results on all
> embedding tasks. MTEB comes with open-source code and a public leaderboard
> at [this https URL](<https://github.com/embeddings-benchmark/mteb>).

Comments: | 24 pages, 14 tables, 6 figures

---|---

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

Cite as: | [arXiv:2210.07316](https://arxiv.org/abs/2210.07316) [cs.CL]

| (or [arXiv:2210.07316v3](https://arxiv.org/abs/2210.07316v3) [cs.CL] for this version)

| <https://doi.org/10.48550/arXiv.2210.07316> Focus to learn more arXiv-issued DOI via DataCite

Submission history

From: Niklas Muennighoff [[view email](/show-email/22b4e350/2210.07316)]

[[v1]](/abs/2210.07316v1) Thu, 13 Oct 2022 19:42:08 UTC (20,387 KB)

[[v2]](/abs/2210.07316v2) Sun, 5 Feb 2023 15:59:49 UTC (3,993 KB)

[v3] Sun, 19 Mar 2023 13:37:01 UTC (3,994 KB)

Full-text links:

Access Paper:

View a PDF of the paper titled MTEB: Massive Text Embedding Benchmark, by

Niklas Muennighoff and 3 other authors

* [View PDF](/pdf/2210.07316)

* [TeX Source](/src/2210.07316)

* [Other Formats](/format/2210.07316)

[view license](http://arxiv.org/licenses/nonexclusive-distrib/1.0/ "Rights to this article")

Current browse context:

cs.CL

[< prev](/prevnext?id=2210.07316&function=prev&context=cs.CL "previous in cs.CL \(\accesskey p\)") | [next >](/prevnext?id=2210.07316&function=next&context=cs.CL "next in cs.CL \(\accesskey n\)")

[new](/list/cs.CL/new) | [recent](/list/cs.CL/recent) | [2022-10](/list/cs.CL/2022-10)

Change to browse by:

[cs](/abs/2210.07316?context=cs)

[cs.IR](/abs/2210.07316?context=cs.IR)

[cs.LG](/abs/2210.07316?context=cs.LG)

References & Citations

- * [NASA ADS](https://ui.adsabs.harvard.edu/abs/arXiv:2210.07316)
- * [Google Scholar](https://scholar.google.com/scholar_lookup?arxiv_id=2210.07316)
- * [Semantic Scholar](https://api.semanticscholar.org/arXiv:2210.07316)

[1 blog link](/tb/2210.07316)

([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/2210.07316&description=MTEB:

Massive Text Embedding Benchmark "Bookmark on BibSonomy") [![Reddit logo](/static/browse/0.3.4/images/icons/social/reddit.png)

](https://reddit.com/submit?url=https://arxiv.org/abs/2210.07316&title=MTEB:

Massive Text Embedding Benchmark "Bookmark on Reddit")

Bibliographic Tools

Bibliographic and Citation Tools

Bibliographic Explorer Toggle

Bibliographic Explorer [_\(\[What is the Explorer?\]\(https://info.arxiv.org/labs/showcase.html#arxiv-bibliographic-explorer\)\)_](https://info.arxiv.org/labs/showcase.html#arxiv-bibliographic-explorer)

Connected Papers Toggle

Connected Papers [_\(\[What is Connected Papers?\]\(https://www.connectedpapers.com/about\)\)_](https://www.connectedpapers.com/about)

Litmaps Toggle

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

scite.ai Toggle

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

Code, Data, Media

Code, Data and Media Associated with this Article

alphaXiv Toggle

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

Spaces Toggle

Hugging Face Spaces [_\(\[What is Spaces?\]\(https://huggingface.co/docs/hub/spaces\)\)_](https://huggingface.co/docs/hub/spaces)

Spaces Toggle

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

Related Papers

Recommenders and Search Tools

Link to Influence Flower

Influence Flower [_\(\[What are Influence Flowers?\]\(https://influencemap.cmlab.dev/\)\)_](https://influencemap.cmlab.dev/)

Core recommender toggle

CORE Recommender [_\(\[What is CORE?\]\(https://core.ac.uk/services/recommender\)\)_](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? [\[**Learn more about arXivLabs**\]\(https://info.arxiv.org/labs/index.html\)](https://info.arxiv.org/labs/index.html).

[\[Which authors of this paper are endorsers?\]\(/auth/show-endorsers/2210.07316\)](#) | [\[Disable MathJax\]\(javascript:setMathjaxCookie\\(\\)\)](#) [\(\[What is MathJax?\]\(https://info.arxiv.org/help/mathjax.html\)\)](#)

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

* [contact arXiv](https://info.arxiv.org/help/contact.html)Click here to contact arXiv [[Contact](https://info.arxiv.org/help/contact.html)](https://info.arxiv.org/help/contact.html)

* [subscribe to arXiv mailings](https://info.arxiv.org/help/subscribe)Click here to subscribe [[Subscribe](https://info.arxiv.org/help/subscribe)](https://info.arxiv.org/help/subscribe)

* [\[Copyright\]](https://info.arxiv.org/help/license/index.html)(https://info.arxiv.org/help/license/index.html)

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

* [\[Web Accessibility Assistance\]](https://info.arxiv.org/help/web_accessibility.html)(https://info.arxiv.org/help/web_accessibility.html)

* [\[arXiv Operational Status \]](https://status.arxiv.org)(https://status.arxiv.org)

Get status notifications via

[\[email\]](https://subscribe.sorryapp.com/24846f03/email/new)(https://subscribe.sorryapp.com/24846f03/email/new) or

[\[slack\]](https://subscribe.sorryapp.com/24846f03/slack/new)(https://subscribe.sorryapp.com/24846f03/slack/new)