


Skip to main content

 (</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, [\[member institutions\]\(https://info.arxiv.org/about/ourmembers.html\)](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:2210.15191

[\[Help\]\(https://info.arxiv.org/help\)](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

 (</static/browse/0.3.4/images/arxiv-logomark-small-white.svg>) (<https://arxiv.org/>)

 (</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:2210.15191\*\* (cs)

[Submitted on 27 Oct 2022]

# Title:Truncation Sampling as Language Model Desmoothing

Authors:[John

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

[Christopher D.

Manning](https://arxiv.org/search/cs?searchtype=author&query=Manning,+C+D),

[Percy Liang](https://arxiv.org/search/cs?searchtype=author&query=Liang,+P)

View a PDF of the paper titled Truncation Sampling as Language Model

Desmoothing, by John Hewitt and 2 other authors

[View PDF](/pdf/2210.15191)

> Abstract: Long samples of text from neural language models can be of poor  
> quality. Truncation sampling algorithms--like top- $p$  or top- $k$ -- address  
> this by setting some words' probabilities to zero at each step. This work  
> provides framing for the aim of truncation, and an improved algorithm for  
> that aim. We propose thinking of a neural language model as a mixture of a  
> true distribution and a smoothing distribution that avoids infinite  
> perplexity. In this light, truncation algorithms aim to perform desmoothing,  
> estimating a subset of the support of the true distribution. Finding a good  
> subset is crucial: we show that top- $p$  unnecessarily truncates high-  
> probability words, for example causing it to truncate all words but Trump  
> for a document that starts with Donald. We introduce  $\eta$ -sampling, which  
> truncates words below an entropy-dependent probability threshold. Compared  
> to previous algorithms,  $\eta$ -sampling generates more plausible long  
> English documents according to humans, is better at breaking out of  
> repetition, and behaves more reasonably on a battery of test distributions.

Comments: | Findings of EMNLP, + small fixes

---|---

Subjects: | Computation and Language (cs.CL)

Cite as: | [arXiv:2210.15191](https://arxiv.org/abs/2210.15191) [cs.CL]

| (or [arXiv:2210.15191v1](https://arxiv.org/abs/2210.15191v1) [cs.CL] for this version)

| <https://doi.org/10.48550/arXiv.2210.15191> Focus to learn more arXiv-issued DOI via DataCite

## ## Submission history

From: John Hewitt [\[view email\]](/show-email/61939dc7/2210.15191/)

**[v1]** Thu, 27 Oct 2022 05:52:35 UTC (800 KB)

## Full-text links:

## ## Access Paper:

View a PDF of the paper titled Truncation Sampling as Language Model

Desmoothing, by John Hewitt and 2 other authors

\* [\[View PDF\]](/pdf/2210.15191/)

\* [\[TeX Source\]](/src/2210.15191/)

\* [\[Other Formats\]](/format/2210.15191/)

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

## Current browse context:

cs.CL

[\[< prev\]](/prevnext?id=2210.15191&function=prev&context=cs.CL) "previous in cs.CL [\\(\accesskey p\\)](#)" | [\[next >\]](/prevnext?id=2210.15191&function=next&context=cs.CL) "next in cs.CL [\\(\accesskey p\\)](#)"

n\)"

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

Change to browse by:

[cs](/abs/2210.15191?context=cs)

### ### References & Citations

- \* [NASA ADS](https://ui.adsabs.harvard.edu/abs/arXiv:2210.15191)
- \* [Google Scholar](https://scholar.google.com/scholar\_lookup?arxiv\_id=2210.15191)
- \* [Semantic Scholar](https://api.semanticscholar.org/arXiv:2210.15191)

[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/2210.15191&description=Truncation

Sampling as Language Model Desmoothing "Bookmark on BibSonomy") [ ![Reddit logo](/static/browse/0.3.4/images/icons/social/reddit.png)

](https://reddit.com/submit?url=https://arxiv.org/abs/2210.15191&title=Truncation Sampling as Language Model Desmoothing "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 [\\_\(\[What are Smart Citations?\]\(https://www.scite.ai/\)\)\\_](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 [\\_\(\[What is CatalyzeX?\]\(https://www.catalyzex.com\)\)\\_](https://www.catalyzex.com/)

DagsHub Toggle

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

GotitPub Toggle

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

Huggingface Toggle

Hugging Face [\\_\(\[What is Huggingface?\]\(https://huggingface.co/huggingface\)\)\\_](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 [\\_\(\[What are Influence Flowers?\]\)\(https://influencemap.cmlab.dev/\)\)\\_](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).

[\[Which authors of this paper are endorsers?\]\(/auth/show-endorsers/2210.15191\)](/auth/show-endorsers/2210.15191) | [\[Disable MathJax\]\(javascript:setMathjaxCookie\\(\\)\)](#) [\(\[What is MathJax?\]\(https://info.arxiv.org/help/mathjax.html\)\)](#)

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

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

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

\* [subscribe to arXiv mailingsClick 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)