

[Skip to main content](#)> [cs](#) > arXiv:2009.01325 

## quick links

- [Login](#)
- [Help Pages](#)
- [About](#)

Computer Science &gt; Computation and Language

arXiv:2009.01325 (cs)

*[Submitted on 2 Sep 2020 ([v1](#)), last revised 15 Feb 2022 (this version, v3)]*

# Learning to summarize from human feedback

[Nisan Stiennon](#), [Long Ouyang](#), [Jeff Wu](#), [Daniel M. Ziegler](#), [Ryan Lowe](#), [Chelsea Voss](#), [Alec Radford](#), [Dario Amodei](#), [Paul Christiano](#)

[Download PDF](#)


As language models become more powerful, training and evaluation are increasingly bottlenecked by the data and metrics used for a particular task. For example, summarization models are often trained to predict human reference summaries and evaluated using ROUGE, but both of these metrics are rough proxies for what we really care about -- summary quality. In this work, we show that it is possible to significantly improve summary quality by training a model to optimize for human preferences. We collect a large, high-quality dataset of human comparisons between summaries, train a model to predict the human-preferred summary, and use that model as a reward function to fine-tune a summarization policy using reinforcement learning. We apply our method to a version of the TL;DR dataset of Reddit posts and find that our models significantly outperform both human reference summaries and much larger models fine-tuned with supervised learning alone. Our models also transfer to CNN/DM news articles, producing summaries nearly as good as the human reference without any news-specific fine-tuning. We conduct extensive analyses to understand our human feedback dataset and fine-tuned models. We establish that our reward model generalizes to new datasets, and that optimizing our reward model results in better summaries than optimizing ROUGE according to humans. We hope the evidence from our paper

motivates machine learning researchers to pay closer attention to how their training loss affects the model behavior they actually want.

Comments: NeurIPS 2020

Subjects: **Computation and Language (cs.CL)**; Artificial Intelligence (cs.AI); Machine Learning (cs.LG)

Cite as: [arXiv:2009.01325](https://arxiv.org/abs/2009.01325) [cs.CL]  
(or [arXiv:2009.01325v3](https://arxiv.org/abs/2009.01325v3) [cs.CL] for this version)  
<https://doi.org/10.48550/arXiv.2009.01325>

 Focus to learn more

arXiv-issued DOI via DataCite

## Submission history

From: Ryan Lowe T. [[view email](#)]

[\[v1\]](#) Wed, 2 Sep 2020 19:54:41 UTC (7,777 KB)

[\[v2\]](#) Tue, 27 Oct 2020 22:19:53 UTC (7,781 KB)

[\[v3\]](#) Tue, 15 Feb 2022 19:09:36 UTC (3,889 KB)

☐ Bibliographic Tools

## Bibliographic and Citation Tools

☐ Bibliographic Explorer Toggle

Bibliographic Explorer ([What is the Explorer?](#))

☐ Litmaps Toggle

Litmaps ([What is Litmaps?](#))

☐ scite.ai Toggle

scite Smart Citations ([What are Smart Citations?](#))

☒ Code, Data, Media

## Code, Data and Media Associated with this Article

☐ Links to Code Toggle

CatalyzeX Code Finder for Papers ([What is CatalyzeX?](#))

☐ DagsHub Toggle

DagsHub ([What is DagsHub?](#))

☐ Links to Code Toggle

Papers with Code ([What is Papers with Code?](#))

☐ ScienceCast Toggle

ScienceCast ([What is ScienceCast?](#))

☐ Demos

## Demos

☐ Replicate Toggle

Replicate ([What is Replicate?](#))

☐ Spaces Toggle

Hugging Face Spaces ([What is Spaces?](#))

☐ Related Papers

# Recommenders and Search Tools

☐ Link to Influence Flower

Influence Flower ([What are Influence Flowers?](#))

☐ Connected Papers Toggle

Connected Papers ([What is Connected Papers?](#))

☐ Core recommender toggle

CORE Recommender ([What is CORE?](#))

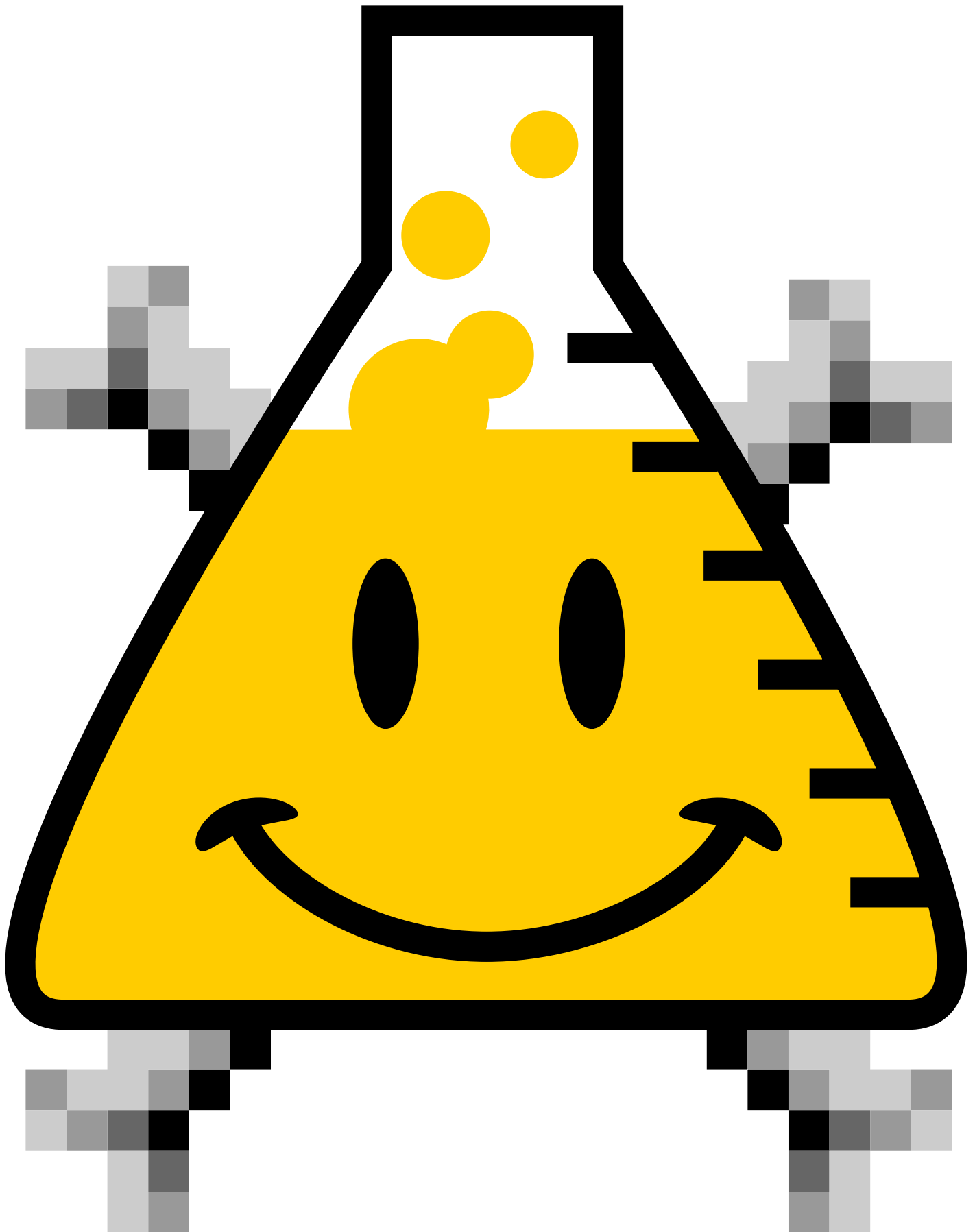
☐ About arXivLabs

## 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](#).



[Which authors of this paper are endorsers?](#) | [Disable MathJax](#) ([What is MathJax?](#))