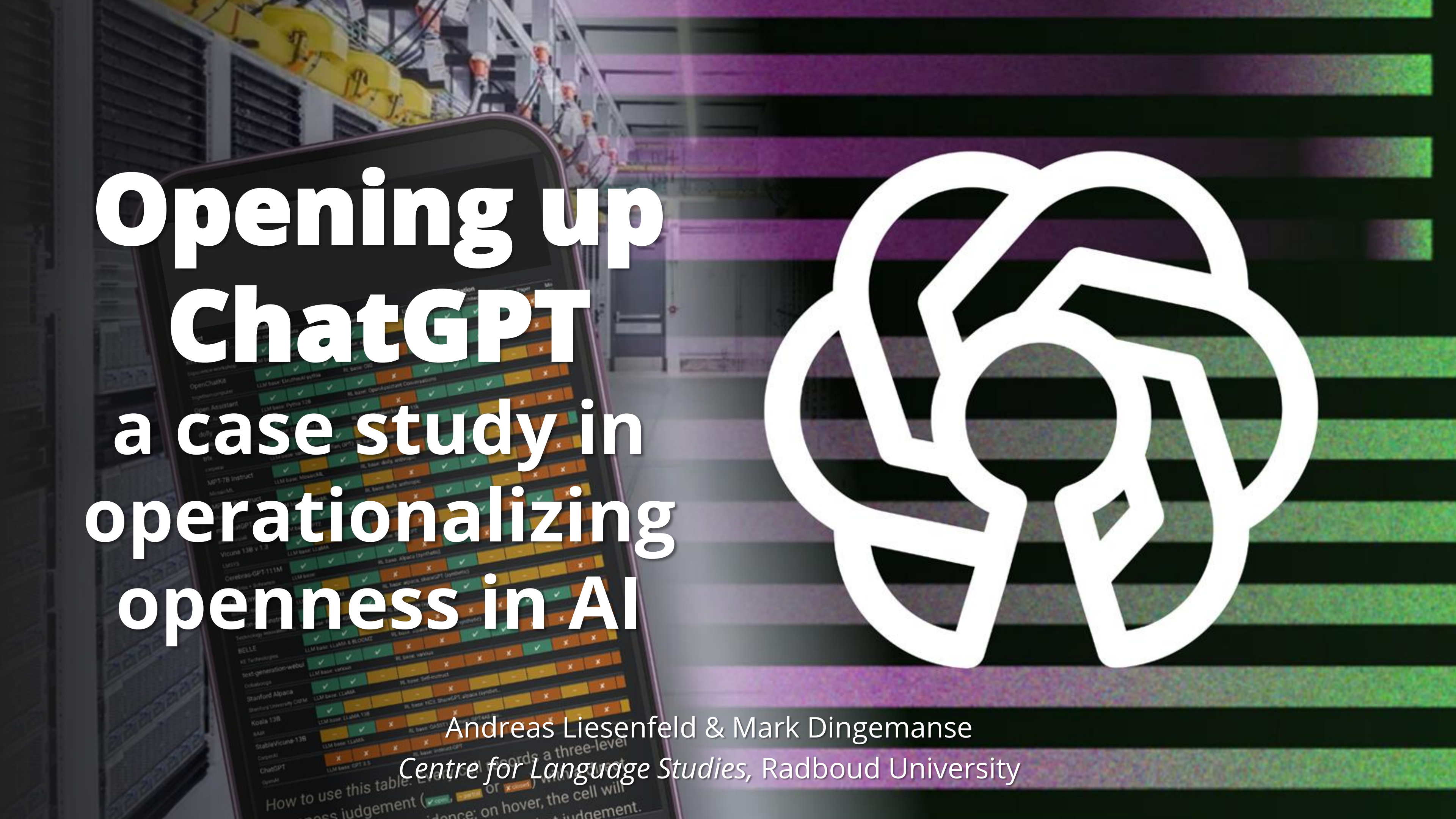


Opening up ChatGPT

a case study in operationalizing openness in AI

Andreas Liesenfeld & Mark Dingemanse

Centre for Language Studies, Radboud University

Germany

University of Rostock

“Open source will be a cornerstone of Germany’s digital state.”

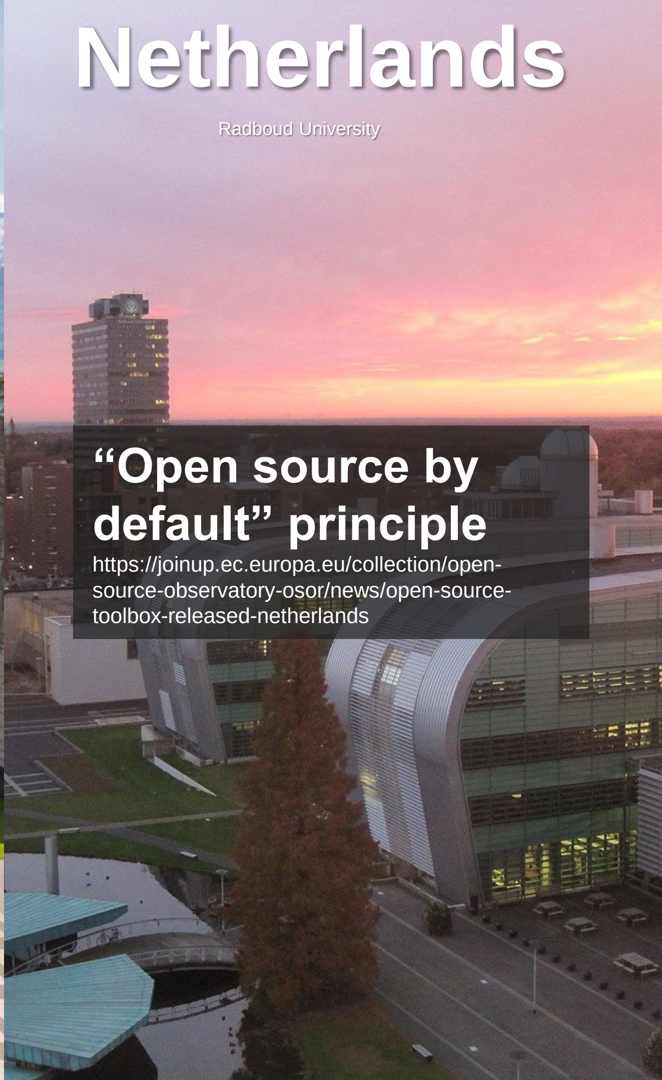
https://joinup.ec.europa.eu/collection/open-source-observatory-osor/news/open-source-be-norm-german-public-procurement

France

Grenoble Alpes University

“Open source as a critical component of scientific research”

https://joinup.ec.europa.eu/collection/open-source-observatory-osor/news/open-source-software-supported-french-open-science-policy



Netherlands

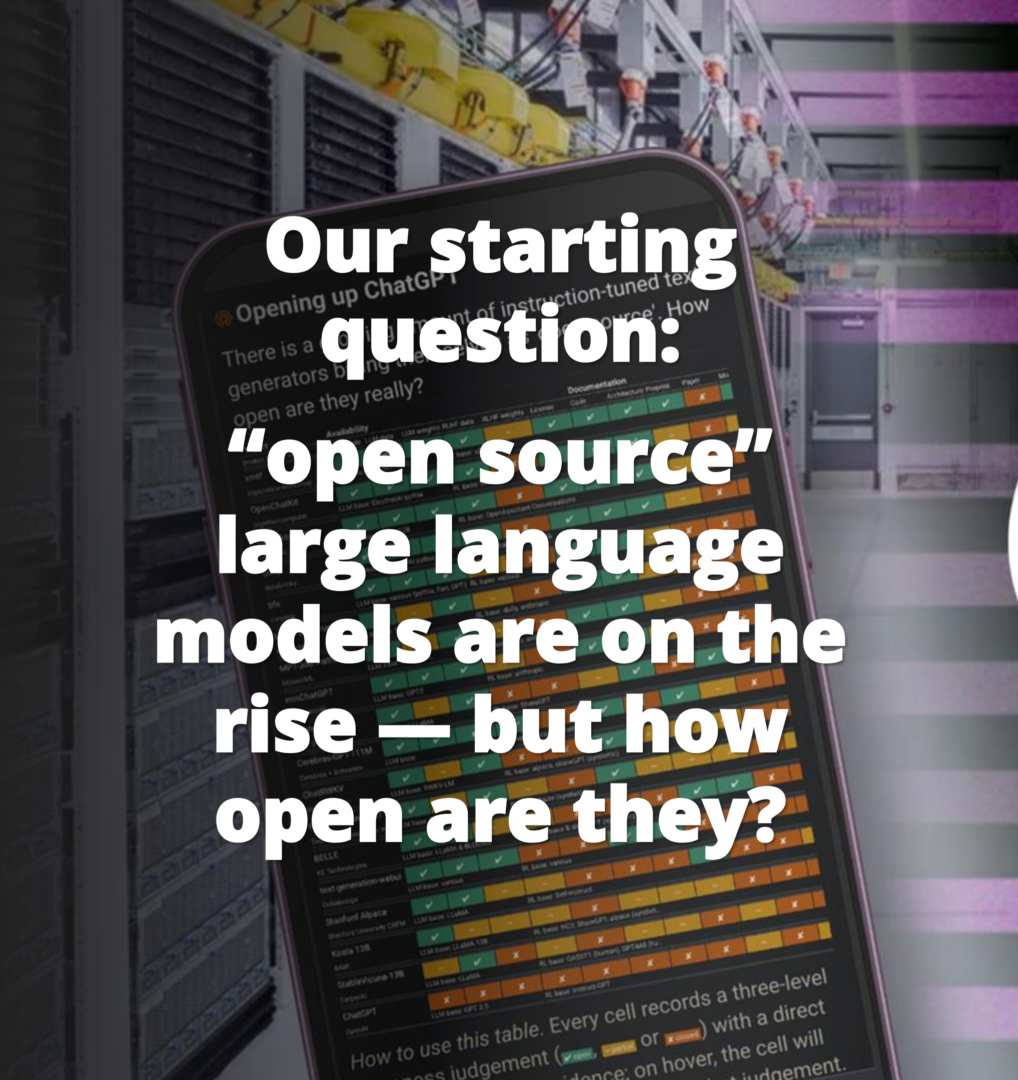
Radboud University

“Open source by default” principle

https://joinup.ec.europa.eu/collection/open-source-observatory-osor/news/open-source-toolbox-released-netherlands

Our starting question:

“open source” large language models are on the rise — but how open are they?



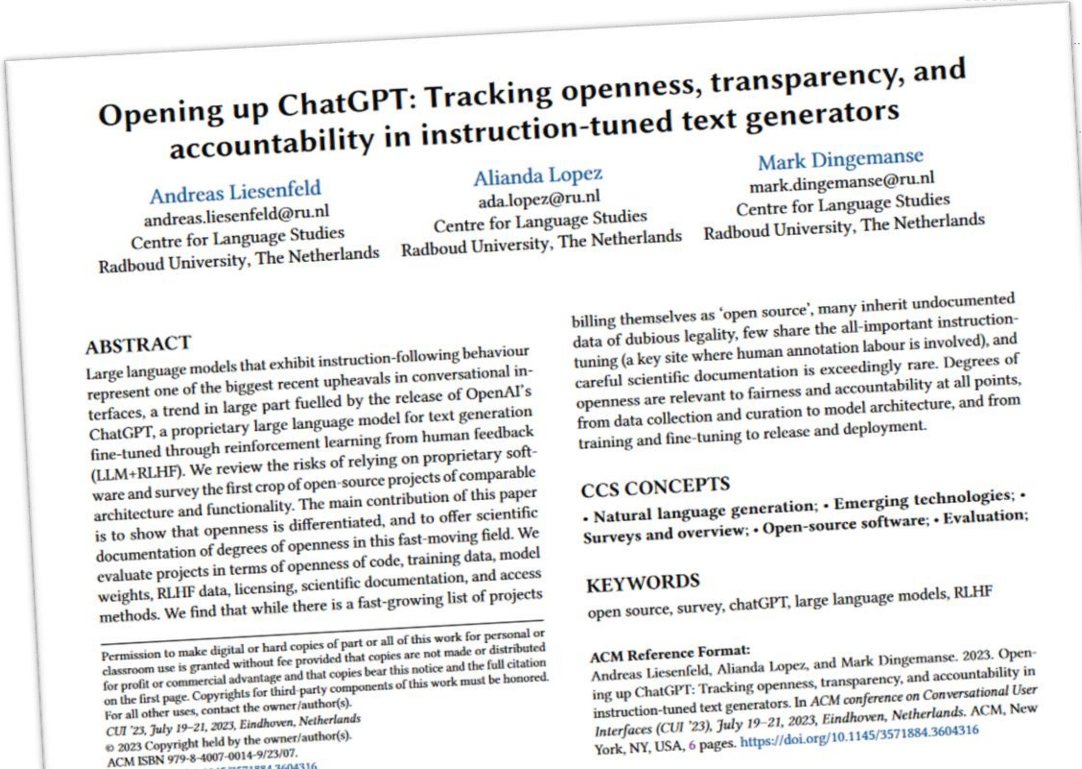
Opening up ChatGPT

There is a [illegible] of instruction-tuned text generators [illegible] 'open source'. How open are they really?

How to use this table. Every cell records a three-level [illegible] judgement (✓ open, ~ partial or ✗ closed) with a direct [illegible] on hover, the cell will [illegible] judgement.

I. Peer-reviewed paper

Liesenfeld, Lopez & Dingemanse (2023) *ACM Conference on Conversational User Interfaces (CUI '23)*. Eindhoven. doi: 10.1145/3571884.3604316



Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators

Andreas Liesenfeld
andreas.liesenfeld@ru.nl
Centre for Language Studies
Radboud University, The Netherlands

Alianda Lopez
ada.lopez@ru.nl
Centre for Language Studies
Radboud University, The Netherlands

Mark Dingemanse
mark.dingemanse@ru.nl
Centre for Language Studies
Radboud University, The Netherlands

ABSTRACT

Large language models that exhibit instruction-following behaviour represent one of the biggest recent upheavals in conversational interfaces, a trend in large part fuelled by the release of OpenAI's ChatGPT, a proprietary large language model for text generation fine-tuned through reinforcement learning from human feedback (LLM+RLHF). We review the risks of relying on proprietary software and survey the first crop of open-source projects of comparable architecture and functionality. The main contribution of this paper is to show that openness is differentiated, and to offer scientific documentation of degrees of openness in this fast-moving field. We evaluate projects in terms of openness of code, training data, model weights, RLHF data, licensing, scientific documentation, and access methods. We find that while there is a fast-growing list of projects billing themselves as 'open source', many inherit undocumented data of dubious legality, few share the all-important instruction-tuning (a key site where human annotation labour is involved), and careful scientific documentation is exceedingly rare. Degrees of openness are relevant to fairness and accountability at all points, from data collection and curation to model architecture, and from training and fine-tuning to release and deployment.

CCS CONCEPTS

• **Natural language generation;** • **Emerging technologies;** • **Surveys and overview;** • **Open-source software;** • **Evaluation;**

KEYWORDS

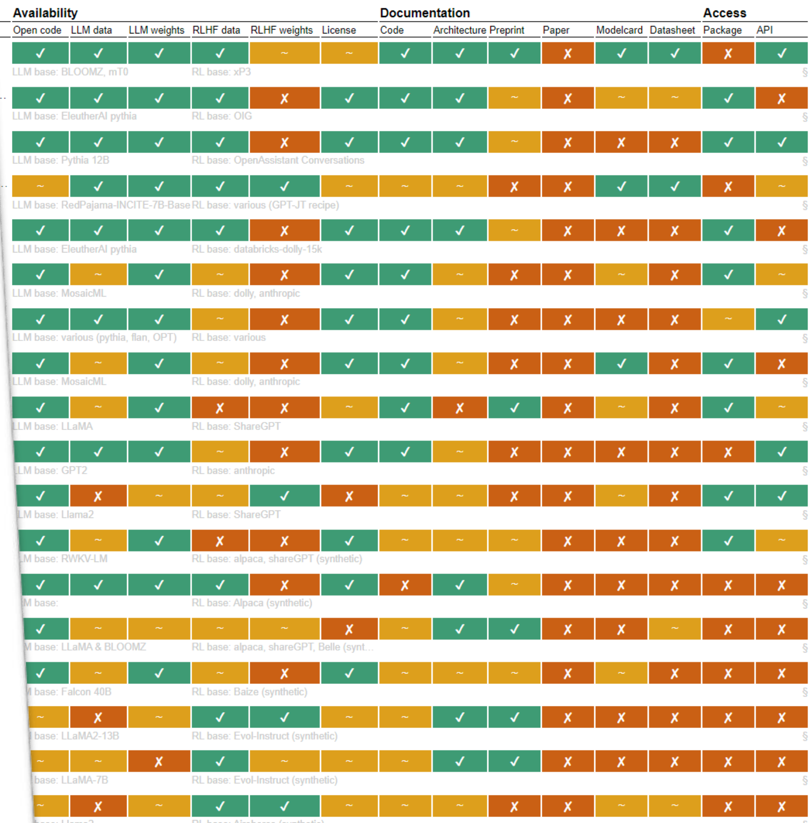
open source, survey, chatGPT, large language models, RLHF

ACM Reference Format:
Andreas Liesenfeld, Alianda Lopez, and Mark Dingemanse. 2023. Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. In *ACM conference on Conversational User Interfaces (CUI '23), July 19–21, 2023, Eindhoven, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3571884.3604316

CUI '23, July 19–21, 2023, Eindhoven, Netherlands
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0014-9/23/07.

II. Crowd-sourced live tracker

opening-up-chatgpt.io



Project	Availability						Documentation						Access	
(maker, bases, URL)	Open code	LLM data	LLM weights	RLHF data	RLHF weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
BLOOMZ LLM base: BLOOMZ, mT0 RL base: xP3 §	✓	✓	✓	✓	~	~	✓	✓	✓	✗	✓	✓	✗	✓
LLM base: EleutherAI pythia RL base: OIG §	✓	✓	✓	✓	✗	✓	✓	✓	~	✗	~	~	✓	✗
LLM base: Pythia 12B RL base: OpenAssistant Conversations §	✓	✓	✓	✓	✗	✓	✓	✓	~	✗	✗	✗	✓	✓
LLM base: RedPajama-INCITE-7B-Base RL base: various (GPT-JT recipe) §	~	✓	✓	✓	✓	~	~	~	✗	✗	✓	✓	✗	~
LLM base: EleutherAI pythia RL base: databricks-dolly-15k §	✓	✓	✓	✓	✗	✓	✓	✓	~	✗	✗	✗	✓	✗
LLM base: MosaicML RL base: dolly, anthropic §	✓	~	✓	~	✗	✓	✓	~	✗	✗	~	✗	✓	~
LLM base: various (pythia, flan, OPT) RL base: various §	✓	✓	✓	~	✗	✓	✓	~	✗	✗	✗	✗	~	✓
LLM base: MosaicML RL base: dolly, anthropic §	✓	~	✓	~	✗	✓	✓	~	✗	✗	✓	✗	✓	✗
LLM base: LLaMA RL base: ShareGPT §	✓	~	✓	✗	✗	~	✓	✗	✓	✗	~	✗	✓	~
LM base: GPT2 RL base: anthropic §	✓	✓	✓	~	✗	✓	✓	~	✗	✗	✗	✗	✗	✓
LM base: Llama2 RL base: ShareGPT §	✓	✗	~	~	✓	✗	~	~	✗	✗	~	✗	✓	✓
M base: RWKV-LM RL base: alpaca, shareGPT (synthetic) §	✓	~	✓	✗	✗	✓	~	~	~	✗	✗	✗	✓	~
M base: RL base: Alpaca (synthetic) §	✓	✓	✓	✓	✗	✓	✗	✓	~	✗	✗	✗	✗	✗
M base: LLaMA & BLOOMZ RL base: alpaca, shareGPT, Belle (synt… §	✓	~	~	~	~	✗	~	✓	✓	✗	✗	~	✗	✗
base: Falcon 40B RL base: Baize (synthetic) §	✓	~	✓	~	✗	✓	~	~	~	✗	~	✗	✗	✗
base: LLaMA2-13B RL base: Evol-Instruct (synthetic) §	~	✗	~	✓	✓	~	~	✓	✓	✗	✗	✗	✗	✗
base: LLaMA-7B RL base: Evol-Instruct (synthetic) §	~	~	✗	✓	~	~	~	✓	✓	✗	✗	✗	✗	✗
	~	✗	~	✓	✓	~	~	~	✗	✗	~	~	✗	✗

Surveying “openness” in ChatGPT-like text generators

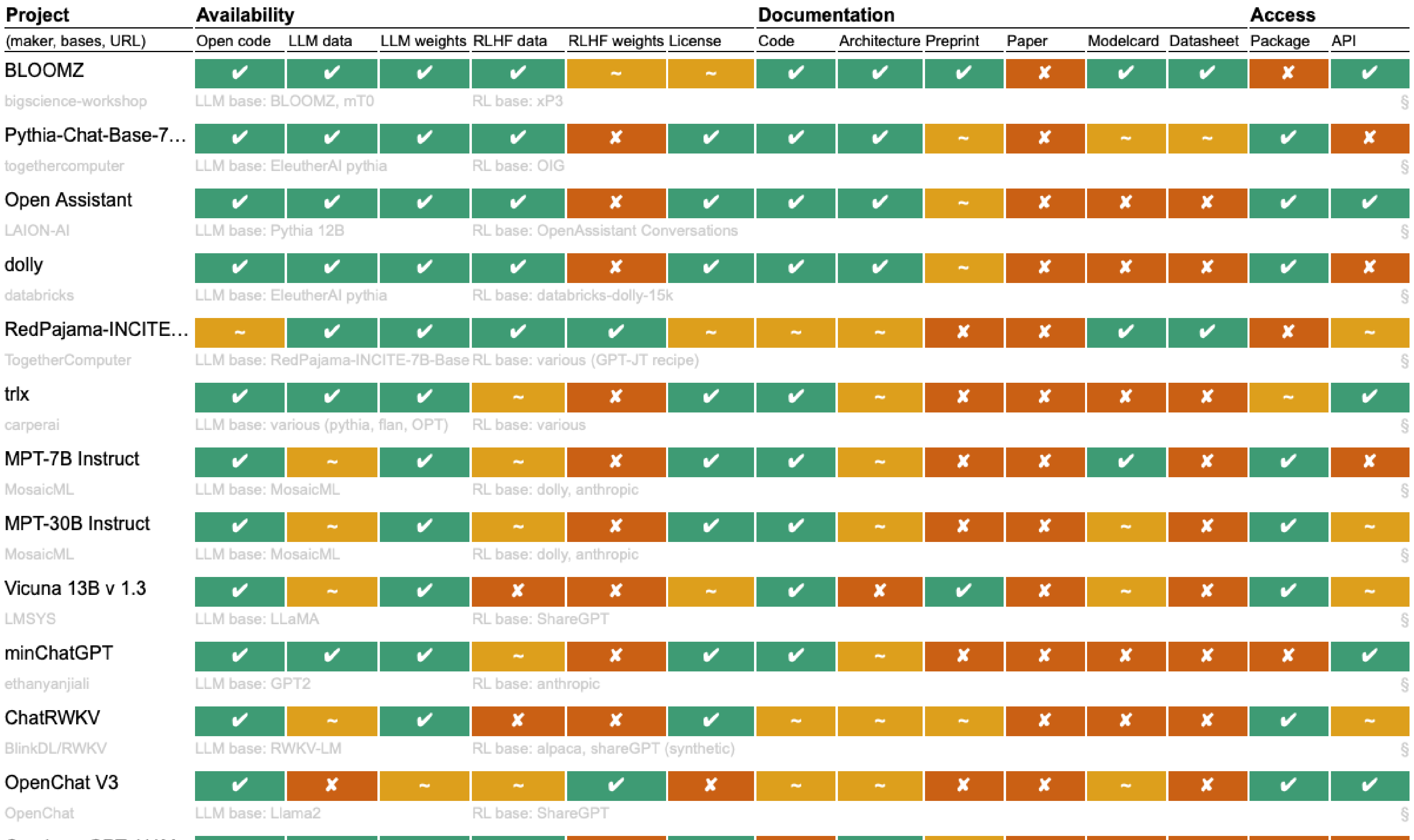
- in complex AI systems, openness is never all-or-nothing
- our approach: decompose into relevant *constituent elements*

Availability	Documentation	User access

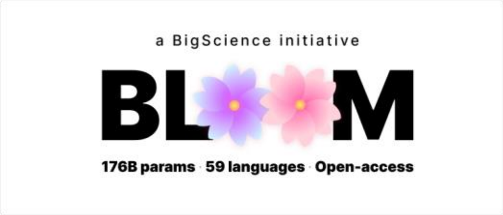
Surveying “openness” in ChatGPT-like text generators

- in complex AI systems, openness is never all-or-nothing
- our approach: decompose into relevant *constituent elements*
- for each element, record *degree of openness*

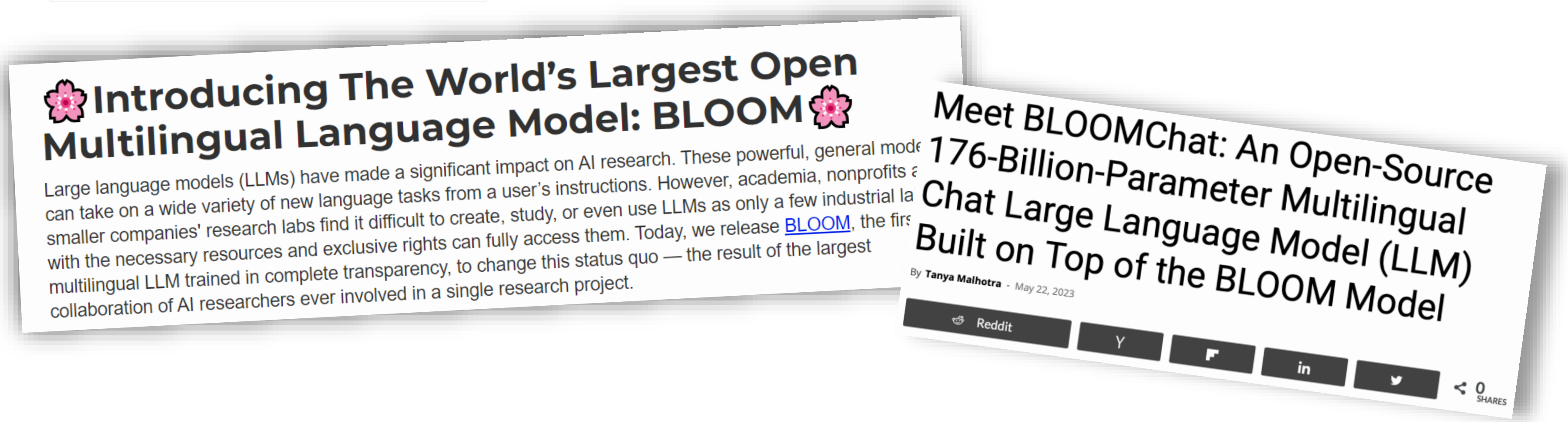
Availability	Documentation	User access
Open code	Code	Package
Base model data	Architecture	API
Base model weights	Preprint	
RLHF data	Paper	
RLHF weights	Model card	
License	Data sheet	



Project	Availability						Documentation						Access	
(maker, bases, URL)	Open code	LLM data	LLM weights	RLHF data	RLHF weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
BLOOMZ	✔	✔	✔	✔	~	~	✔	✔	✔	✘	✔	✔	✘	✔
bigscience-workshop	LLM base: BLOOMZ, mT0			RL base: xP3										§
Pythia-Chat-Base-7…	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	~	~	✔	✘
togethercomputer	LLM base: EleutherAI pythia			RL base: OIG										§
Open Assistant	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	✘	✘	✔	✔
LAION-AI	LLM base: Pythia 12B			RL base: OpenAssistant Conversations										§
dolly	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	✘	✘	✔	✘
databricks	LLM base: EleutherAI pythia			RL base: databricks-dolly-15k										§
RedPajama-INCITE…	~	✔	✔	✔	✔	~	~	~	✘	✘	✔	✔	✘	~
TogetherComputer	LLM base: RedPajama-INCITE-7B-Base			RL base: various (GPT-JT recipe)										§
trlx	✔	✔	✔	~	✘	✔	✔	~	✘	✘	✘	✘	~	✔
carperai	LLM base: various (pythia, flan, OPT)			RL base: various										§
MPT-7B Instruct	✔	~	✔	~	✘	✔	✔	~	✘	✘	✔	✘	✔	✘
MosaicML	LLM base: MosaicML			RL base: dolly, anthropic										§
MPT-30B Instruct	✔	~	✔	~	✘	✔	✔	~	✘	✘	~	✘	✔	~
MosaicML	LLM base: MosaicML			RL base: dolly, anthropic										§
Vicuna 13B v 1.3	✔	~	✔	✘	✘	~	✔	✘	✔	✘	~	✘	✔	~
LMSYS	LLM base: LLaMA			RL base: ShareGPT										§
minChatGPT	✔	✔	✔	~	✘	✔	✔	~	✘	✘	✘	✘	✘	✔
ethanyanjiali	LLM base: GPT2			RL base: anthropic										§
ChatRWKV	✔	~	✔	✘	✘	✔	~	~	~	✘	✘	✘	✔	~
BlinkDL/RWKV	LLM base: RWKV-LM			RL base: alpaca, shareGPT (synthetic)										§
OpenChat V3	✔	✘	~	~	✔	✘	~	~	✘	✘	~	✘	✔	✔
OpenChat	LLM base: Llama2			RL base: ShareGPT										§



Bloom(z) by BigScience Workshop



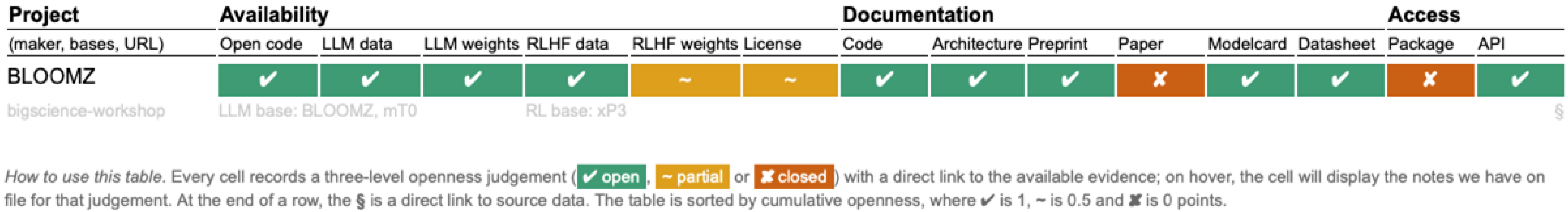
🌸Introducing The World's Largest Open Multilingual Language Model: BLOOM🌸

Large language models (LLMs) have made a significant impact on AI research. These powerful, general mode can take on a wide variety of new language tasks from a user's instructions. However, academia, nonprofits a smaller companies' research labs find it difficult to create, study, or even use LLMs as only a few industrial la with the necessary resources and exclusive rights can fully access them. Today, we release BLOOM, the firs multilingual LLM trained in complete transparency, to change this status quo — the result of the largest collaboration of AI researchers ever involved in a single research project.

Meet BLOOMChat: An Open-Source 176-Billion-Parameter Multilingual Chat Large Language Model (LLM) Built on Top of the BLOOM Model

By Tanya Malhotra - May 22, 2023

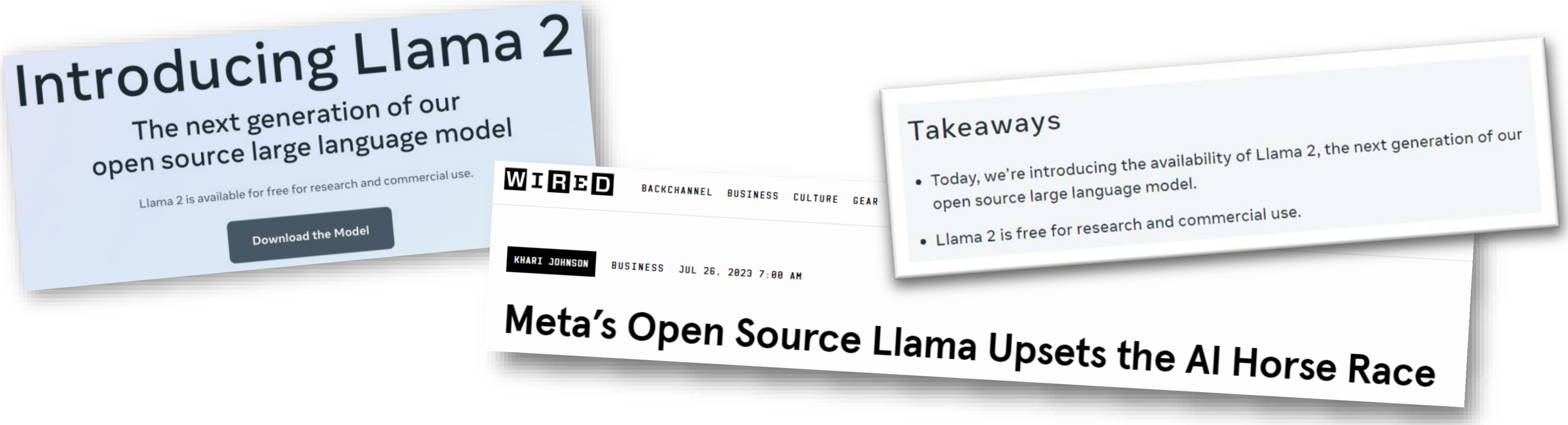
Reddit
Y
in
0
SHARES



Project	**Availability**						**Documentation**						**Access**	
(maker, bases, URL)	Open code	LLM data	LLM weights	RLHF data	RLHF weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
BLOOMZ	✔	✔	✔	✔	~	~	✔	✔	✔	✘	✔	✔	✘	✔
bigscience-workshop	LLM base: BLOOMZ, mT0			RL base: xP3										§

How to use this table. Every cell records a three-level openness judgement (✔ open, ~ partial or ✘ closed) with a direct link to the available evidence; on hover, the cell will display the notes we have on file for that judgement. At the end of a row, the § is a direct link to source data. The table is sorted by cumulative openness, where ✔ is 1, ~ is 0.5 and ✘ is 0 points.

Llama2 by Meta Platforms, Inc.



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

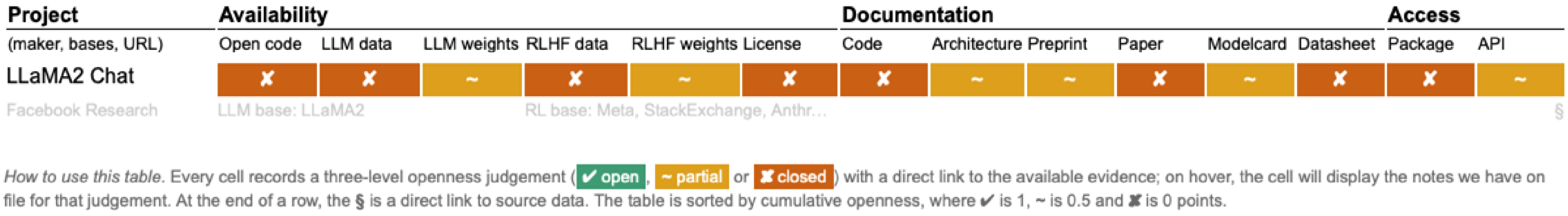
WIRED BACKCHANNEL BUSINESS CULTURE GEAR

KHARI JOHNSON BUSINESS JUL 26, 2023 7:00 AM

Meta's Open Source Llama Upsets the AI Horse Race

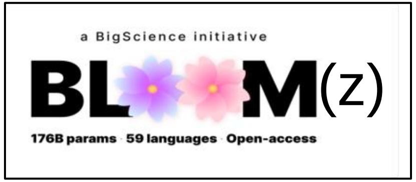
Takeaways

- Today, we're introducing the availability of Llama 2, the next generation of our open source large language model.
- Llama 2 is free for research and commercial use.



Project	**Availability**						**Documentation**						**Access**	
(maker, bases, URL)	Open code	LLM data	LLM weights	RLHF data	RLHF weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
LLaMA2 Chat	✘	✘	~	✘	~	✘	✘	~	~	✘	~	✘	✘	~
Facebook Research	LLM base: LLaMA2			RL base: Meta, StackExchange, Anthr...										§

How to use this table. Every cell records a three-level openness judgement (✓ open, ~ partial or ✘ closed) with a direct link to the available evidence; on hover, the cell will display the notes we have on file for that judgement. At the end of a row, the § is a direct link to source data. The table is sorted by cumulative openness, where ✓ is 1, ~ is 0.5 and ✘ is 0 points.

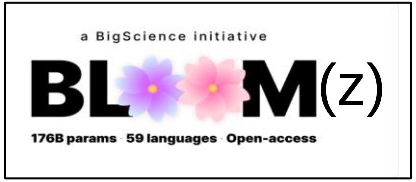


Fully shared code to enable reproduction

Open code
LLM data
LLM weights
RLHF data
RLHF weights
License
Code
Architecture
Preprint
Paper
Modelcard
Datasheet
Package
API

No training code available

Is the source code openly available?

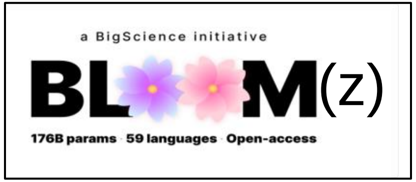


Fully shared code to enable reproduction
Training data shared

Open code
LLM data
LLM weights
RLHF data
RLHF weights
License
Code
Architecture
Preprint
Paper
Modelcard
Datasheet
Package
API

No training code available
No training data available

Is the pretraining dataset documented and available?

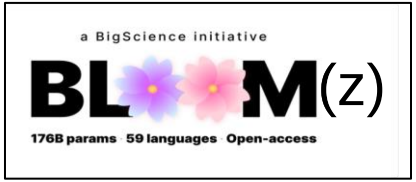


Fully shared code to enable reproduction
Training data shared
Access to base LLM without instruction tuning

Open code
LLM data
LLM weights
RLHF data
RLHF weights
License
Code
Architecture
Preprint
Paper
Modelcard
Datasheet
Package
API

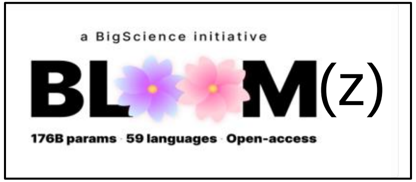
No training code available
No training data available
Accessible after registration

Are the model weights openly available?



Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	**RLHF data**	No data
	RLHF weights	
	License	
	Code	
	Architecture	
	Preprint	
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Are the instruction-tuning datasets documented and available?



Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	**RLHF weights**	Accessible after registration
	License	
	Code	
	Architecture	
	Preprint	
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Are the instruction-tuned model weights made available?

a BigScience initiative

BLOOM(z)

176B params · 59 languages · Open-access

Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	**License**	“Community license”, non OSI
	Code	
	Architecture	
	Preprint	
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Is the system released under an open license?

Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: "RAIL", non OSI	**License**	"Community license", non OSI
	Code	
	Architecture	
	Preprint	
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Is "unlimited" always best?

- Responsible AI License (RAIL) aims to address the moral dilemmas of harmful and unintended uses of tech (Contractor et al. 2022 *FAccT*)
- Restricts particular use cases (e.g. "don't use to exploit vulnerabilities of a specific group")

Responsibility: two approaches

- Llama2: you may not "represent that Llama 2 outputs are human-generated" (a low bar)
- RAIL: you may not "generate content without expressly and intelligibly disclaiming that the text is machine generated"

Is the system released under an open license?

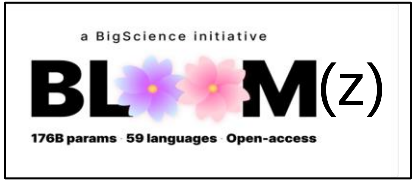
a BigScience initiative

BLOOM(z)

176B params · 59 languages · Open-access

Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	**Code**	Only minimal examples
	Architecture	
	Preprint	
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Is the codebase well-maintained and documented?



Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	**Architecture**	Sketched in preprint
	Preprint	
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Is the system architecture clearly documented?

a BigScience initiative BLOOM(z) 176B params · 59 languages · Open-access		Llama-2-7B Llama-2-13B Llama-2-70B
Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	Architecture	Sketched in preprint
Multiple detailed preprints	**Preprint**	Corporate preprint only
	Paper	
	Modelcard	
	Datasheet	
	Package	
	API	

Is there a preprint providing scientific documentation of the system?

a BigScience initiative

BLOOM(z)

176B params · 59 languages · Open-access

Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	Architecture	Sketched in preprint
Multiple detailed preprints	Preprint	Corporate preprint only
No paper	**Paper**	No paper
	Modelcard	
	Datasheet	
	Package	
	API	

Has the system been scrutinized under rigorous peer-review?

a BigScience initiative

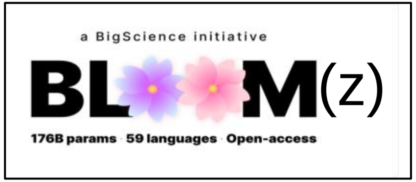
BLOOM(z)

176B params · 59 languages · Open-access

Llama-2-7B
Llama-2-13B
Llama-2-70B

Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	Architecture	Sketched in preprint
Multiple detailed preprints	Preprint	Corporate preprint only
No paper	Paper	No paper
Available	**Modelcard**	Only minimal detail provided
	Datasheet	
	Package	
	API	

Is the model described in a model card? *(Mitchell et al. 2019)*



Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	Architecture	Sketched in preprint
Multiple detailed preprints	Preprint	Corporate preprint only
No paper	Paper	No paper
Available	Modelcard	Only minimal detail provided
Available	**Datasheet**	No datasheet
	Package	
	API	

Is there a data sheet documenting data collection & curation? *(McMillan Major et al. 2023)*

a BigScience initiative

BLOOM(z)

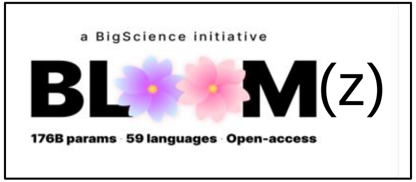
176B params 59 languages Open-access

Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	Architecture	Sketched in preprint
Multiple detailed preprints	Preprint	Corporate preprint only
No paper	Paper	No paper
Available	Modelcard	Only minimal detail provided
Available	Datasheet	No datasheet
No package	**Package**	No package
	API	

Is there a packaged release available?

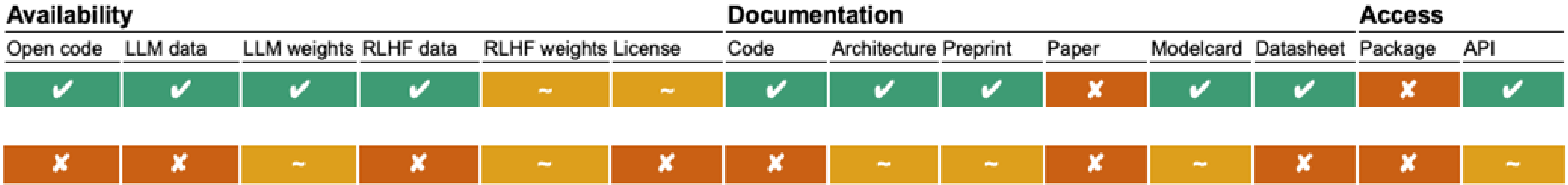
a BigScience initiative BLOOM(z) 176B params 59 languages Open-access		Llama-2-7B Llama-2-13B Llama-2-70B
Fully shared code to enable reproduction	Open code	No training code available
Training data shared	LLM data	No training data available
Access to base LLM without instruction tuning	LLM weights	Accessible after registration
Accessible	RLHF data	No data
Training checkpoint available to download	RLHF weights	Accessible after registration
Code: Apache 2, Model: “RAIL”, non OSI	License	“Community license”, non OSI
Accessible and well-maintained	Code	Only minimal examples
Accessible and documented in preprint	Architecture	Sketched in preprint
Multiple detailed preprints	Preprint	Corporate preprint only
No paper	Paper	No paper
Available	Modelcard	Only minimal detail provided
Available	Datasheet	No datasheet
No package	Package	No package
“Petals API” available via huggingface	**API**	Limited access, sign-up required

Is there an openly available API with unrestricted access?



Two extremes

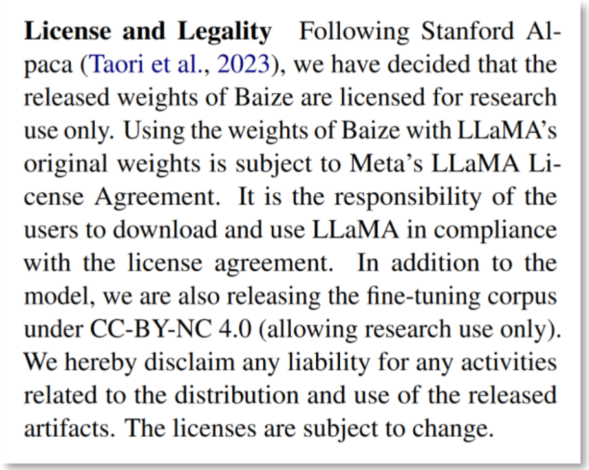
- Both claim to be "open source" — only one is
- Drilling into details makes differences visible
- Evidence-based judgements help
  - to credit initiatives for care taken in developing and releasing AI technology
  - to puncture corporate hype
  - to call out hijacking of terms like "open source"



Availability						Documentation						Access	
Open code	LLM data	LLM weights	RLHF data	RLHF weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
✔	✔	✔	✔	~	~	✔	✔	✔	✘	✔	✔	✘	✔
✘	✘	~	✘	~	✘	✘	~	~	✘	~	✘	✘	~

Surveying 25+ text generators: recurring issues

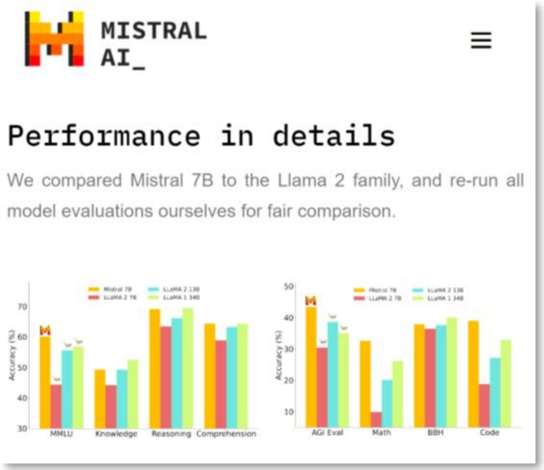
1. Inherited data is common & legal murkiness ensues
2. Synthetic data is on the rise, with unknown consequences
3. “Release by blogpost” should not be accepted as sufficient



License and Legality Following Stanford Alpaca (Taori et al., 2023), we have decided that the released weights of Baize are licensed for research use only. Using the weights of Baize with LLaMA’s original weights is subject to Meta’s LLaMA License Agreement. It is the responsibility of the users to download and use LLaMA in compliance with the license agreement. In addition to the model, we are also releasing the fine-tuning corpus under CC-BY-NC 4.0 (allowing research use only). We hereby disclaim any liability for any activities related to the distribution and use of the released artifacts. The licenses are subject to change.

>40% of LLMs we survey now use *synthetic data** for instruction-tuning

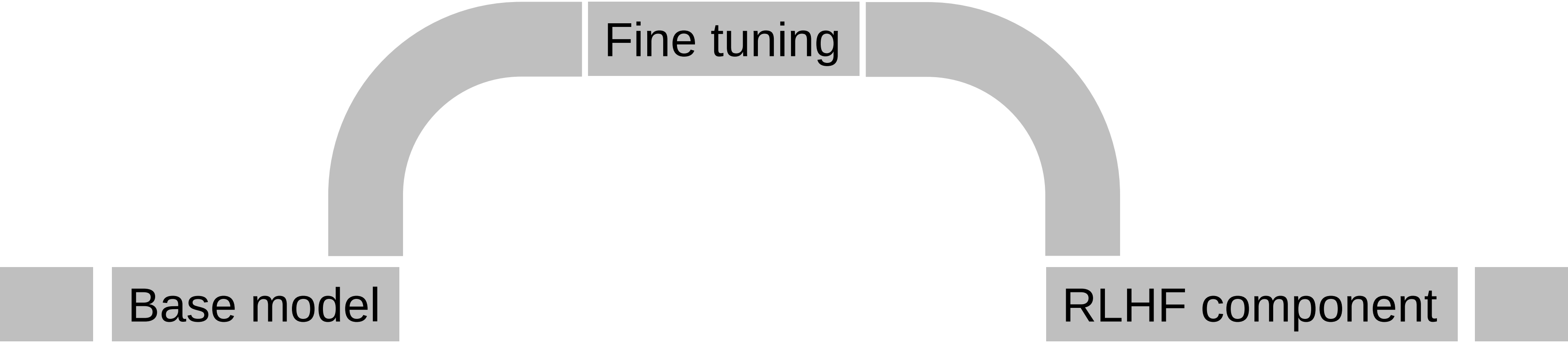
* prompts, responses, or ratings harvested from other LLMs



MISTRAL AI_

Performance in details

We compared Mistral 7B to the Llama 2 family, and re-run all model evaluations ourselves for fair comparison.

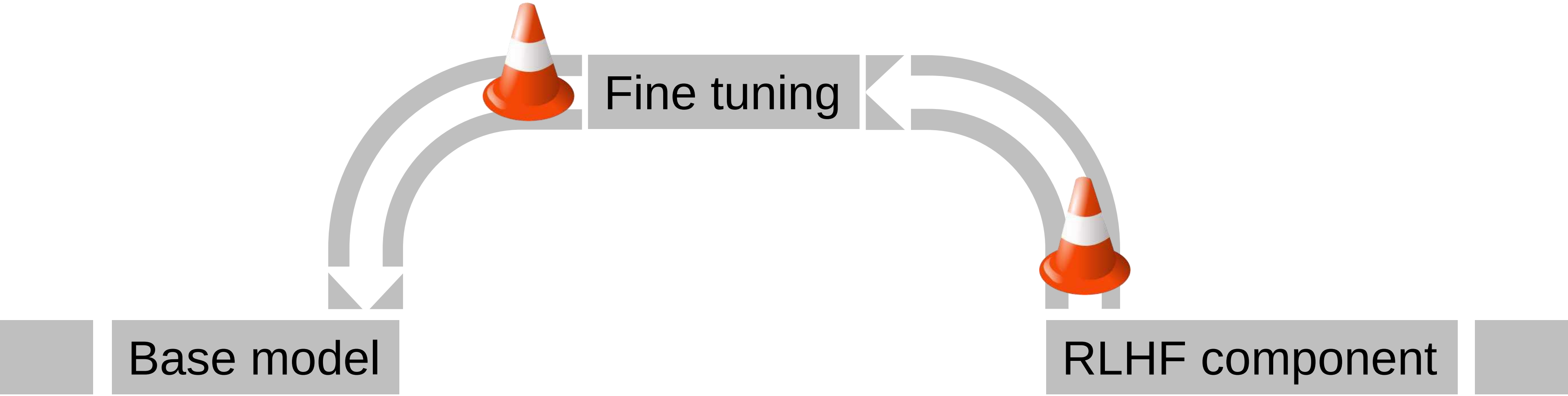


Fine tuning

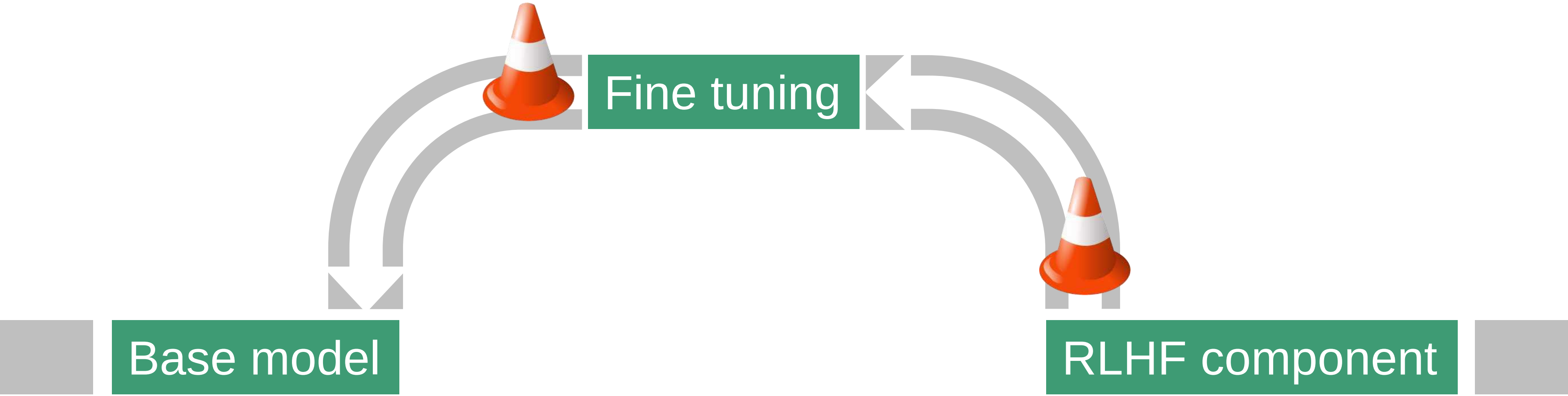
Base model

RLHF component

- Current AI systems are complex and multi-part



- Current AI systems are complex and multi-part — how to reverse engineer?
- Downstream elements can **obstruct** access to earlier parts



- Current AI systems are complex and multi-part — how to reverse engineer?
- Downstream elements can **obstruct** access to earlier parts (“roadblocks”)
- True openness only possible if intermediate steps **documented & opened up**
- Supply source at each roadblock to preserve reverse engineerability

Conclusions

Our approach

- Isolate most relevant dimensions of openness (relative to system)
- Provide evidence-based judgements of openness on those
- All work done out in the open: opening-up-chatgpt.io

Towards a definition of "open" AI systems

- For any genAI system, openness will be composite & graded
- *No one-size fits all solution*: domain knowledge needed to identify relevant dimensions
- Preserve the spirit of reverse engineerability