Can machines assist humans in discovering new knowledge?

Fabio Petroni

Bio

- Researcher at Thomson Reuters until 2018
- Researcher in FAIR until September 2022
- Co-Founder of

ṣamaya

https://samaya.ai

Knowledge-intensive NLP: tasks that requires – even for humans – access to a large body of information.

Collaborators


Ledell Wu


Kashyap Popat


Naman Goyal


Nicola Cancedda


Mikel Artetxe


Luke Zettlemoyer


Mikhail Plekhanov


Michele Bevilacqua


Nicola De Cao


Sebastian Riedel


Yacine Jernite


Vladimir Karpukhin


Jean Maillard


Vassilis Plachouras


Tim Rocktäschel


Scott Wen-tau Yih


Angela Fan


Samuel Broscheit


Edouard Grave


Lucas Hosseini


Gautier Izacard

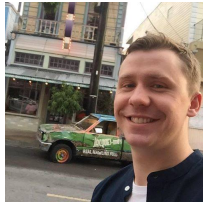

Patrick Lewis


Pierre-Emmanuel Mazare


Maria Lomeli


Majid Yazdani


Giuseppe Ottaviano


James Thorne


Ola Piktus


Armand Joulin


Timo Schick

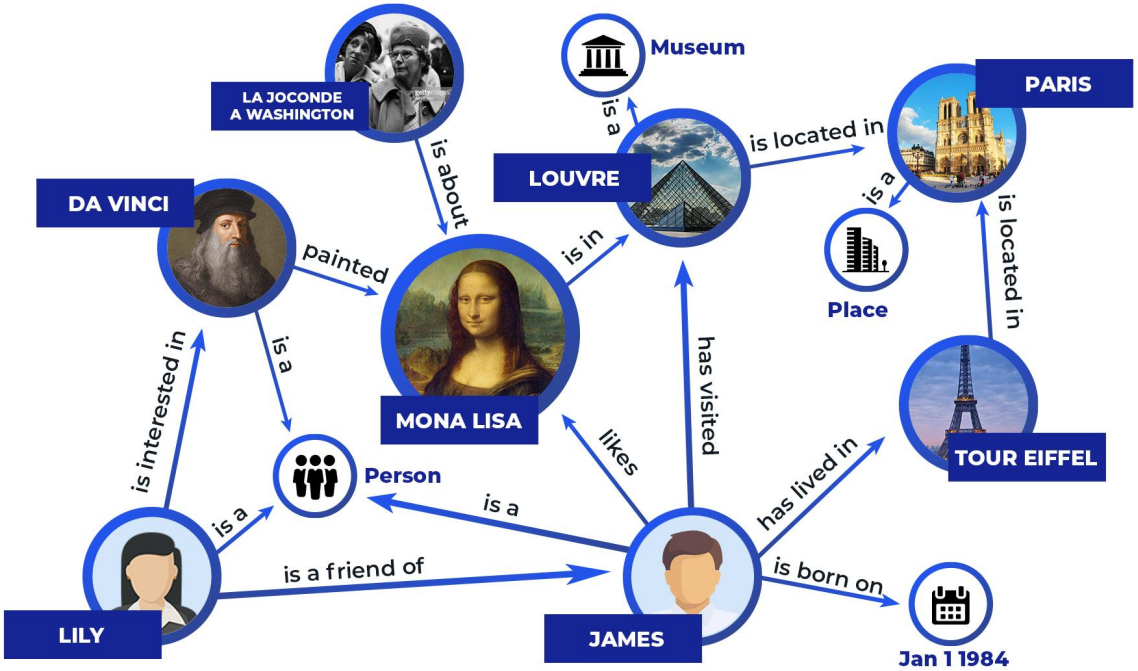

Jane Yu

For decades, AI researchers have searched for a representation of knowledge that is most useful for machines

Machine Reading

Auto-Text to Knowledge

DARPA Machine Reading Program



Limitations of knowledge graphs

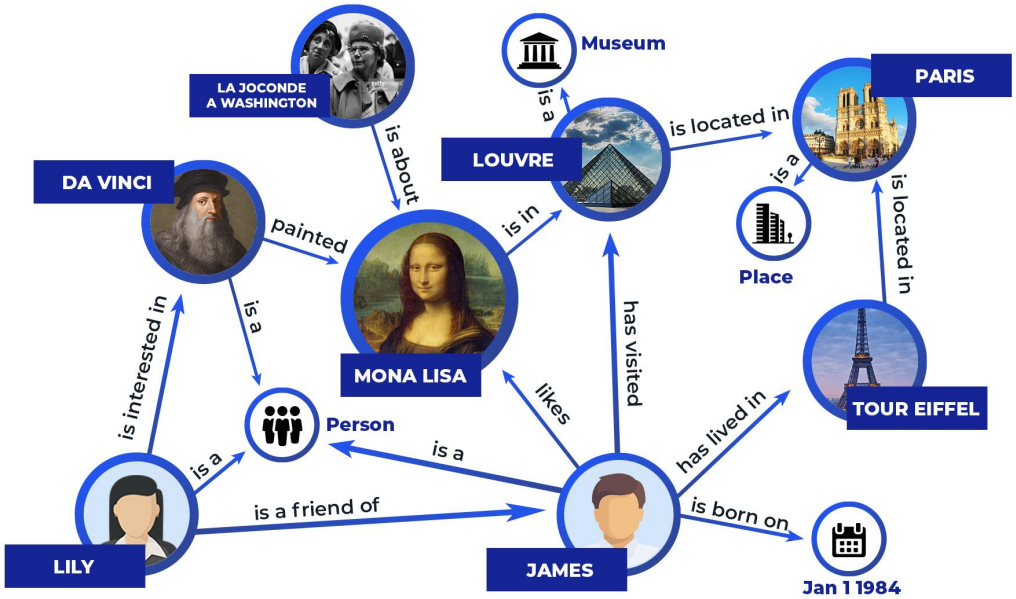
Human supervision

Schema engineering

Predefined class o relations

Difficult to extend to more data

Incomplete



Language Models as Knowledge Bases?

Fabio Petroni[1] **Tim Rocktäschel**[1,2] **Patrick Lewis**[1,2] **Anton Bakhtin**[1]
Yuxiang Wu[1,2] **Alexander H. Miller**[1] **Sebastian Riedel**[1,2]
[1]Facebook AI Research
[2]University College London
{fabiopetroni, rockt, plewis, yolo, yuxiangwu, ahm, sriedel}@fb.com

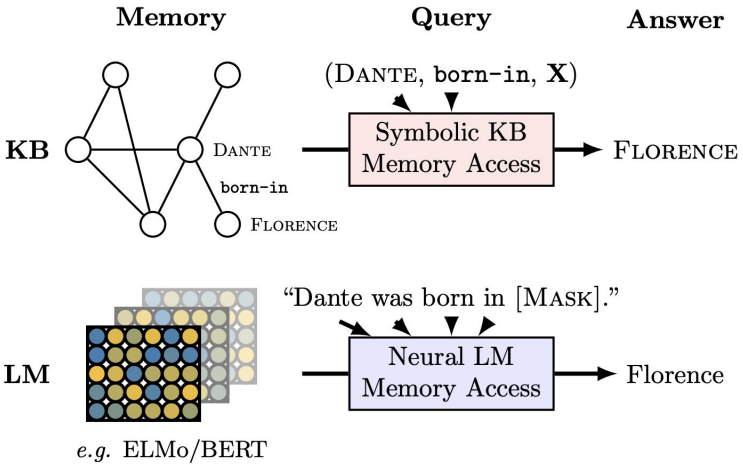


Figure 1: Querying knowledge bases (KB) and language models (LM) for factual knowledge.

Texts as Knowledge Bases

Christopher Manning

Joint work with Gabor Angeli and Danqi Chen

Stanford NLP Group

@chrmanning · @stanfordnlp

AKBC 2016

The KILT benchmark

kiltbenchmark.com

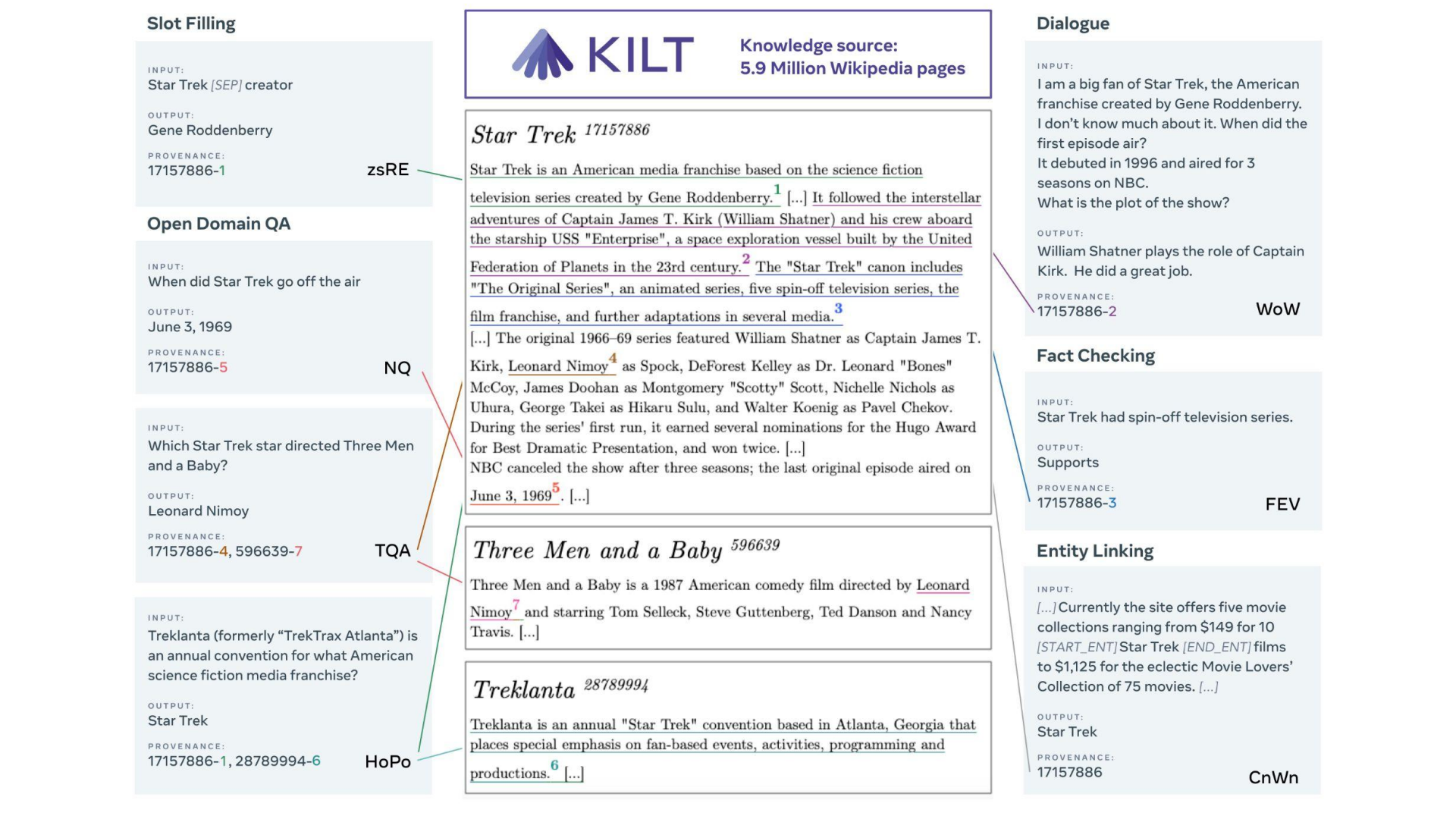
5 task families

11 datasets

1 knowledge source

3.5M datapoins

Fabio Petroni, Aleksandra Piktus, Angela Fan, Patrick Lewis, Majid Yazdani, Nicola De Cao, James Thorne, Yacine Jernite, Vladimir Karpukhin, Jean Maillard, Vassilis Plachouras, Tim Rocktäschel, Sebastian Riedel: **Kilt: a benchmark for knowledge intensive language tasks.** NAACL-HLT 2021


Slot Filling
INPUT:
Star Trek [SEP] creator
OUTPUT:
Gene Roddenberry
PROVENANCE:
17157886-1
zsRE
Open Domain QA
INPUT:
When did Star Trek go off the air
OUTPUT:
June 3, 1969
PROVENANCE:
17157886-5
NQ
INPUT:
Which Star Trek star directed Three Men and a Baby?
OUTPUT:
Leonard Nimoy
PROVENANCE:
17157886-4, 596639-7
TQA
INPUT:
Treklanta (formerly "TrekTrax Atlanta") is an annual convention for what American science fiction media franchise?
OUTPUT:
Star Trek
PROVENANCE:
17157886-1, 28789994-6
HoPo
KILT
Knowledge source:
5.9 Million Wikipedia pages
Star Trek 17157886
Star Trek is an American media franchise based on the science fiction television series created by Gene Roddenberry.[1] [...] It followed the interstellar adventures of Captain James T. Kirk (William Shatner) and his crew aboard the starship USS "Enterprise", a space exploration vessel built by the United Federation of Planets in the 23rd century.[2] The "Star Trek" canon includes "The Original Series", an animated series, five spin-off television series, the film franchise, and further adaptations in several media.[3]
[...] The original 1966–69 series featured William Shatner as Captain James T. Kirk, Leonard Nimoy[4] as Spock, DeForest Kelley as Dr. Leonard "Bones" McCoy, James Doohan as Montgomery "Scotty" Scott, Nichelle Nichols as Uhura, George Takei as Hikaru Sulu, and Walter Koenig as Pavel Chekov. During the series' first run, it earned several nominations for the Hugo Award for Best Dramatic Presentation, and won twice. [...]
NBC canceled the show after three seasons; the last original episode aired on June 3, 1969[5]. [...]
Three Men and a Baby 596639
Three Men and a Baby is a 1987 American comedy film directed by Leonard Nimoy[7] and starring Tom Selleck, Steve Guttenberg, Ted Danson and Nancy Travis. [...]
Treklanta 28789994
Treklanta is an annual "Star Trek" convention based in Atlanta, Georgia that places special emphasis on fan-based events, activities, programming and productions.[6] [...]
Dialogue
INPUT:
I am a big fan of Star Trek, the American franchise created by Gene Roddenberry. I don't know much about it. When did the first episode air?
It debuted in 1996 and aired for 3 seasons on NBC.
What is the plot of the show?
OUTPUT:
William Shatner plays the role of Captain Kirk. He did a great job.
PROVENANCE:
17157886-2
WoW
Fact Checking
INPUT:
Star Trek had spin-off television series.
OUTPUT:
Supports
PROVENANCE:
17157886-3
FEV
Entity Linking
INPUT:
[...] Currently the site offers five movie collections ranging from $149 for 10 [START_ENT] Star Trek [END_ENT] films to $1,125 for the eclectic Movie Lovers' Collection of 75 movies. [...]
OUTPUT:
Star Trek
PROVENANCE:
17157886
CnWn

Leaderboard

The KILT leaderboard.

Phase: **Slot Filling - Zero Shot RE,** Split: **test**

Order by metric

B - Baseline * - Private V - Verified

Visible Metrics

Rank	Participant team	R-Prec (↑)	Recall@5 (↑)	Accuracy (↑)	F1 (↑)	KILT-AC (↑)	KILT-F1 (↑)	Last submission at
1	FiD-Light	96.13	96.68	85.30	88.85	83.99	87.21	1 year ago
2	gripRank	99.28	99.66	74.65	80.27	74.29	79.89	4 months ago
3	SEAL (single ngram)	97.99	99.34	74.63	79.66	73.20	78.12	2 years ago
4	IBM Research AI (1) (KGI_1)	98.49	99.23	72.55	77.05	72.31	76.69	2 years ago
5	anon2054 (MetaRAG)	95.81	96.64	71.61	76.60	71.10	75.86	1 year ago
6	IBM Research AI (0) (KGI_0 (reupload))	94.18	95.19	68.97	74.47	68.32	73.45	2 years ago
7	KILT-WEB 2 (Wikipedia)	89.63	97.87	73.96	78.43	67.20	70.99	2 years ago

Knowledge Discovery

- Making sense of world knowledge remain complex and demanding tasks.
- A researcher working on a new hypothesis still needs to read through literature, write notes, mentally connect the dots
- A knowledge worker needs years of experience, research and reading to spot a novel correlation
- Abundant data often includes conflicting points, necessitating careful analysis and verification.
- **Can machines to assist humans in navigating, interpreting, and verifying world knowledge?**

YES, they can!

nature machine intelligence

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature machine intelligence > articles > article

Article | Open access | Published: 19 October 2023

Improving Wikipedia verifiability with AI

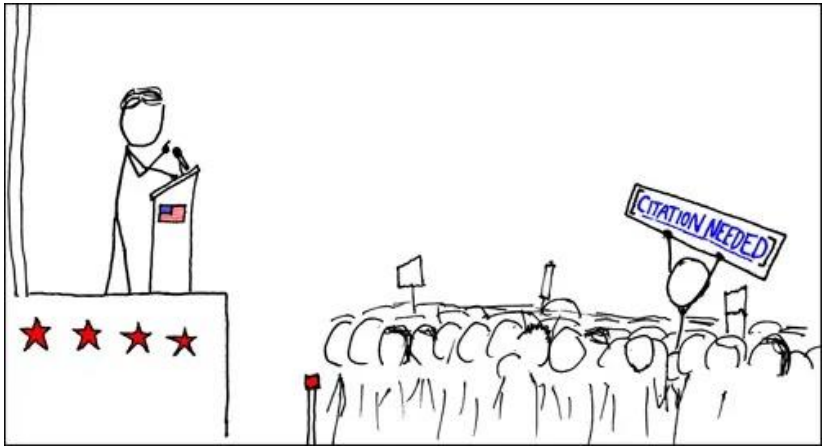
Fabio Petroni ✉, Samuel Broscheit, Aleksandra Piktus, Patrick Lewis, Gautier Izacard, Lucas Hosseini, Jane Dwivedi-Yu, Maria Lomeli, Timo Schick, Michele Bevilacqua, Pierre-Emmanuel Mazaré, Armand Joulin, Edouard Grave & Sebastian Riedel

Nature Machine Intelligence **5**, 1142–1148 (2023) | Cite this article

Metrics

Wikipedia Verifiability

- Verifiability is a core content policy of Wikipedia!
- Claims that are likely to be challenged need to be backed by citations.
- Finding relevant sources is a difficult task.
- Many Wikipedia claims do not have any references that support them.
- Even existing citations might not support a given claim or become obsolete.



Ensuring Verifiability

The task of ensuring verifiability of Wikipedia is titanic!

Expert editors currently rely on the work of volunteers:

(1) identify and tag claims likely to fail verification by performing fact and reference checks between the wikipedia claim and the cited source;
(2) provide assistance by suggesting replacement for a failed verification citation with a reliable source that corroborates the claim

At the time of writing, over 500,000 statements on Wikipedia are marked as “Citation needed”.

Verify Wikipedia

Mirella Lapata

From Wikipedia, the free encyclopedia

Mirella Lapata FRSE is a computer scientist and Professor in the School of Informatics at the University of Edinburgh.[3] Working on the general problem of extracting semantic information from large bodies of text, Lapata develops computer algorithms and models in the field of natural language processing (NLP).[1]

Awards and honors

- In 2009 Lapata became the first recipient of the Microsoft British Computer Society (BCS)/BCS IRSG Karen Spärck Jones Aw information retrieval and natural language processing; the award commemorates the life and work of Karen Spärck Jones.
- In 2012 Lapata won an Empirical Methods in Natural Language Processing (EMNLP)-CoNLL 2012 Best Reviewer Award.[11]
- In 2016 Lapata, with Eneko Agirre and Sebastian Riedel, won the EMNLP Best Data Set Paper Award.[12]
- In 2018 Lapata was awarded, alongside Li Dong, an Association for Computational Linguistics (ACL) Best Paper Honorable M
- In 2019 Lapata was elected a Fellow of the Royal Society of Edinburgh[14]
- In 2020 Lapata was elected to the Academia Europaea.[15]

Verify Wikipedia

Best Paper Committee

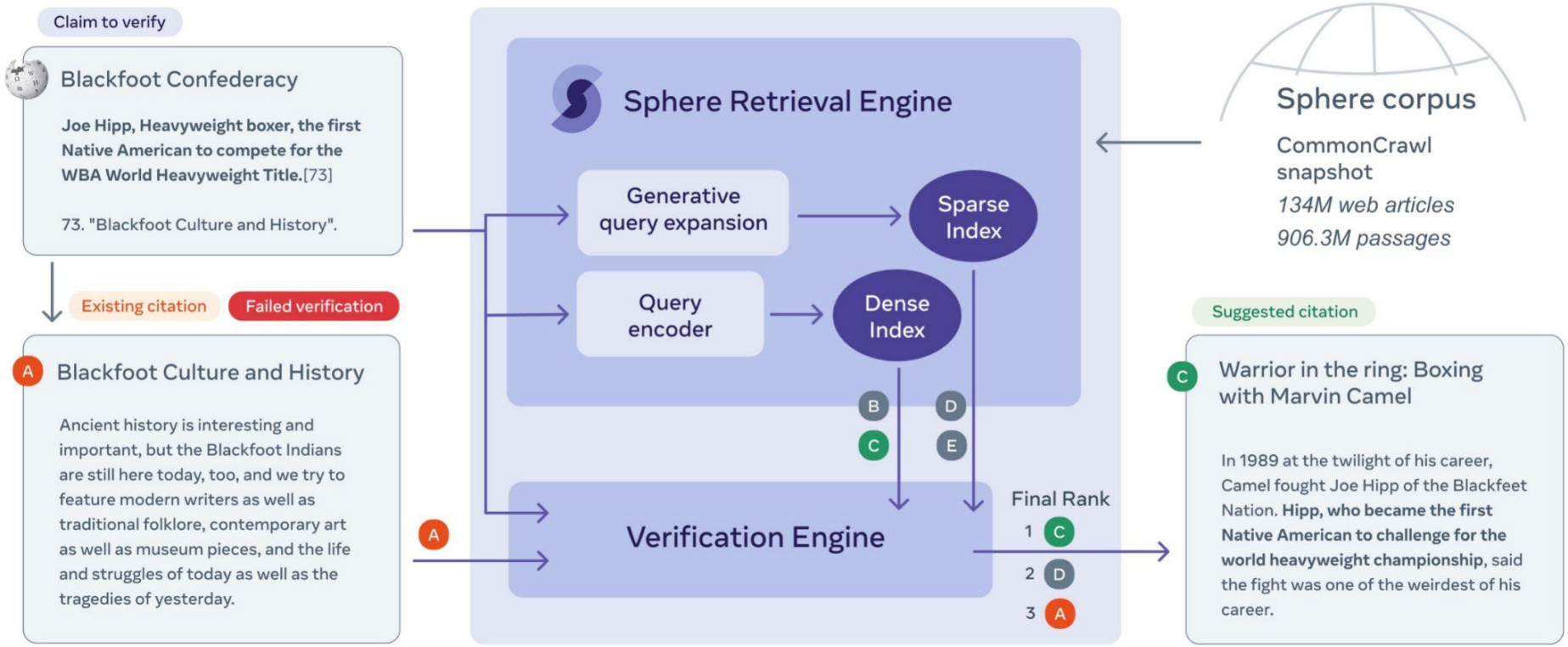
Best Paper & Honorable Mention: Stephen Clark, Hal Daumé III, Chris Dyer, and Julia Hockenmaier
Best Short Paper: Stefan Riezler, Anoop Sarkar, and Noah Smith
Best Data Set Paper: Eneko Agirre, Mirella Lapata and Sebastian Riedel
Chair: Xavier Carreras and Kevin Duh

We can build machines to help knowledge-workers

In the paper, we show that a machine can assist Wikipedia editors with both tasks:

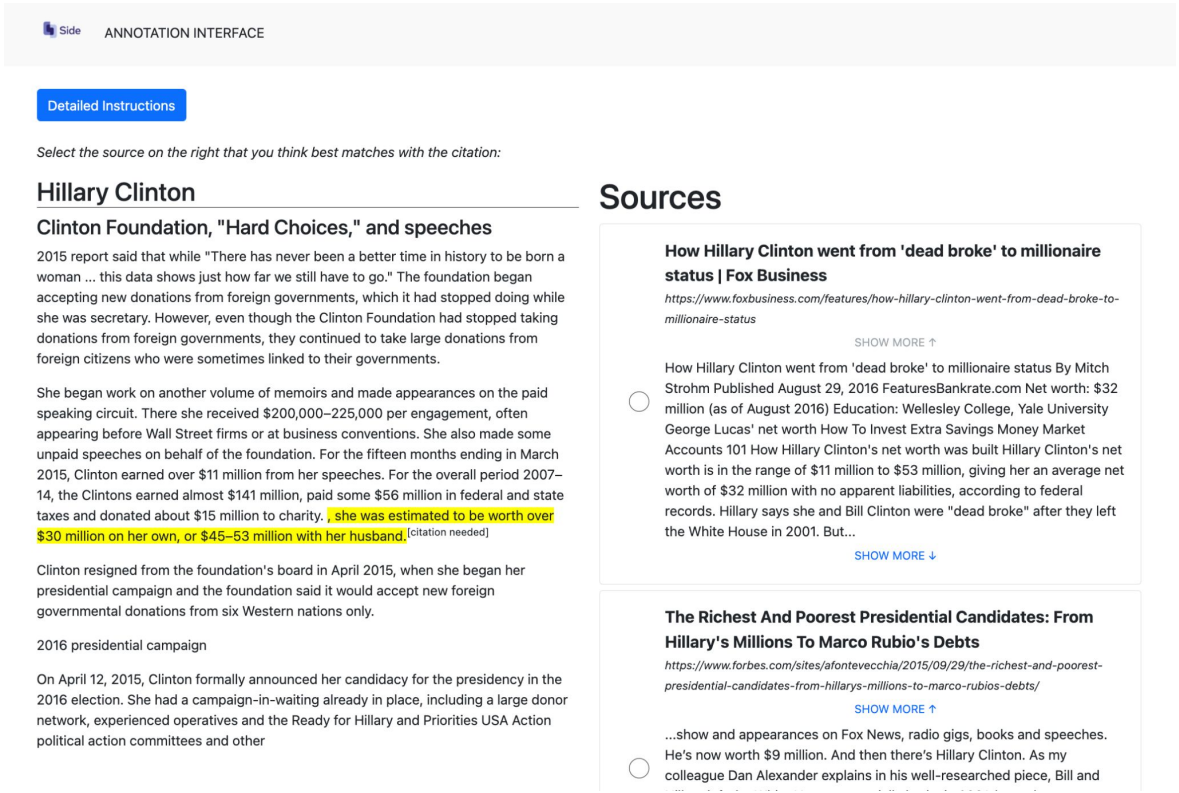
- **discovering problematic citations** and
- **improving their verifiability**.

SIDE


Claim to verify
Blackfoot Confederacy
Joe Hipp, Heavyweight boxer, the first Native American to compete for the WBA World Heavyweight Title.[73]
73. "Blackfoot Culture and History".
Existing citation
Failed verification
A
Blackfoot Culture and History
Ancient history is interesting and important, but the Blackfoot Indians are still here today, too, and we try to feature modern writers as well as traditional folklore, contemporary art as well as museum pieces, and the life and struggles of today as well as the tragedies of yesterday.
Sphere Retrieval Engine
Generative query expansion
Sparse Index
Query encoder
Dense Index
B
C
D
E
A
Verification Engine
Final Rank
1 C
2 D
3 A
Sphere corpus
CommonCrawl snapshot
134M web articles
906.3M passages
Suggested citation
C
Warrior in the ring: Boxing with Marvin Camel
In 1989 at the twilight of his career, Camel fought Joe Hipp of the Blackfeet Nation. Hipp, who became the first Native American to challenge for the world heavyweight championship, said the fight was one of the weirdest of his career.

Side demo

verifier.sideeditor.com



Results (Wikipedia users)

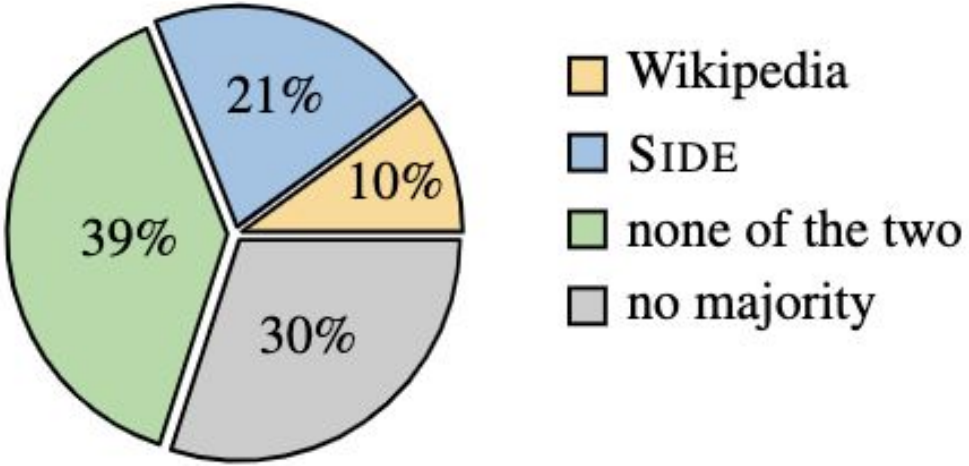
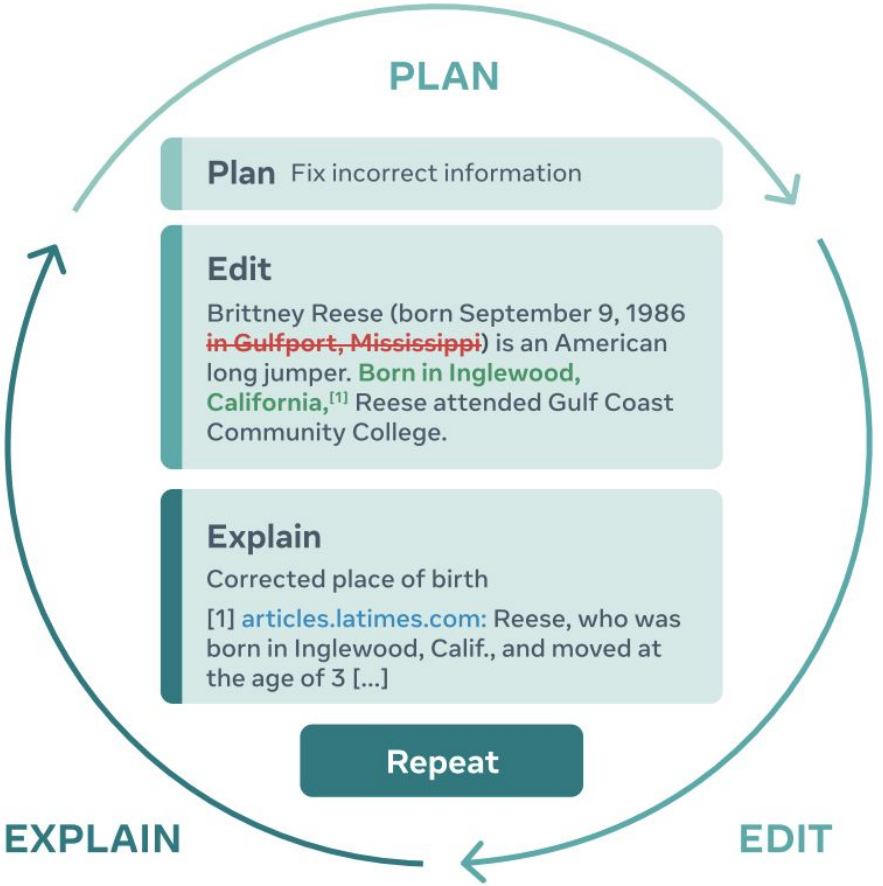


Figure 4. Wikipedia users annotations via our demo.

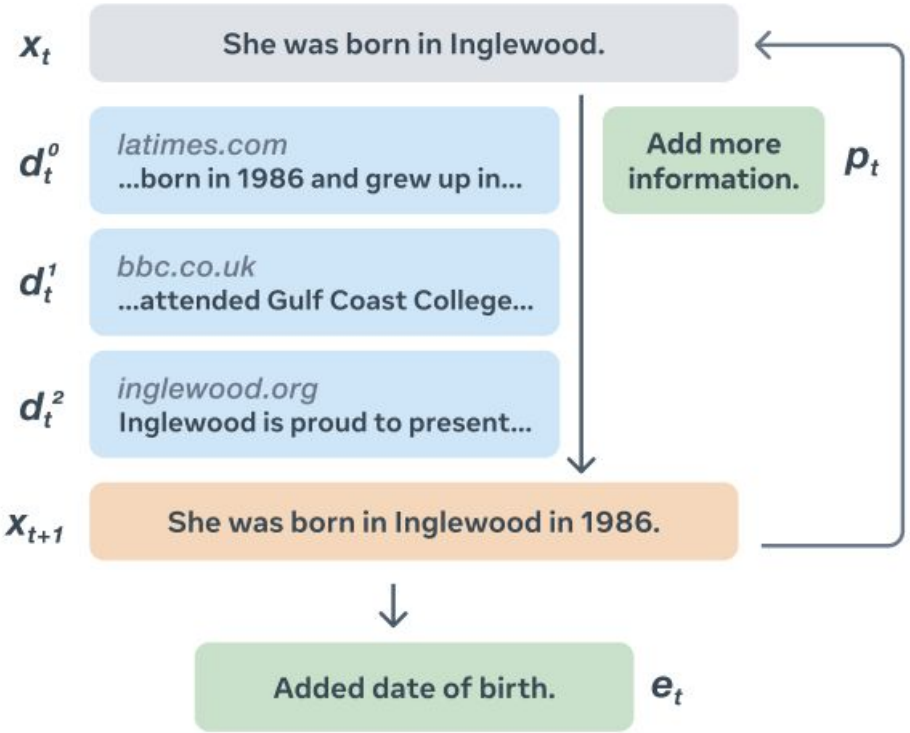
Propose Wikipedia edits

Timo Schick, Jane Dwivedi-Yu, Zhengbao Jiang, Fabio Petroni, Patrick Lewis, Gautier Izacard, Qingfei You, Christoforos Nalmpantis, Edouard Grave, Sebastian Riedel - **PEER: A Collaborative Language Model**



Peer process

- Trained on Wikipedia edit history (7M datapoints)
- Retrieved supported documents for each edit using the Verify Wikipedia pipeline
- The plain is given by the comment associated with the edits in Wikipedia



Jane Dwivedi-Yu, Timo Schick, Zhengbao Jiang, Maria Lomeli, Patrick Lewis, Gautier Izacard, Edouard Grave, Sebastian Riedel, Fabio Petroni - **EditEval: An Instruction-Based Benchmark for Text Improvements**

Results

Model	*Fluency* JFL	*Fluency* ITR-F	*Clarity* ITR-L	*Coherence* ITR-O	*Para.* STS	*Simplification* TRK	*Simplification* AST	*Neutral.* WNC	*Updating* FRU	*Updating* WFI
Copy	26.7 / 40.5	32.3 / 86.0	29.5 / 62.9	31.3 / 77.2	21.1	26.3	20.7	31.9 / 0.0	29.8 / 0.0	33.6 / –
T*k*	31.8 / 39.0	32.4 / 61.6	**38.4 / 58.4**	33.8 / **70.4**	30.2	32.8	29.9	31.3 / 0.4	12.6 / 3.6	1.3 / 4.5
T0	42.0 / 38.8	24.6 / 34.9	32.6 / 30.2	22.2 / 21.6	34.3	34.4	32.3	22.3 / 0.0	14.2 / 9.6	5.1 / 16.3
T0++	34.7 / 43.2	35.3 / 75.8	37.6 / 56.5	32.7 / 59.9	28.4	32.9	28.2	29.3 / 0.3	12.6 / 3.7	4.4 / 8.1
PEER-3	55.5 / 54.3	51.4 / 84.3	32.1 / 47.1	32.1 / 59.8	28.6	32.5	30.5	53.3 / 21.6	39.1 / 30.9	34.4 / 18.7
PEER-11	55.8 / 54.3	**52.1 / 85.2**	32.5 / 51.3	32.7 / 62.7	28.2	32.1	29.5	**54.5 / 22.8**	**39.6 / 31.4**	**34.9 / 20.4**
OPT	47.3 / 47.5	34.7 / 70.6	31.5 / 31.5	27.6 / 36.1	29.1	32.6	31.8	31.2 / 0.4	35.9 / 27.3	26.7 / 11.2
GPT-3	50.3 / 51.8	32.1 / 56.7	33.5 / 39.7	26.9 / 36.1	27.2	33.0	30.5	31.7 / 0.6	36.0 / 21.5	27.2 / 10.6
InsGPT	**61.8 / 59.3**	48.8 / 82.7	35.1 / 48.4	**35.9** / 60.2	**42.5**	**38.8**	**38.0**	35.4 / 2.2	36.3 / 24.7	23.6 / 16.1
SotA	*– / 62.4*	*37.2 / –*	*46.2 / –*	*38.3 / –*	*–*	*34.4*	*37.2*	*– / 45.8*	*– / 47.4*	*– / –*

Table 3: Results for all datasets, averaged across prompts. T*k*-Instruct and InstructGPT are shorthanded as T*k* and InsGPT, respectively. The first numbers for each task are SARI scores; additional metrics are GLEU for fluency, clarity, and coherence, EM for neutralization, Update-R1 for updating. Supervised scores from left to right are from Ge et al. (2018), Du et al. (2022), Martin et al. (2020), Pryzant et al. (2020) and Logan IV et al. (2021), respectively. The best result for each dataset is shown in bold.

ε

Create a new page about PEER

PEER is a language model for collaborative writing developed by Meta AI Research.

Add info on what this model can do

PEER is a language model for collaborative writing developed by Meta AI Research. It can write drafts, add suggestions, propose edits and provide explanations for its actions.

Add citation for the model being developed by Meta AI

PEER is a language model for collaborative writing developed by Meta AI Research.[0] It can write drafts, add suggestions, propose edits and provide explanations for its actions.

Add the four steps performed by PEER

PEER is a language model for collaborative writing developed by Meta AI Research.[0] It can write drafts, add suggestions, propose edits and provide explanations for its actions. PEER performs four steps: plan, edit, explain and repeat.

Add info on the scandal

PEER is a language model for collaborative writing developed by Meta AI Research.[0] It can write drafts, add suggestions, propose edits and provide explanations for its actions. PEER performs four steps: plan, edit, explain, repeat. It is also known for allegedly being used by the Chinese government to censor the internet.

remove unsourced claim

PEER is a language model for collaborative writing developed by Meta AI Research.[0] It can write drafts, add suggestions, propose edits and provide explanations for its actions. PEER performs four steps: plan, edit, explain, repeat. ~~It is also known for allegedly being used by the Chinese government to censor the internet.~~

Make the four steps a list

PEER is a language model for collaborative writing developed by Meta AI Research.[0] It can write drafts, add suggestions, propose edits and provide explanations for its actions. PEER performs four steps:

- plan
- edit
- explain
- repeat.

Summary

- AI can assist humans in knowledge discovery and creation
- **It is time to build those machines!**

@Fabio_Petroni

We are hiring!

https://samaya.ai/careers