ORES
Scalable, community-driven, auditable,
machine learning-as-a-service
WIKIMEDIA
FOUNDATION

In this presentation, we lay out ideas...

This is a conversation starter, not a conversation ender. We put forward our best ideas and thinking based on our partial perspectives.

We look forward to hearing yours.

Outline

- Opportunities and fears of artificial intelligence
- ORES: A break-away success
- The proposal & asks

WIKIMEDIA
FOUNDATION

We face a significant challenge...

There's been a significant decline in the number of active contributors to the English-language Wikipedia. Contributors have fallen by 40% over the past eight years, to about 30,000.

Hamstrung by our own success?

“Research indicates that the problem may be rooted in Wikipedians’ complex bureaucracy and their often hard-line responses to newcomers’ mistakes, enabled by semi-automated tools that make deleting new changes easy.” - MIT Technology Review (December 1, 2015)

It’s the revert-new-editors’-first-few-edits-and-alienate-them problem.

Imagine an algorithm that can...

- Judge whether an edit was made in good faith, so that new editors don't have their first contributions wiped out.
- Detect bad-faith edits and help fight vandalism.

Machine learning and prediction have the power to support important work in open contribution projects like Wikipedia at massive scales.

But it's far from a no-brainer. Quite the opposite.

On the bright side,
Machine learning has the potential to help our projects scale
by reducing the workload of editors
and enhancing the value of our content.

by Luiz Augusto Fornasier Milani under CC BY-SA 4.0,
from Wikimedia Commons.

On the dark side,
AIs have the potential to
perpetuate biases and silence voices
in novel and insidious ways.

by Paulo Barcellos Jr.under CC BY-SA 2.0, from Wikimedia Commons.

Any solution would need to balance efficiency and accuracy with transparency, ethics, and fairness.

But if we got it right...

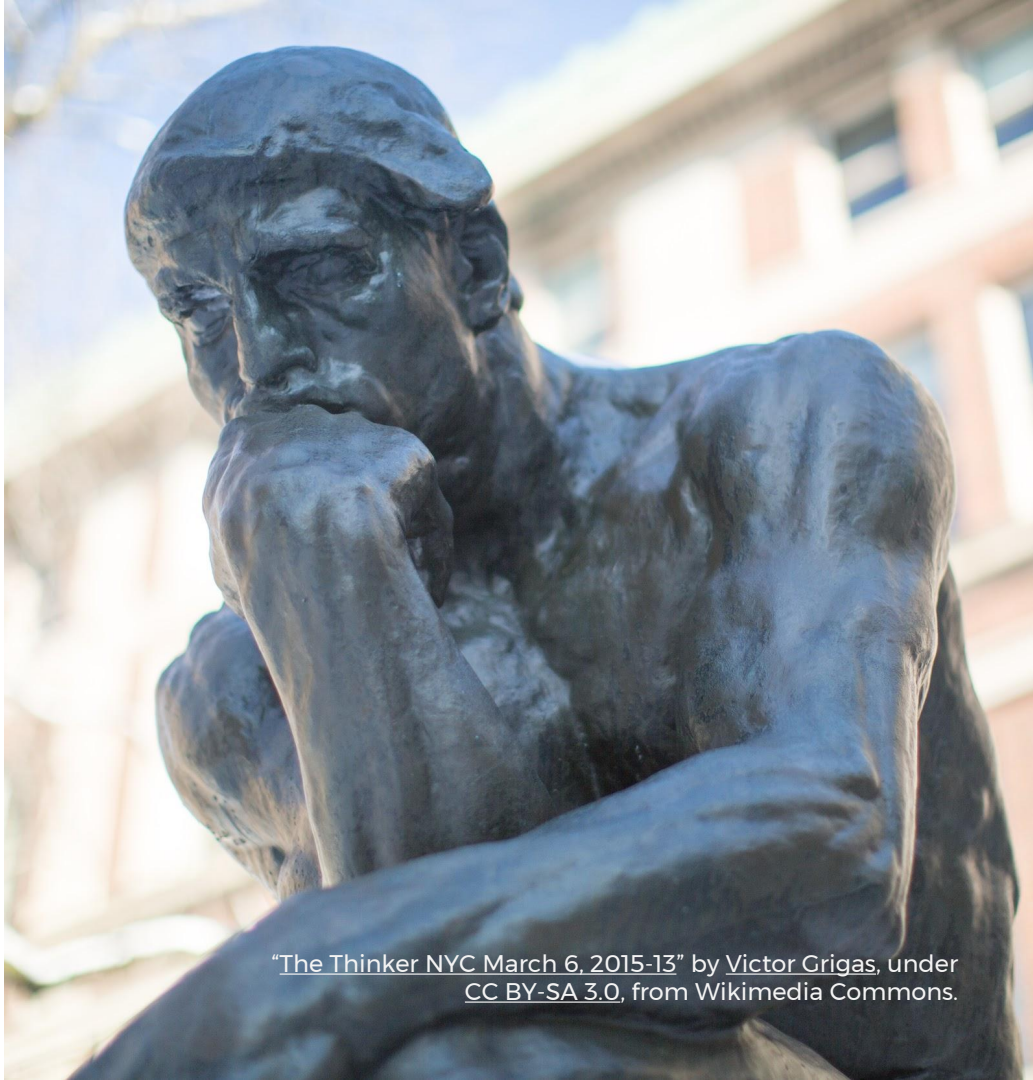
It could help retain editors and fight vandalism.
Plus, we could join other leaders in ethical, effective AI for social impact.

Enter ORES

A research focused, community supported, AI-as-a-service that began by deploying algorithms to detect good and bad faith edits.

ORES is a high capacity, machine learning prediction service that's been online since 2014.

A collection of technologies that enact an AI-as-a-service platform.



"The Thinker NYC March 6, 2015-13" by Victor Grigas, under CC BY-SA 3.0, from Wikimedia Commons.

Our communities support this initiative

From the Community Wishlist:

"On the English Wikipedia, where the new-article volume is very high, managing the influx of new articles consumes a large amount of volunteer time. (...) The proposed solution is to use an ORES-style machine learning mechanism to score newly created articles."

–Opabinia regalis (talk) 22:51, 19 November 2016 (UTC)

This quarter, we built this prediction model. ✓

Voting - Quality scoring for new articles

1. ⊕ **Support**--Shizhao (talk) 03:02, 28 November 2016 (UTC)
2. ⊕ **Support** --Kudpung (talk) 12:24, 28 November 2016 (UTC)
3. ⊕ **Support** MichaelMaggs (talk) 19:58, 28 November 2016 (UTC)
4. ⊕ **Support** FoCuSandLeArN (talk) 17:27, 29 November 2016 (UTC)
5. ⊕ **Support** --*Telaneo* (User talk page) 21:56, 29 November 2016 (UTC)
6. ⊕ **Support** --Anthonyhcole (talk) 06:47, 30 November 2016 (UTC)
7. ⊕ **Support** Noyster (talk) 00:18, 1 December 2016 (UTC)
8. ⊕ **Support** MER-C (talk) 05:23, 1 December 2016 (UTC)
9. ⊕ **Support** as proposer. Opabinia regalis (talk) 03:59, 4 December 2016 (UTC)
10. ⊕ **Support** I am amazed this obvious great idea doesn't have a lot more supp December 2016 (UTC)
11. ⊖ **Oppose** Per my vote under #New pages Feed/Page Curation. --Liuxinyu97 Liuxinyu970226, that vote appeared to have nothing to do with the subject, nor 15:10, 6 December 2016 (UTC)
12. ⊕ **Support**, per Kudpung comments above. Wikid77 (talk) 00:04, 6 December
13. ⊕ **Support**--Ranjithsiji (talk) 11:17, 6 December 2016 (UTC)
14. ⊕ **Support** **Blue Rasberry** (talk) 19:26, 6 December 2016 (UTC)
15. ⊕ **Support** **Ks0stm** (T·C·G) 21:22, 6 December 2016 (UTC)
16. ⊕ **Support** I'm surprised this extremely good idea isn't getting more traction. 2016 (UTC)
17. ⊕ **Support** ***Miniapolis*** 20:10, 9 December 2016 (UTC)

ORES is a *platform*.

ORES is a platform-level intervention—an infrastructure for other work.

Our users are developers.

As a platform, it spans many different use cases (e.g. a single model has been used in 25 different applications thus far). It's a functionality that has broad applications and a whole ecosystem of features.

ORES is a collection of technologies...

that enact AI as a service platform.

revscoring
A generalized MediaWiki prediction model builder

ORES
A scalable model hosting service

Wiki labels
A labeled-data gathering interface

https://commons.wikimedia.org/wiki/File:Revision_Scoring_as_a_Service_logo.svg
(PD, Frater5)

https://commons.wikimedia.org/wiki/File:Objective_Revision_Evaluation_Service_logo.svg
(PD, Helder, Danilo.mac, Offnfopt)

https://commons.wikimedia.org/wiki/File:Wiki_Labels_X_logo.svg
(CC-BY-SA 3.0, Perhelion)

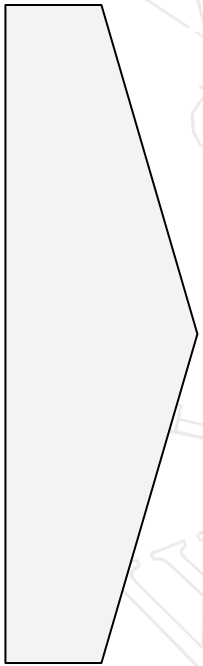
ORES is already deployed and a break-away success.

It’s already making life easier for vandal fighters.

267 hours

Amount of time that English Wikipedians used to spend reviewing RecentChanges for vandalism and damage *every day.*

That's 33 people working 8 hours each.



27 hours

Amount of time that English Wikipedians now spend reviewing the RecentChanges for vandalism and damage *every day* **WITH ORES.**

18,050

The number of editors who have **manually** enabled ORES Review Tool enabled across 8 wikis (English, Portuguese, Persian, Turkish, Russian, Polish, Dutch, and Wikidata)

(12,000 in English Wikipedia alone)

Who uses ORES?

1. Editors

- 18,050 of them

2. Tool developers:

- :pt:Wikipédia:Scripts/FastButtons
- mw-gadget-ScoredRevisions
- tools-Raun
- mw-gadget-RTRC
- :en:User:Reports_bot
- WikiEdu student dashboard (report)
- WikiEdu article finder
- :en:User:SuggestBot
- tools-crosswatch
- Huggle
- :en:User:DataflowBot
- tools-copypatrol

3. Wikimedia Product:

- Edit review improvements (Collaboration)
- Quality ranking experiments (Search)
- *In progress* (Reading Dept.)

other accomplishments

- Basic support to 24 wikis
- Deployed the beta feature to 8 wikis for 18k users
- Increased system scalability by a factor of 4
- Developed basic capabilities for 4 new models
- Supported 3 external research collaborations
- Organized 3 workshops on use of AI and ethical considerations
- Presented about bias detection and other socio-technical issues

WIKIMEDIA
FOUNDATION

ORES has been welcomed by the tech community and press.

- Wikipedia Deploys AI to Expand Its Ranks of Human Editors (Wired)
- Artificial Intelligence Aims to Make Wikipedia Friendlier and Better (MIT Technology Review)
- Wikipedia launches edit-checking artificial intelligence (BBC Technology)

... and many more at m:Research:Revision_scoring_as_a_service/Media

ORES has done all of this despite *virtually no resources*.

$46,875 from two Individual Engagement Grants
+ Staff research scientist @ 50% since July, 2015

1. https://meta.wikimedia.org/wiki/Grants:IEG/Revision_scoring_as_a_service
2. https://meta.wikimedia.org/wiki/Grants:IEG/Revision_scoring_as_a_service/Renewal

In part because the design of the ORES team is unique.

The ORES team is grounded in a network of external researchers. These researchers are critical to scaling the work. Without them, we'd have to in-source all of that brain power. That's tough on a non-profit budget.

These researchers are not ordinary volunteers.
They collaborate formally because of their own work.
It's in line with work that they have already academically committed to.

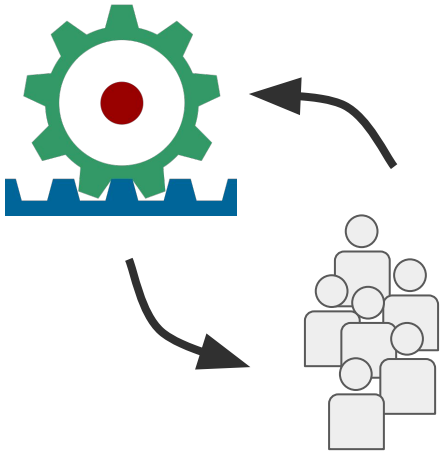
Now it’s time to scale.

We need to hire:

- An engineering manager
- An engineer
- An engineer
- A technical writer
- A community liaison

2017?

Accountability mechanisms



WIKIMEDIA FOUNDATION

https://commons.wikimedia.org/wiki/File:Wiktionary_small.svg
(CC-BY-SA 3.0, Wikimedia)

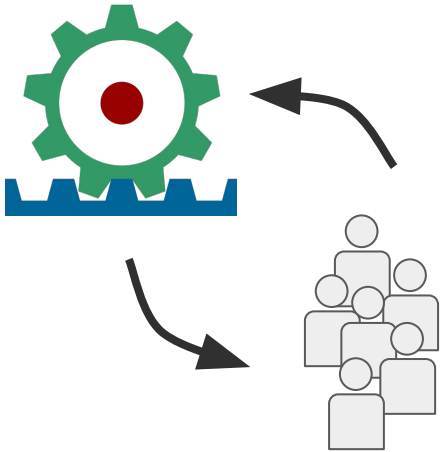
https://commons.wikimedia.org/wiki/File:Wikibooks-logo.svg
(CC-BY-SA 3.0, User:Bastique, User:Ramac)

https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2.svg
(CC-BY-SA 3.0, Wikimedia)

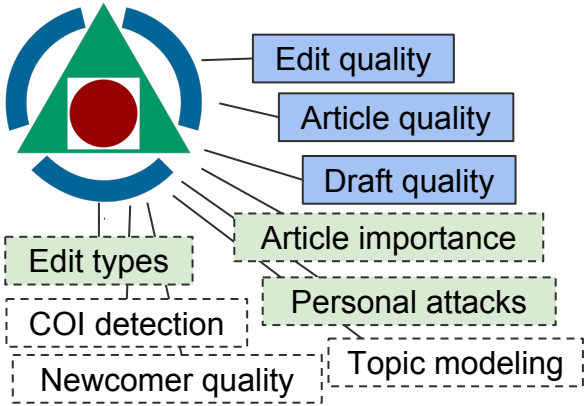
https://commons.wikimedia.org/wiki/File:Notification-icon-Wikidata-logo.svg
(CC-BY-SA 3.0, User:Jdforrester (WMF))

2017?

Accountability mechanisms



More models



https://commons.wikimedia.org/wiki/File:Wiktionary_small.svg
(CC-BY-SA 3.0, Wikimedia)

https://commons.wikimedia.org/wiki/File:Wikibooks-logo.svg
(CC-BY-SA 3.0, User:Bastique, User:Ramac)

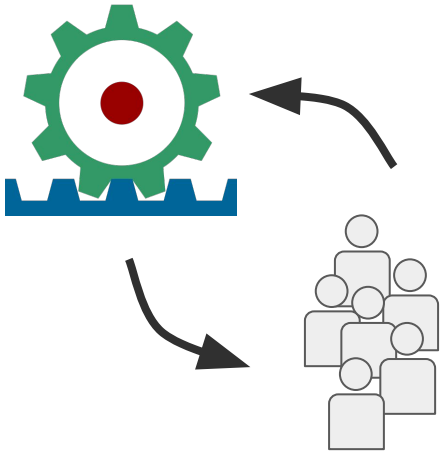
WIKIMEDIA FOUNDATION

https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2.svg
(CC-BY-SA 3.0, Wikimedia)

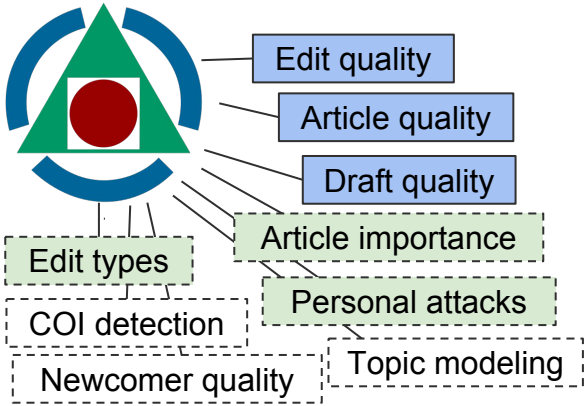
https://commons.wikimedia.org/wiki/File:Notification-icon-Wikidata-logo.svg
(CC-BY-SA 3.0, User:Jdforrester (WMF))

2017?

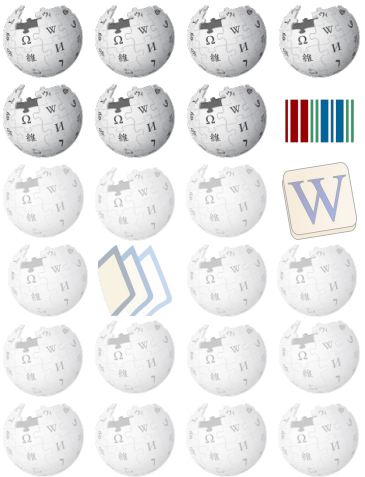
Accountability mechanisms



More models



More wikis



https://commons.wikimedia.org/wiki/File:Wiktionary_small.svg
(CC-BY-SA 3.0, Wikimedia)

https://commons.wikimedia.org/wiki/File:Wikibooks-logo.svg
(CC-BY-SA 3.0, User:Bastique, User:Ramac)

WIKIMEDIA FOUNDATION

https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2.svg
(CC-BY-SA 3.0, Wikimedia)

https://commons.wikimedia.org/wiki/File:Notification-icon-Wikidata-logo.svg
(CC-BY-SA 3.0, User:Jdforrester (WMF))

Next steps

- Obviously, we have a lot to discuss here… then,
- Which stakeholders need to be brought into the conversation?
- Identify funding sources (e.g., is this just something to work into the annual plan? Do we need to pursue a restricted grant?)
- Build team, establish direction and priorities

Q&A

Appendix

in random order

WIKIMEDIA
FOUNDATION

Staffing

8					
7					
6					
5					
4				WMDE contractor	*New staff*
3		WMDE contractor		*New staff*	2018
2		*IEG grantee* *IEG grantee*	WMDE contractor		
1	*Volunteer*	2015	2016	2017	

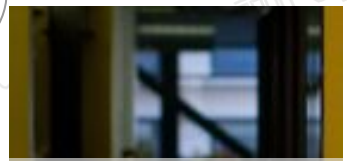
WIKIMEDIA FOUNDATION

Aaron will:
- Set the direction
- Prioritize the work and research
- Pull in external researchers and consult with users (Product & tool developers)

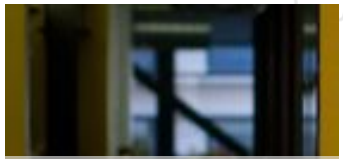
Lead Idea Architect

Engineering manager

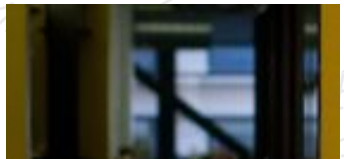
An engineering manager will handle the managerial work as well as code, so that Aaron stays focused and generative on the most important problems.



Engineer



Technical writer



CL

A network of collaborating, volunteer researchers

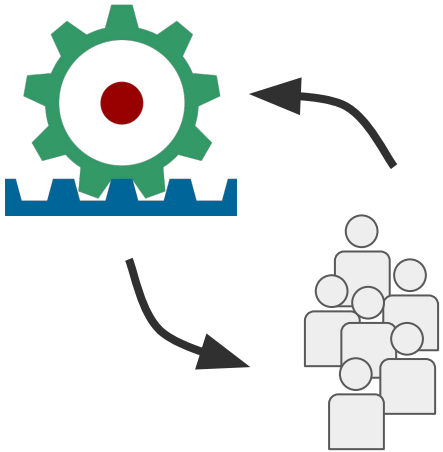
Why should Aaron lead on ideas?

- He’s the inventor.
- He knows the challenge better than anyone. He’s studied the challenge and surrounding complications for years.
- He understands the core issues: what is important, what isn't, where the opportunities are, and what is dangerous.
- He found the money to experiment and succeed under unusual resource constraints (a grant).
- Done a good job working with Ops, Research, editors, tool developers, and Product.
- He made this work with an organized network of researchers--that’s leadership!
- He cultivates and maintains researcher relationships.
- It shows others at WMF with good ideas that their efforts will be rewarded with some agency to work with others toward their vision.

Got a rough sketch of a road map?

2017?

Accountability mechanisms



WIKIMEDIA FOUNDATION

https://commons.wikimedia.org/wiki/File:Wiktionary_small.svg
(CC-BY-SA 3.0, Wikimedia)

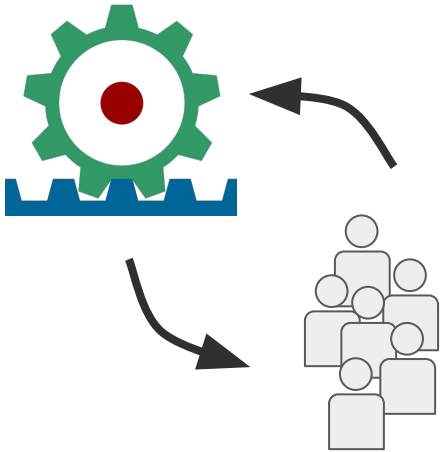
https://commons.wikimedia.org/wiki/File:Wikibooks-logo.svg
(CC-BY-SA 3.0, User:Bastique, User:Ramac)

https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2.svg
(CC-BY-SA 3.0, Wikimedia)

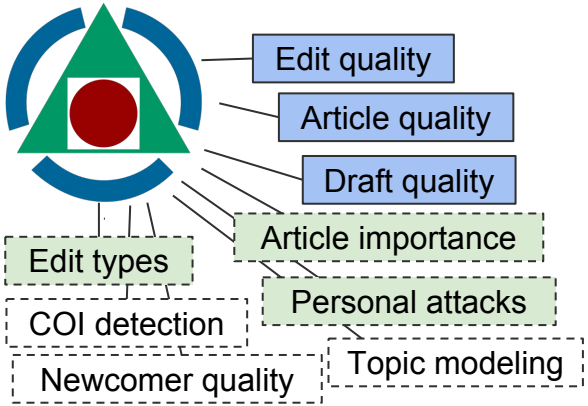
https://commons.wikimedia.org/wiki/File:Notification-icon-Wikidata-logo.svg
(CC-BY-SA 3.0, User:Jdforrester (WMF))

2017?

Accountability mechanisms



More models



https://commons.wikimedia.org/wiki/File:Wiktionary_small.svg
(CC-BY-SA 3.0, Wikimedia)

https://commons.wikimedia.org/wiki/File:Wikibooks-logo.svg
(CC-BY-SA 3.0, User:Bastique, User:Ramac)

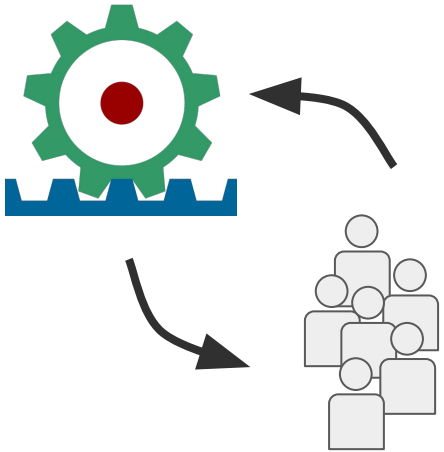
WIKIMEDIA FOUNDATION

https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2.svg
(CC-BY-SA 3.0, Wikimedia)

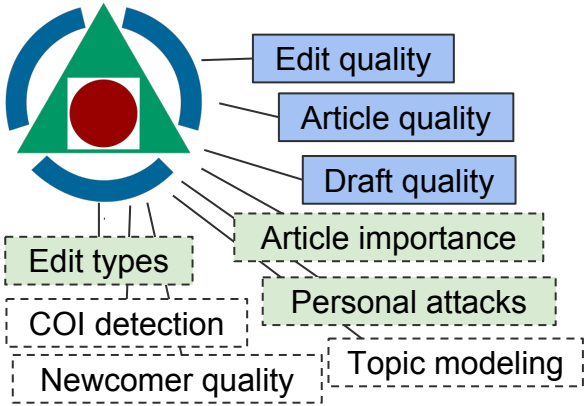
https://commons.wikimedia.org/wiki/File:Notification-icon-Wikidata-logo.svg
(CC-BY-SA 3.0, User:Jdforrester (WMF))

2017?

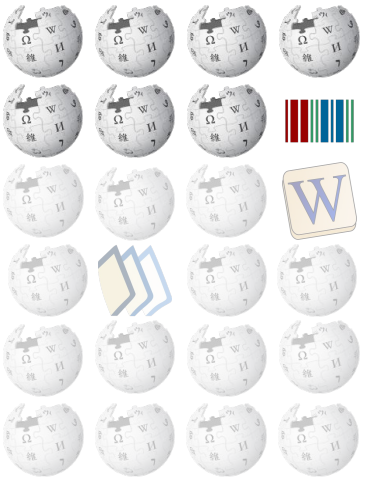
Accountability mechanisms



More models



More wikis



https://commons.wikimedia.org/wiki/File:Wiktionary_small.svg
(CC-BY-SA 3.0, Wikimedia)

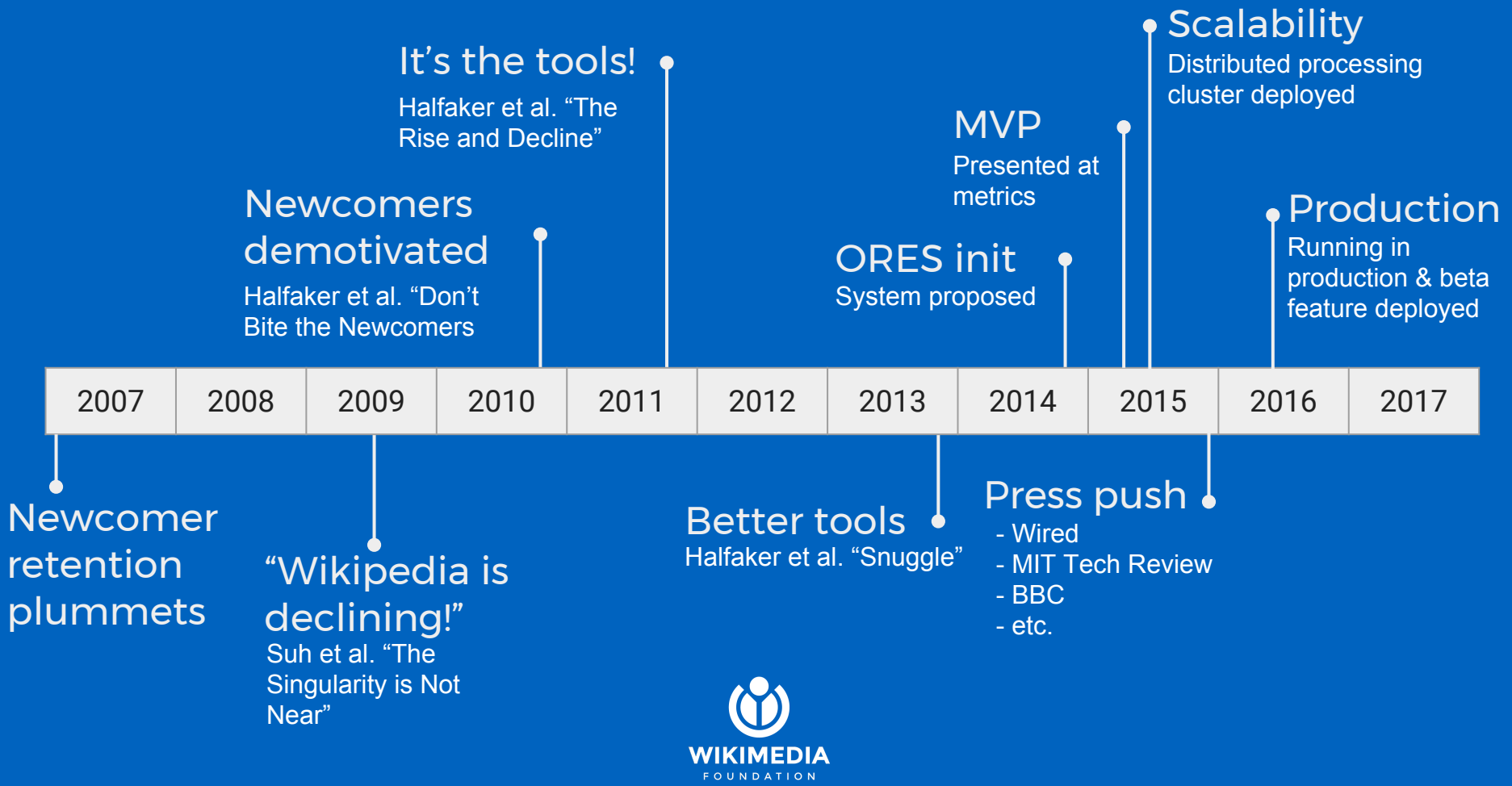
https://commons.wikimedia.org/wiki/File:Wikibooks-logo.svg
(CC-BY-SA 3.0, User:Bastique, User:Ramac)

WIKIMEDIA FOUNDATION

https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2.svg
(CC-BY-SA 3.0, Wikimedia)

https://commons.wikimedia.org/wiki/File:Notification-icon-Wikidata-logo.svg
(CC-BY-SA 3.0, User:Jdforrester (WMF))

Timeline


Scalability
Distributed processing cluster deployed
It's the tools!
Halfaker et al. "The Rise and Decline"
MVP
Presented at metrics
Production
Running in production & beta feature deployed
Newcomers demotivated
Halfaker et al. "Don't Bite the Newcomers
ORES init
System proposed
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
Newcomer retention plummets
"Wikipedia is declining!"
Suh et al. "The Singularity is Not Near"
Better tools
Halfaker et al. "Snuggle"
Press push
- Wired
- MIT Tech Review
- BBC
- etc.
WIKIMEDIA
FOUNDATION

Isn’t basic AI a solved problem?

WIKIMEDIA
FOUNDATION

“Subjective algorithms”

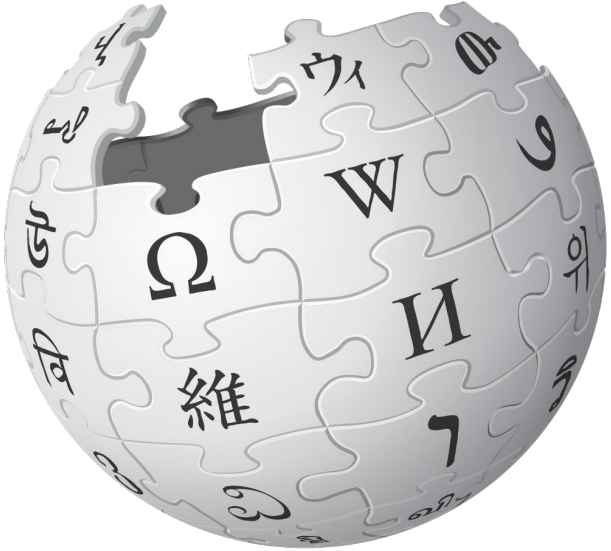
"algorithms, often aided by big data, now make decisions in subjective realms where there is **no right decision**, and no anchor with which to judge outcomes."

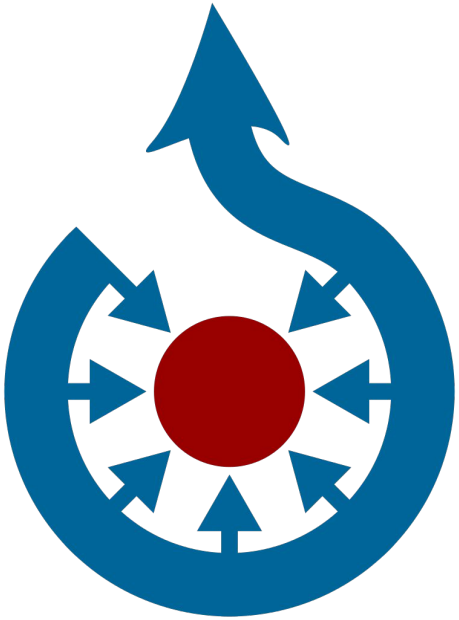
Tufekci, Z. (2015). Algorithms in our Midst: Information, Power and Choice when Software is Everywhere. CSCW (pp. 1918-1918). ACM.

What is good? relevant?

important? desirable? valuable?

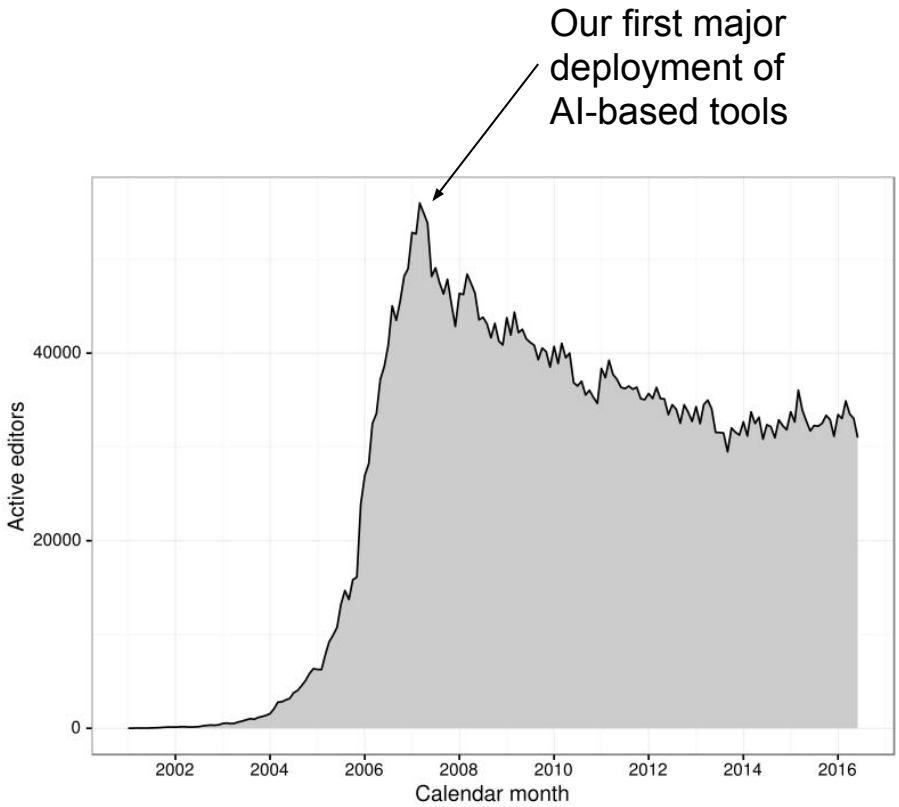
https://commons.wikimedia.org/wiki/File:Zeynep_Tufekci_2_(5805394290).jpg

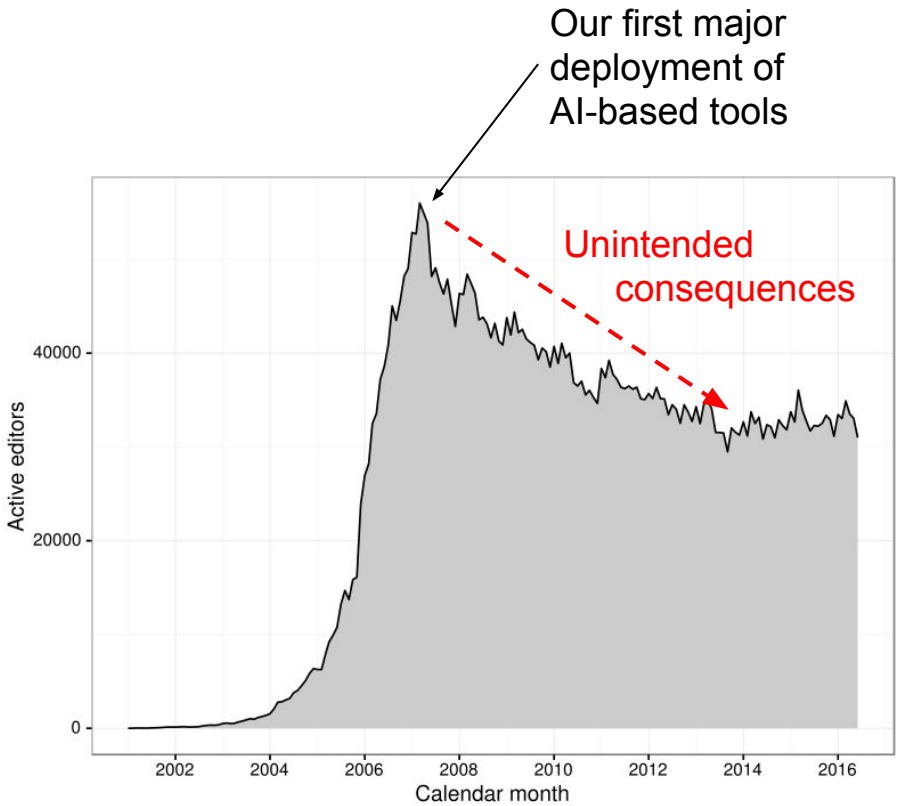




Who is allowed to participate? Who gets labeled “bad-faith”?

What types of contributions will be labeled “damaging”?


Our first major deployment of AI-based tools
Active editors
40000
20000
0
2002
2004
2006
2008
2010
2012
2014
2016
Calendar month



https://commons.wikimedia.org/wiki/File:Baby_vs._Bathwater_Annotated.JPG (PD)

1. Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, *57*(5), 664-688.
2. Halfaker, A., Geiger, R. S., & Terveen, L. G. (2014, April). Snuggle: Designing for efficient socialization and ideological critique. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 311-320). ACM.
3. Collier, B., & Bear, J. (2012, February). Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. InProceedings of the ACM 2012 conference on computer supported cooperative work (pp. 383-392). ACM.
4. VANDALISM, C. (2012). Bots and cyborgs: Wikipedia's immune system. (p. 79 - 82) IEEE Communting.

What's the bigger vision?

WIKIMEDIA
FOUNDATION

We, Wikimedia, can be a leading voice in effective and ethical AI.

We've come so far with minimal investment, let's scale it as the next step.

Cumulus clouds in fair weather.jpeg By Michael Jastremski - legacy.openphoto.net, CC BY-SA 2.0

How do you use ORES?

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

“English Wikipedia”

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

Is this edit damaging?

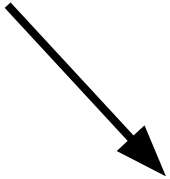
http://ores.wmflabs.org/scores/enwiki/damaging/638307884

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

In 2013 Wilson, in collaboration with [[Mikhail Baryshnikov]] and co-starring [[Willem Dafoe]], developed "[[**The Old Woman (play)**|The Old Woman]]", an adaptation of the work by the Russian author [[Daniil Kharms]]. The play premiered at MIF13, Manchester International Festival.<ref>{{Cite web|url = http://www.mif.co.uk/event/the-old-woman|title = The Old Woman Robert Wilson, Mikhail Baryshnikov, Willem Dafoe|date = 2013|accessdate = 2014-11-23|website = Manchester International Festival|publisher = Manchester International Festival|last = Jansch|first = Lucie}}</ref> Wilson wrote that he and [[Mikhail Baryshnikov|Baryshnikov]] have discussed creating a play together for years, perhaps based on a Russian text.<ref>{{Cite web|url = http://calperformances.org/learn/program_notes/2014/pn_oldwoman.pdf|title = Cal Performances University of California, Berkeley Playbill|date = 2014|accessdate = 2014-

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

In 2013 Wilson, in collaboration with [[Mikhail Baryshnikov]] and co-starring [[Willem Dafoe]], developed "[[**The Old Woman (play)**|The Old Woman]]", an adaptation of the work by the Russian author [[Daniil Kharms]]. The play premiered at MIF13, Manchester International Festival.<ref>{{Cite web|url = http://www.mif.co.uk/event/the-old-woman|title = The Old Woman Robert Wilson, Mikhail Baryshnikov, Willem Dafoe|date = 2013|accessdate = 2014-11-23|website = Manchester International Festival|publisher = Manchester International Festival|last = Jansch|first = Lucie}}</ref> Wilson wrote that he and [[Mikhail Baryshnikov|Baryshnikov]] have discussed creating a play together for years, perhaps based on a Russian text.<ref>{{Cite web|url = http://calperformances.org/learn/program_notes/2014/pn_oldwoman.pdf|title = Cal Performances University of California, Berkeley Playbill|date = 2014|accessdate = 2014-




```
"638307884": {
  "prediction": false,
  "probability": {
    "false": 0.942,
    "true": 0.058
  }
}
```

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

In 2013 Wilson, in collaboration with [[Mikhail Baryshnikov]] and co-starring [[Willem Dafoe]], developed "[[**The Old Woman (play)**|The Old Woman]]", an adaptation of the work by the Russian author [[Daniil Kharms]]. The play premiered at MIF13, Manchester International Festival.<ref>{{Cite web|url = http://www.mif.co.uk/event/the-old-woman|title = The Old Woman Robert Wilson, Mikhail Baryshnikov, Willem Dafoe|date = 2013|accessdate = 2014-11-23|website = Manchester International Festival|publisher = Manchester International Festival|last = Jansch|first = Lucie}}</ref> Wilson wrote that he and [[Mikhail Baryshnikov|Baryshnikov]] have discussed creating a play together for years, perhaps based on a Russian text.<ref>{{Cite web|url = http://calperformances.org/learn/program_notes/2014/pn_oldwoman.pdf|title = Cal Performances University of California, Berkeley Playbill|date = 2014|accessdate = 2014-


```
"638307884": {
  "prediction": false,
  "probability": {
    "false": 0.942,
    "true": 0.058
  }
}
```


http://ores.wmflabs.org/scores/enwiki/damaging/642215410

\-

"Intelligence" is hard to define, whether in dogs, other animals, or humans. The ability to learn quickly might be taken as a sign of [[intelligence (trait)|intelligence]], but such evidence must be interpreted with care, because learning speed may be affected by such things as the effectiveness of the rewards used in training or the motivation or activity level of the dog. For example, some breeds, such as [[Siberian Husky|Siberian Huskies]], are said to be not particularly rewarded by pleasing their owners, but quickly learn to escape from yards or catch small animals, often using ingenious ways of doing both.
<ref>http://siberianhuskycentral.com/siberian-husky-dogs/are-siberian-huskies-good-dogs</ref>

\+

"Intelligence" is hard to define, whether in dogs, other animals, or humans. The ability to learn quickly might be taken as a sign of [[intelligence (trait)|intelligence]], but such evidence must be interpreted with care, because learning speed may be affected by such things as the effectiveness of the rewards used in training or the motivation or activity level of the dog. For example, some breeds, such as [[Siberian Husky|Siberian Huskies]], are said to be not particularly rewarded by pleasing their owners, but quickly learn to escape from yards or catch small animals, often using ingenious ways of doing both. **LLAMAS GROW ON TREES**

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

In 2013 Wilson, in collaboration with [[Mikhail Baryshnikov]] and co-starring [[Willem Dafoe]], developed "[[**The Old Woman (play)**|The Old Woman]]", an adaptation of the work by the Russian author [[Daniil Kharms]]. The play premiered at MIF13, Manchester International Festival.<ref>{{Cite web|url = http://www.mif.co.uk/event/the-old-woman|title = The Old Woman Robert Wilson, Mikhail Baryshnikov, Willem Dafoe|date = 2013|accessdate = 2014-11-23|website = Manchester International Festival|publisher = Manchester International Festival|last = Jansch|first = Lucie}}</ref> Wilson wrote that he and [[Mikhail Baryshnikov|Baryshnikov]] have discussed creating a play together for years, perhaps based on a Russian text.<ref>{{Cite web|url = http://calperformances.org/learn/program_notes/2014/pn_oldwoman.pdf|title = Cal Performances University of California, Berkeley Playbill|date = 2014|accessdate = 2014-


```
"638307884": {
  "prediction": false,
  "probability": {
    "false": 0.942,
    "true": 0.058
  }
}
```


http://ores.wmflabs.org/scores/enwiki/damaging/642215410

\- "Intelligence" is hard to define, whether in dogs, other animals, or humans. The ability to learn quickly might be taken as a sign of [[intelligence (trait)|intelligence]], but such evidence must be interpreted with care, because learning speed may be affected by such things as the effectiveness of the rewards used in training or the motivation or activity level of the dog. For example, some breeds, such as [[Siberian Husky|Siberian Huskies]], are said to be not particularly rewarded by pleasing their owners, but quickly learn to escape from yards or catch small animals, often using ingenious ways of doing both. **<ref>http://siberianhuskycentral.com/siberian-husky-dogs/are-siberian-huskies-good-dogs</ref>**

\+ "Intelligence" is hard to define, whether in dogs, other animals, or humans. The ability to learn quickly might be taken as a sign of [[intelligence (trait)|intelligence]], but such evidence must be interpreted with care, because learning speed may be affected by such things as the effectiveness of the rewards used in training or the motivation or activity level of the dog. For example, some breeds, such as [[Siberian Husky|Siberian Huskies]], are said to be not particularly rewarded by pleasing their owners, but quickly learn to escape from yards or catch small animals, often using ingenious ways of doing both. **LLAMAS GROW ON TREES**


```
"642215410": {
  "prediction": true,
  "probability": {
    "false": 0.080,
    "true": 0.920
  }
}
```

Copyright © 2017, Aaron Halfaker and the Wikimedia Foundation.
This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International license.