

ORES - Machine learning on Wikipedia

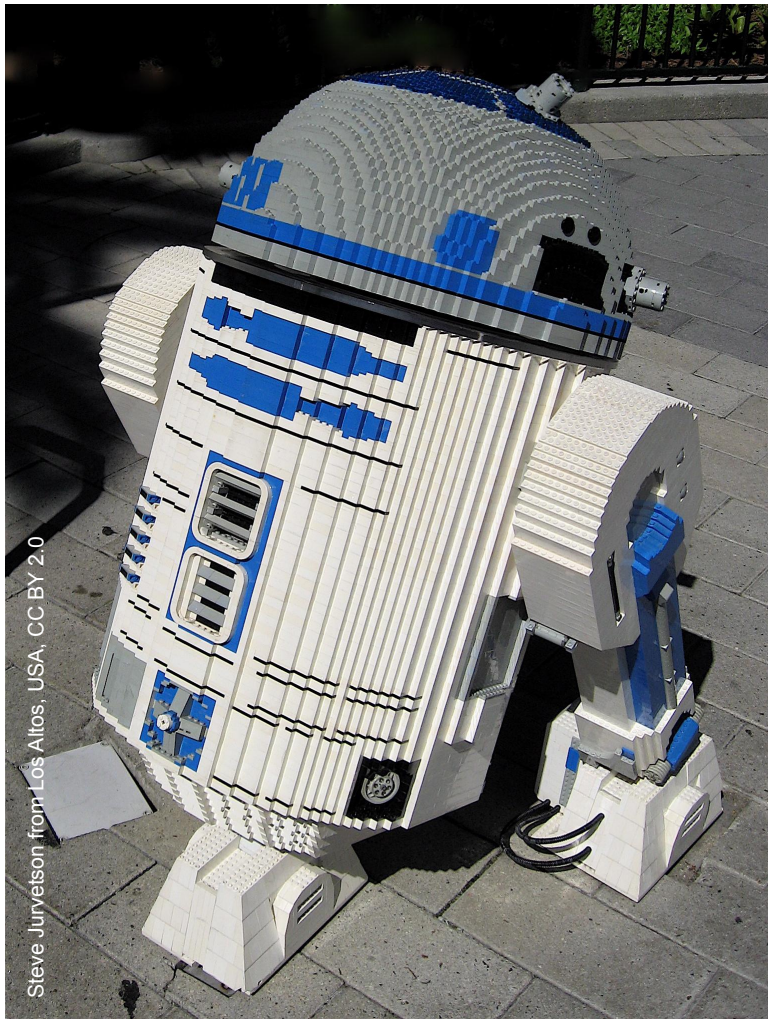
Hi, ik ben Ciell: ik ben begonnen als schrijver op Wikipedia in 2006 en sinds mei 2019 betrokken bij het project rondom de inzet van AI op de Nederlandstalige Wikipedia.

Naast AI houd ik mij onder andere bezig met onderhoudstaken op de Nederlandstalige Wikipedia en Wikimedia Commons, en zet ik mij in om de diverse 'gaps' in informatie zichtbaar te maken en waar mogelijk te verkleinen.



Gebruiker: Ciell

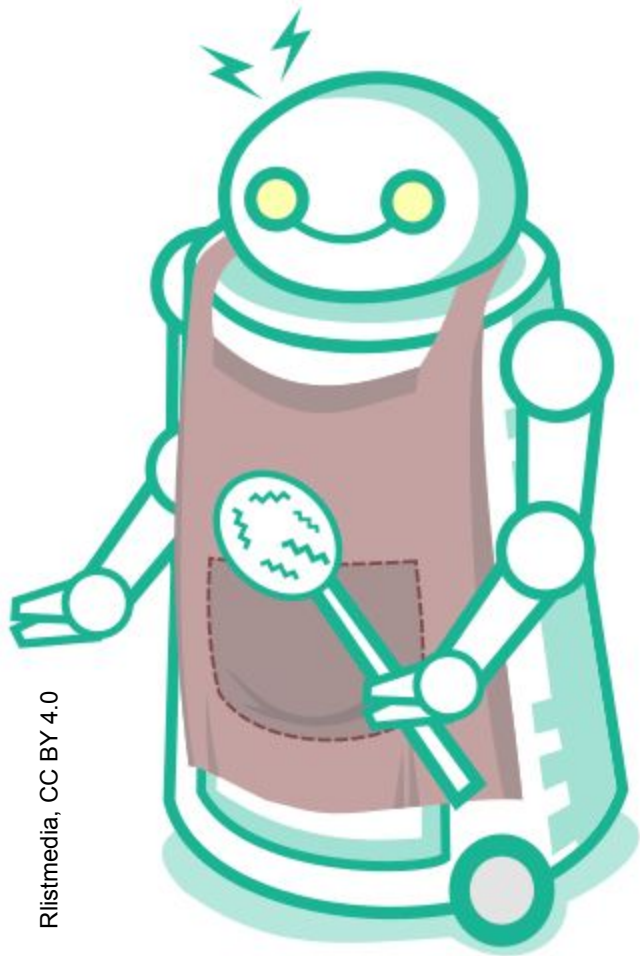
[ciell.wikipedia@gmail.com](mailto:ciell.wikipedia@gmail.com)



## **Bot:**

afkorting voor robot, die herhaalde (en vaak saaie) taken snel uit kan voeren.

- Op de Nederlandstalige Wikipedia zijn 47 accounts met bot-rechten. Dit betekent dat de bestuurder van de bot naar onze inschatting voldoende ervaring en kennis heeft om in een hoog tempo op de achtergrond geautomatiseerd onderhoudstaken uit te voeren.



Versie van 11 nov 2021 01:08 (bewerken | **ongedaan maken** | bedanken)

**MRbazzeman** (overleg | bijdragen | blokkeren)

(→*Huidige rolverdeling*) (weergeven/verbergen)

(Label: Visuele tekstverwerker)

Nieuwere bewerking - **[Als gecontroleerd markeren]**

#### Regel 289:

| [[Joep Onderdelinden]]

| 2013-2017, 2020-heden

+ | -

+ | Chocopiet

+ | [[Rosa da Silva]]

+ | 2021-heden

| -

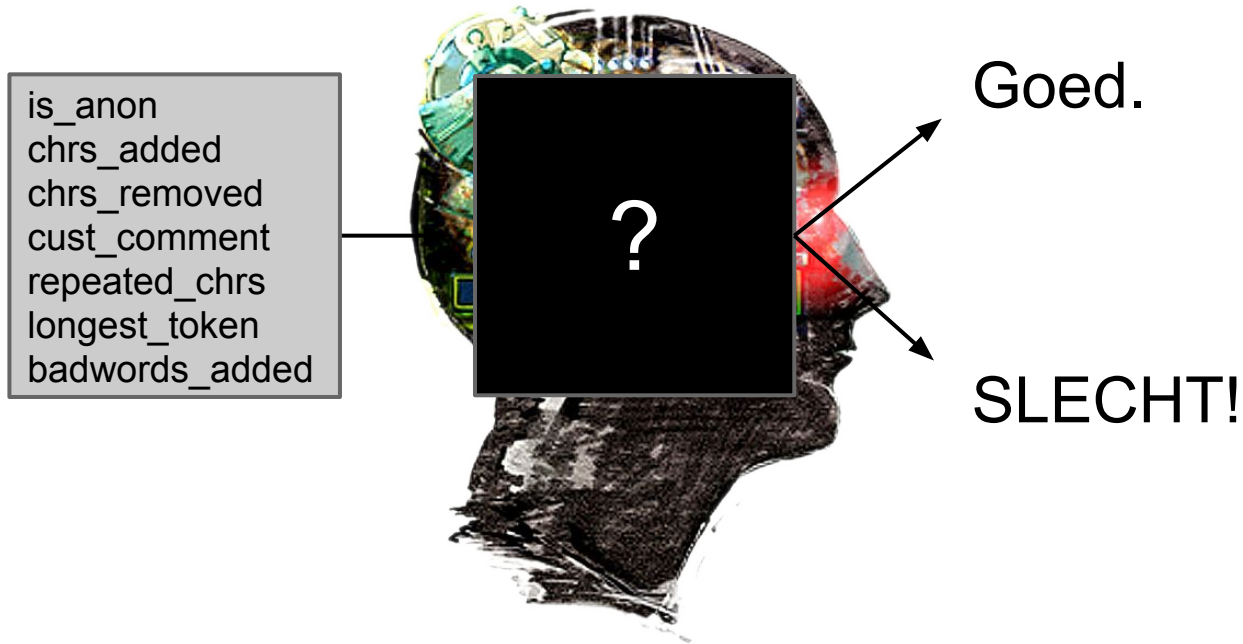
| Piet

# De geautomatiseerde beoordeling

is\_anon  
chrs\_added  
chrs\_removed  
cust\_comment  
repeated\_chrs  
longest\_token  
badwords\_added



# De geautomatiseerde beoordeling

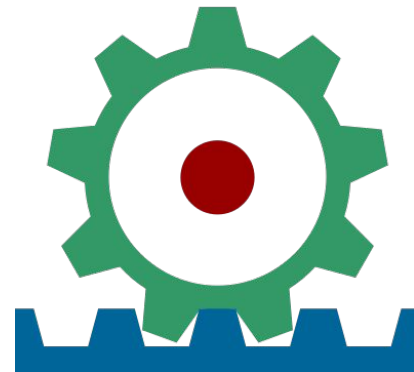


# ORES

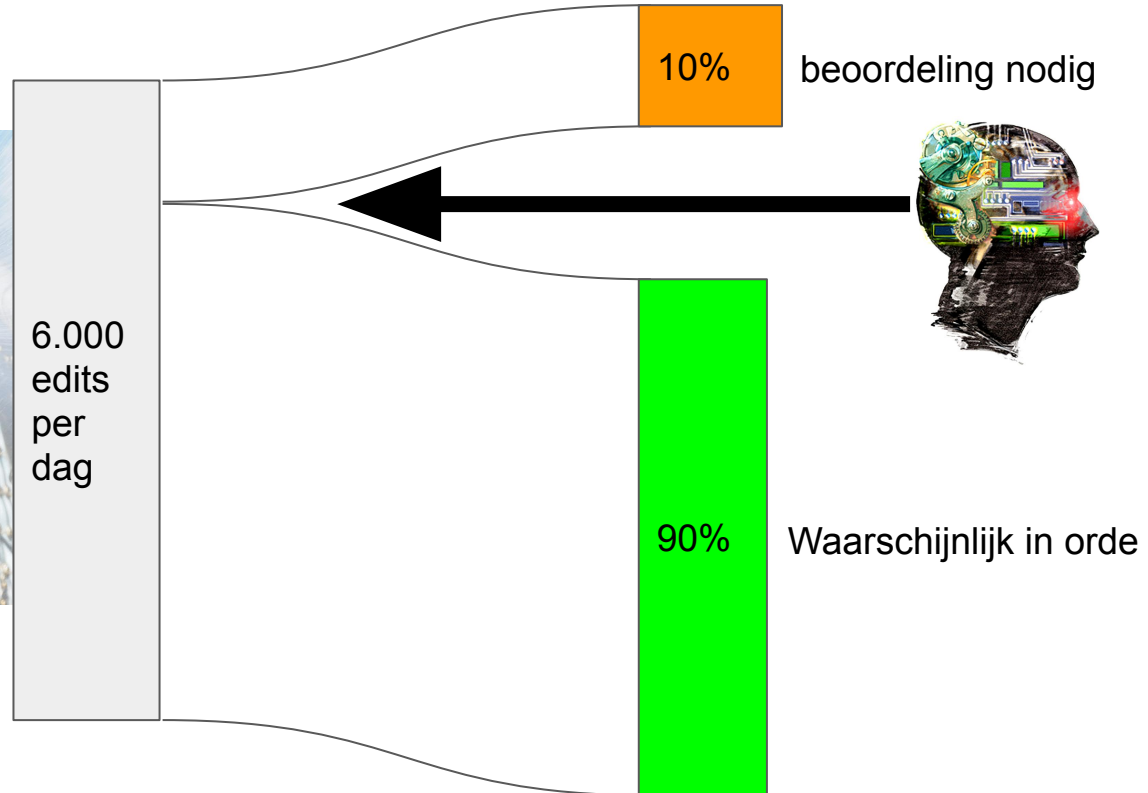
- AI hulp voor de menselijke controleurs van bewerkingen op Wikipedia



MC2 Joshua R. Nistas/U.S. Navy - CC-0  
[https://commons.wikimedia.org/wiki/File:Flickr\\_-\\_Official\\_U.S.\\_Navy\\_Imagery\\_-\\_Sailor%27s\\_daughter\\_operates\\_a\\_fire\\_hose\\_with\\_crew\\_member\\_assistance.jpg](https://commons.wikimedia.org/wiki/File:Flickr_-_Official_U.S._Navy_Imagery_-_Sailor%27s_daughter_operates_a_fire_hose_with_crew_member_assistance.jpg)



# Vandalisme tegengaan

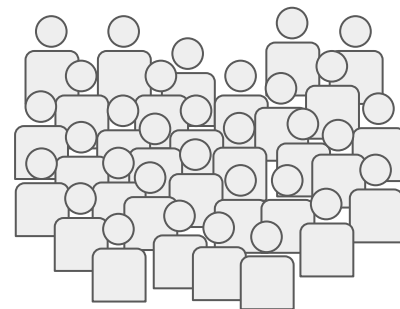




**Zonder ORES:** Beoordelen van 6000 edits per dag...

166 uur

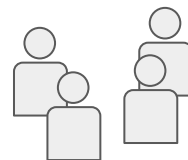
(30 mensen x 5,5 uur per dag)



**Met ORES:** Beoordelen van 600 edits per dag...

16,6 uur

(6 mensen x 2,5 uur per dag)



# Recente wijzigingen

[Andere controlehulpmiddelen](#) ▾

 **Opgeslagen filters** ▾

## Actieve filters

[Verbergen](#)

- Menselijk (geen bot) ✕
- Paginabewerkingen ✕
- Nieuwe pagina's ✕
- Geregistreerde acties ✕
- Heeft zeer waarschijnlijk problemen ✕
- Heeft waarschijnlijk problemen ✕
- Heeft wellicht problemen ✕
- Zeer mogelijk goed ✕
- Niet-geregistreerd ✕
- Nieuwkomers ✕

 **Naamruimten**  **Labels**

## Filters

 **Resultaten markeren**

 **1.000 wijzigingen, 30 dagen** ▾

### Kwaliteitsvoorspellingen voor bijdragen [Hoe werkt dit?](#)

Lijst met afkortingen:[\[Uitvouwen\]](#)

- Zeer mogelijk goed** ● ▾  
Highly accurate at finding almost all problem-free edits.
- Heeft wellicht problemen** ● ▾  
Finds most flawed or damaging edits but with lower accuracy.
- Heeft waarschijnlijk problemen** ● ▾  
With medium accuracy, finds an intermediate fraction of problem edits.
- Heeft zeer waarschijnlijk problemen** ● ▾  
Very highly accurate at finding the most obviously flawed or damaging edits.

### Gebruikersintentievoorspellingen [Hoe werkt dit?](#)

- Zeer mogelijk goede wil** ● ▾

- blokkeren) (Label: Visuele tekstverwerker) [terugdraaien]
- bijdragen | blokkeren) (Pagina voor Scouts 't Hoekske aangemaakt) (Label:
- eg | blokkeren) (→Privé) (Labels: Bewerking via mobiel, Bewerking via mobiele
- g | bijdragen | blokkeren) [terugdraaien]
- erleg | blokkeren) (→Seizoen 5) (Label: Visuele tekstverwerker) [terugdraaien]
- blokkeren) (→Derde generatie (2021)) (Label: Visuele tekstverwerker)
- overleg | blokkeren) (Aanpassing terminologie) (Labels: Bewerking via mobiel,
- :6227 (overleg | blokkeren) (Toevoeging waarheid) (Labels: Bewerking via mobiel,
- bijdragen | blokkeren)

- [ reageer ] [ B Margaretha van Hohenzollern-Neurenberg ] - informatie over eerste testament en grafmonument toegevoegd. HRvO (overleg) 1 jan 2021 15:33 (CET) [ reageer ]
- [ reageer ] [ C Luminize ] - Kroatische rockgroep die uit vrouwen bestaat, stond op TBP - hiro the club is open 6 jan 2021 21:29 (CET) [ reageer ]
- [ reageer ] [ A Belle van Zuylen ] - citaat in het Frans met Nederlandse vertaling - Hansmuller (overleg) 18 jan 2021 22:56 (CET) [ reageer ]
- [ reageer ] [ C Sofia Dragt ] - Nederlandse singer-songwriter, teruggehaald na TBP-sessie - hiro the club is open 25 jan 2021 23:13 (CET) [ reageer ]
- [ reageer ] [ A Belle van Zuylen ] - verwijzing naar vertalingen in het Engels - Hansmuller (overleg) 27 jan 2021 07:45 (CET) [ reageer ]
- [ reageer ] [ D Kristen Schaal ] - Nederlands-Amerikaanse cabaretière enzovoorts, fout afkomst eruit, aangevuld, categorie - Hansmuller (overleg) 3 feb 2021 15:01 (CET) [ reageer ]
- [ reageer ] [ A Belle van Zuylen ] - citaat in het Frans met Nederlandse vertaling - Help mee tegen de censuur van dit citaat. We laten Belle zelf aan het woord. Bij andere schrijvers citeren we ook. Dus kom op voor Belle en behoud haar citaat, dank, Hansmuller (overleg) 3 feb 2021 15:40 (CET) [ reageer ]
- [ reageer ] [ D Selma de Mink ] - Nederlands sterrenkundige, biografie samengevat, aangevuld Hansmuller (overleg) 9 feb 2021 21:17 (CET) [ reageer ]
- [ reageer ] [ B Nina Berberova ] - Russisch dichter en schrijver, aanvulling plus toevoeging Nederlandse vertalingen van haar werk Sjahrazaad (overleg) 14 feb 2021 13:38 (CET) [ reageer ]
- [ reageer ] [ C Maria Versfelt ] - Nederlands courtisane. Informatie uitgebreid Mathijsloo (overleg) 15 feb 2021 10:55 (CET) [ reageer ]
- [ reageer ] [ B Margaretha van Hohenzollern-Neurenberg ] - informatie over muurschildering bij grafmonument toegevoegd Royalty & Nassau Expert (overleg) 6 mrt 2021 22:09 (CET) [ reageer ]
- [ reageer ] [ D Williamina Fleming ] - vroege leven en carrière erbij geplaatst (beginnetje af). algontoverleg 10 mrt 2021 00:26 (CET) [ reageer ]
- [ reageer ] [ D Heleen Mees ] - referenties, Hillary Clinton. Hansmuller (overleg) 16 mrt 2021 13:14 (CET) [ reageer ]
- [ reageer ] [ D Jane Turner ] - publicaties, links. Hansmuller (overleg) 16 mrt 2021 14:25 (CET) [ reageer ]
- [ reageer ] [ C Angelique Van Ombergen ] - noodnummer 1722 - Fiene-Fleur (overleg) 25 mrt 2021 14:26 (CET) [ reageer ]
- [ reageer ] [ D Maria Anna van Portugal (1861-1942) ], groothertogin(-regentes) van Luxemburg, op 28 maart korte paragraaf over haar regentschap toegevoegd - Royalty & Nassau Expert (overleg) 2 apr 2021 16:46 (CEST) [ reageer ]
- [ reageer ] [ B Jorien Wuite ], lid Tweede Kamer - Kattiel (overleg) 3 apr 2021 15:52 (CEST) [ reageer ]
- [ reageer ] [ C Lionstorm ] - Nederlands rapduo - hiro the club is open 28 apr 2021 17:18 (CEST) maar credits voor M.L.Wits die het artikel aanmaakte. [ reageer ]
- [ reageer ] [ B The Mary Tyler Moore Show ] - Amerikaanse tv-serie - Belang als product van de tweede feministische golf beschreven. hiro the club is open 30 mei 2021 04:30 (CEST) [ reageer ]
- [ reageer ] [ D Olena Tokar ] - Oekraïens sopraan. Was verwijderd op grond van crosswiki spam; teruggehaald en uitgebreid. hiro the club is open 9 jun 2021 10:26 (CEST) [ reageer ]
- [ reageer ] [ C Ietje Paalman-de Miranda ], door Gmekli in het kader van Wikipedia evenement EWM-NL
- [ reageer ] [ B Ingrid Daubechies ], door Gmekli in het kader van Wikipedia evenement EWM-NL
- [ reageer ] [ C Tatjana Pavlovna Ehrenfest ], door JanetHaneg in het kader van Wikipedia evenement EWM-NL
- [ reageer ] [ B Mauritia Eleonora van Portugal ] - prinses van Portugal en door huwelijk gravin van Nassau-Siegen. Artikel geheel herschreven aan de hand van bronnen. Roelof Hendrickx (overleg) 22 aug 2021 17:52 (CEST) [ reageer ]
- [ reageer ] [ D Jessica Harper ] - als nieuwe Wikipediiaan (LievAnn) een beginnend artikel verder aangevuld

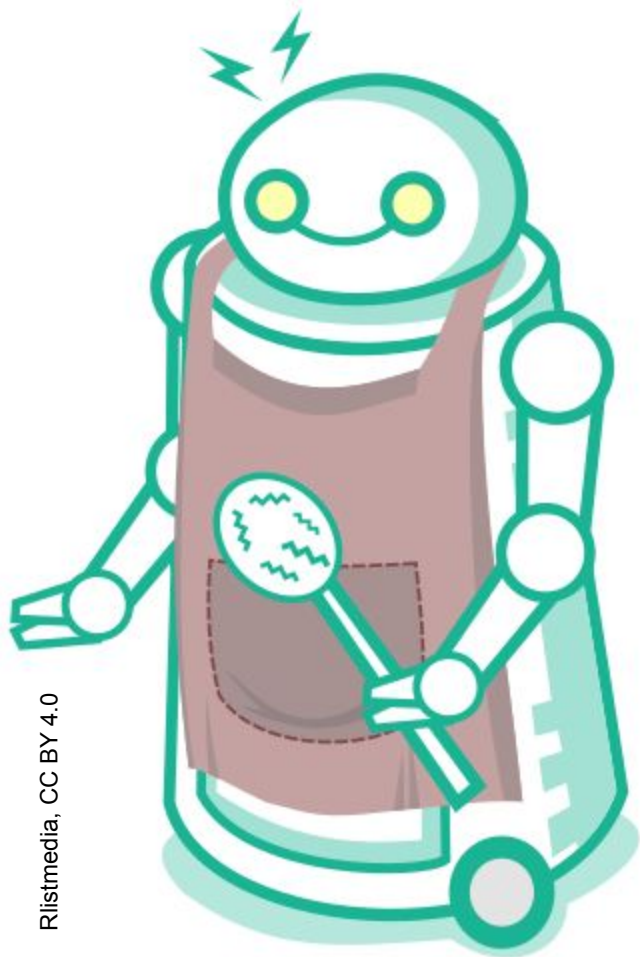


Martha Cunz



Helen Dahm





donderdag 11/11  
#60327829

```
nlwiki:
  models:
    articlequality:
      version: "0.8.0"
    damaging:
      version: "0.5.1"
    goodfaith:
      version: "0.5.1"
  scores:
    60327829:
      articlequality:
        score:
          prediction: "B"
          probability:
            A: 0.027439872609634524
            B: 0.8881107231467956
            C: 0.022009328951280258
            D: 0.04928312624295924
            E: 0.013156949049330345
      damaging:
        score:
          prediction: false
          probability:
            false: 0.9852819864934785
            true: 0.014718013506521459
      goodfaith:
        score:
          prediction: true
          probability:
            false: 0.0016215639561024675
            true: 0.9983784360438975
```

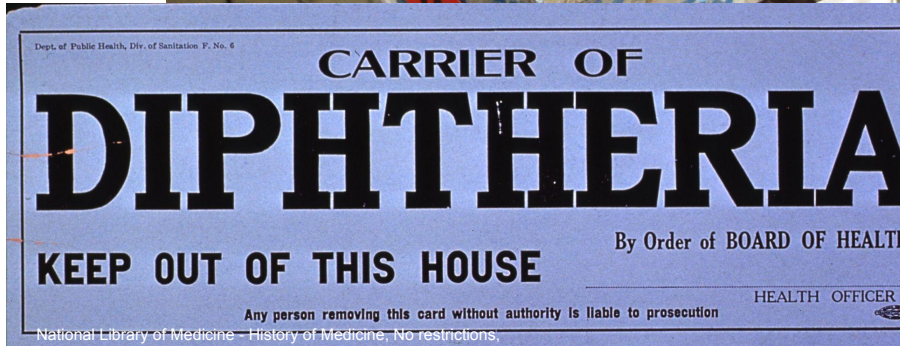


LucaLX, CC BY 3.0



Dorian Wallender from Lake Havasu City, Arizona, USA, CC BY-SA 2.0

Michelle Lai Lai from Tokyo, Japan, CC BY-SA 4.0



Niranjan Arminius, CC BY-SA 2.0



Diego Torres Silvestre from Sao Paulo, Brazil, CC BY 2.0

# Model cards: inzicht

## Enwiki Good Faith Model Card [\[ edit \]](#)



This model card was created and written entirely by the [Algorithmic Accountability Bot](#) account. The bot is operated by [Hal Friedman](#). It regularly creates and updates pages about the provenance and statistical performance of machine learning models and datasets owned by the Wikimedia Foundation.

*Note: Any and all edits to this page will be overwritten the next time it is updated. Please put all questions and discussion of this algorithmic component in the talk page, or else [contact Hal](#) or the [v ML team directly](#).*



This is a **Tier 1** model card. That means it is generated by testing the model against a more recent, indepe test set, which provides the deepest possible dive into how the model is working on MediaWiki right now.

This tier also includes:

- Detailed statistics of model performance at training time
- Full description of model architecture
- An in-depth explanation of the model rationale, owners, creators, provenance, etc.

## Qualitative Analysis [\[ edit \]](#)

### What is the motivation behind creating this model? [\[ edit \]](#)

To prioritize review of potentially damaging edits or vandalism. The model provides a guess at whether or not a given revision is in good faith, ar probabilities to serve as a measure of its confidence level.

### Who created this model? [\[ edit \]](#)

[Aaron Halfaker](#) [@](#)(aaron.halfaker@gmail.com) and [Amir Sarabadani](#) [@](#)(amir.sarabadani@wikimedia.de).

### Who currently owns/is responsible for this model? [\[ edit \]](#)

[WMF Machine Learning Team](#) [@](#)(ml@wikimediafoundation.org)

### Who are the intended users of this model? [\[ edit \]](#)

English Wikipedia uses the model as a service for facilitating efficient edit reviews. On an individual basis, anyone can submit a properly-formatt given revision and get back the result of this model.

### What should this model be used for? [\[ edit \]](#)

This model should be used for prioritizing the review and potential reversion of damaging edits and vandalism on English Wikipedia.

### What should this model not be used for? [\[ edit \]](#)

This model should not be used as an ultimate arbiter of whether or not an edit ought to be considered good faith. It should not be used for any other English- language wiki besides English Wikipedia, and shouldn't be used for other languages.

### What community approval processes has this model gone through? [\[ edit \]](#)

English Wikipedia decided (note: don't know where/when this decision was made, would love to find a link to that discussion) to use this model. Over time, the model has been validated through use in the community. The links below are just an example to show what this product might look like.

### Dates of consideration forums [\[ edit \]](#)

- [2021-09-07](#) [@](#)

### What internal or external changes could make this model deprecated or no longer usable? [\[ edit \]](#)

- Data drift means training data for the model is no longer usable.
- Doesn't meet desired performance metrics in production.
- English Wikipedia community decides to not use this model anymore.

### How should this model be licensed? [\[ edit \]](#)

[Creative Commons Attribution ShareAlike 3.0](#)

### If this model is retrained, can we see how it has changed over time? [\[ edit \]](#)

To my knowledge, this model has not been retrained over time — it still uses the original dataset from 2014-2015.

### How does this model mitigate data drift? [\[ edit \]](#)

This model does not mitigate data drift.

### Which service(s) rely on this model? [\[ edit \]](#)

This model is one of many models that powers ORES, the Wikimedia Foundation's machine machine learning API.

Learn more about ORES [here](#) [@](#)

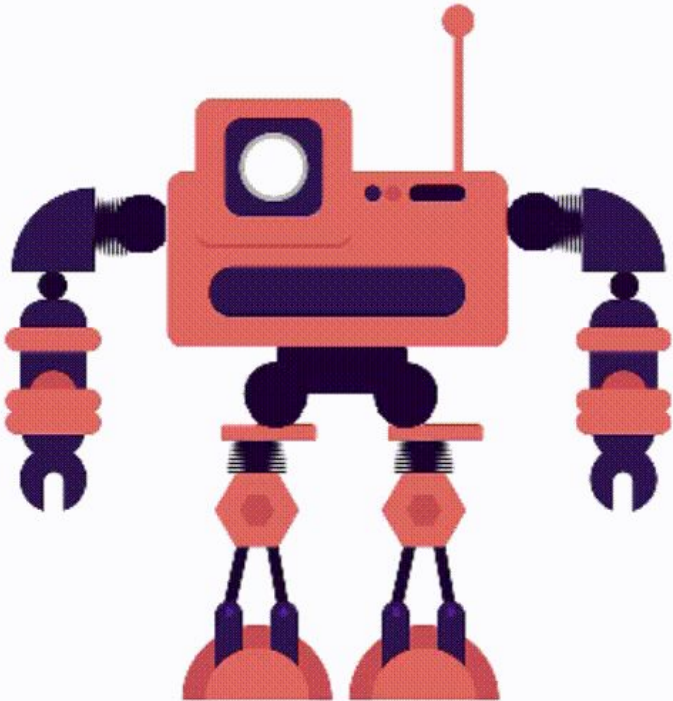
### Which dataset(s) does this model rely on? [\[ edit \]](#)

This model was trained using hand-labeled training data from 2014-2015. It was tested on a small sample of data from a later hand-labeling campaign from 2015-2016.

Train dataset is available for download [here](#) [@](#)

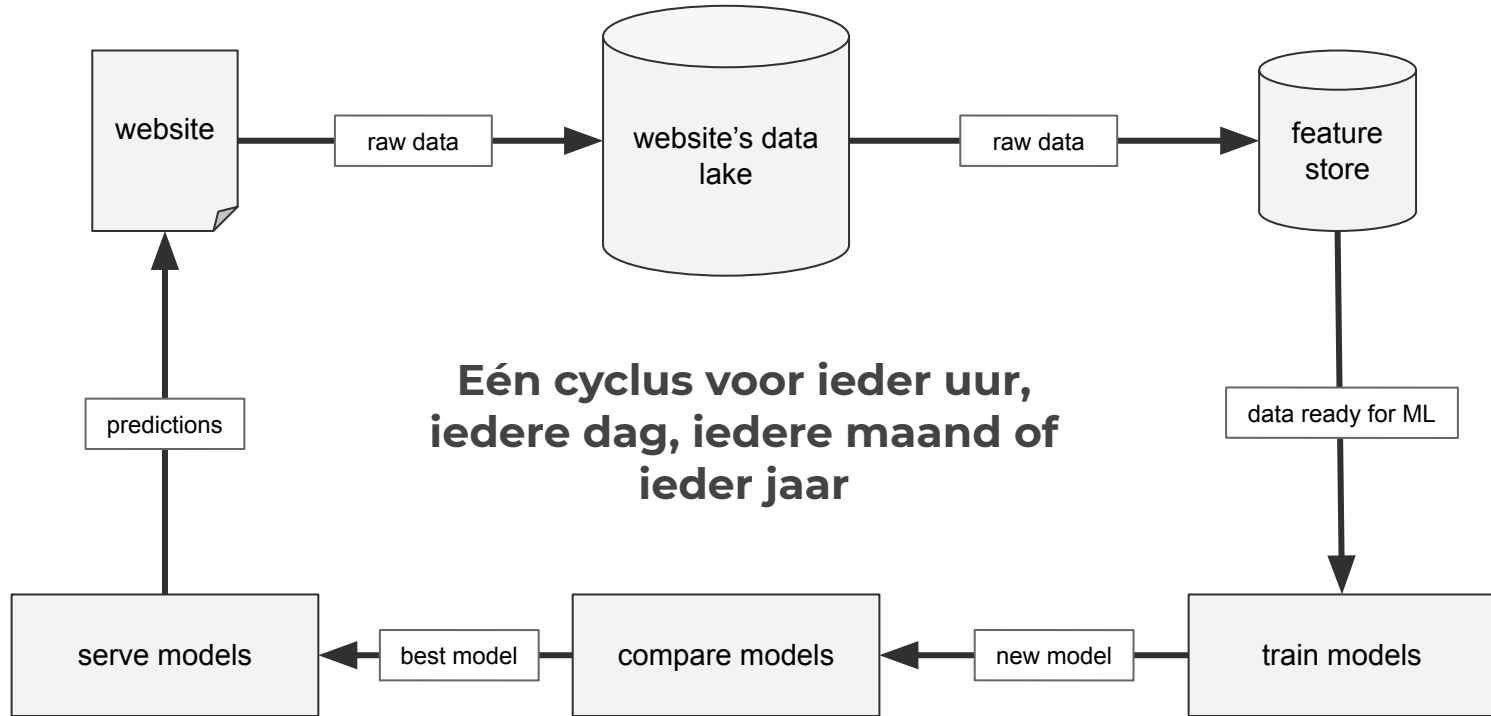
Test dataset is available for download [here](#) [@](#)

# Toekomst



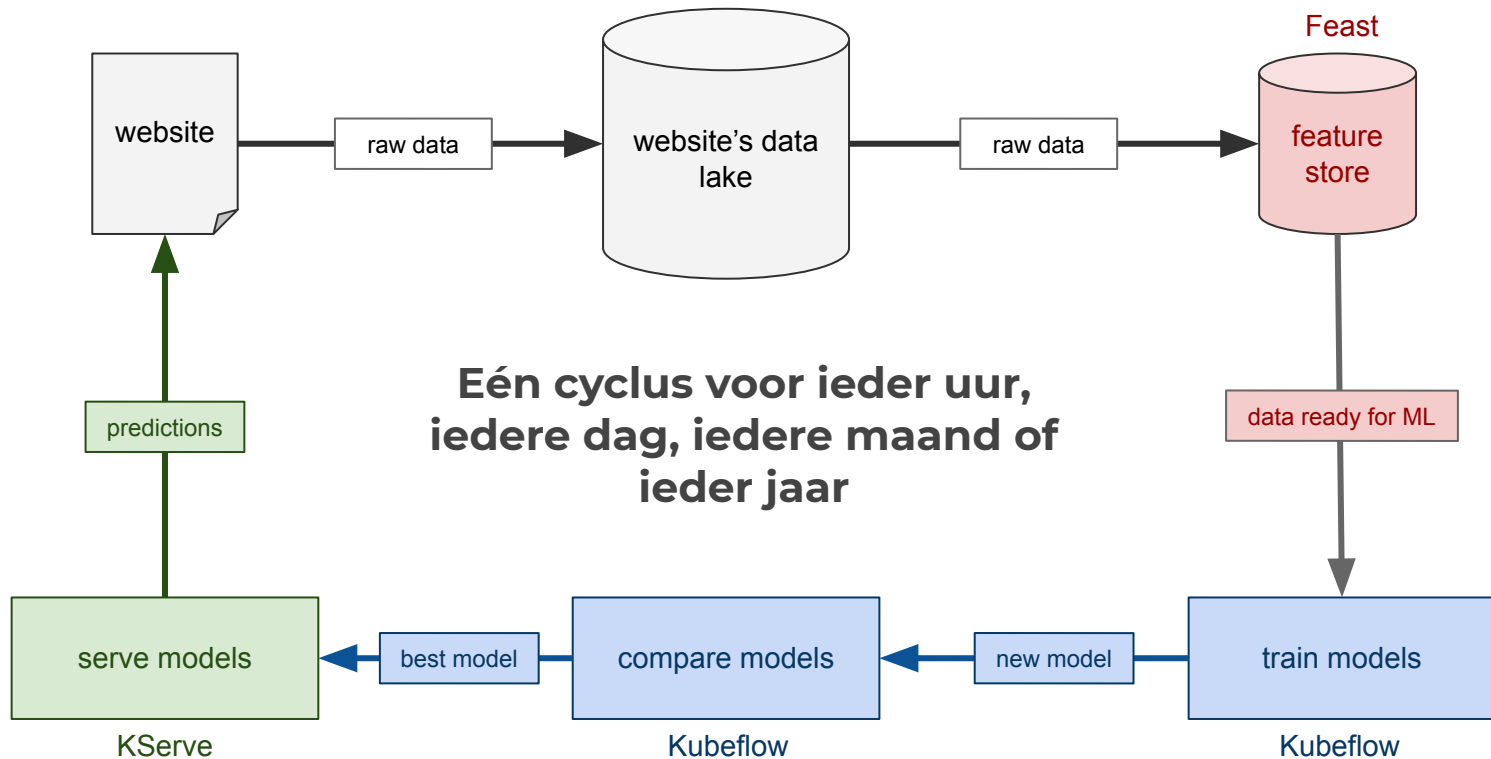
- nieuwe infrastructuur: Lift Wing
- andere server
- andere api
- weg van de statische modellen naar een meer cyclische beoordeling
- gemeenschap krijgt mogelijkheid direct input te leveren op bestaande modellen

# Huidige infrastructuur: MLOps 101



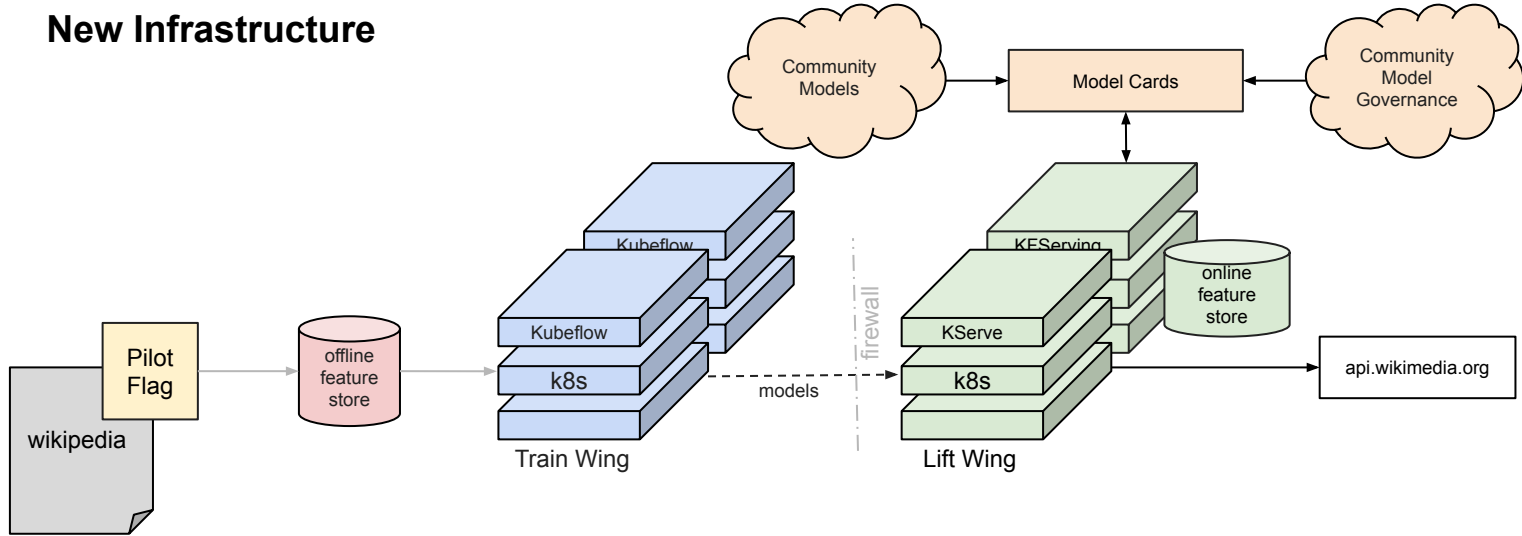


# Lift Wing: snellere voorspellingen door minder serverload



# Einddoel: een nieuwe, moderne infrastructuur

## New Infrastructure



## ORES Infrastructure





Thank you!  
Come back  
to see us again  
soon!

Visitor7, CC BY-SA 3.0 , via Wikimedia Commons

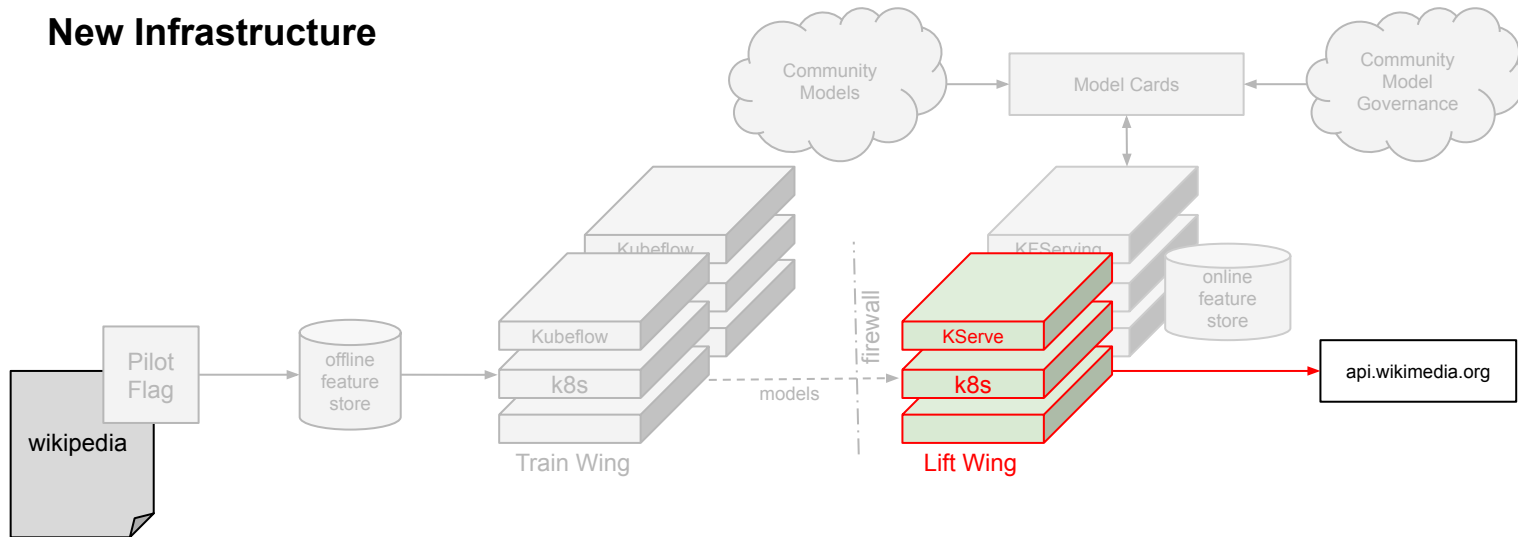
### Handige links:

- Vacature ML engineer op [greenhouse](#)
- Live uitzending iedere 2 weken op donderdag 18.30 via [Twitch kanaal WikimediaML](#)
- [Wikimedia ML op Youtube](#)
- Machine Learning pagina [op Mediawiki](#)
  
- ML/ORES op de Nederlandstalige Wikipedia: [Wikipedia:ORES](#)
- Huidige ORES api zoals voor artikelkwaliteit is beschikbaar via <https://ores.wikimedia.org/scores/nlwiki/><versie nummer>

# Q3 - FY20/21 tot en met Q2 FY21/22

1. Build an MVP model serving system (Lift Wing)
2. Connect to API Gateway
3. Migrate existing machine learning models to Lift Wing, including models hosted on ORES
4. Create Model Card Proof Of Concepts
5. Hire third MLE

## New Infrastructure



## ORES Infrastructure

