AI and Participation

Pablo Aragón

paragon@wikimedia.org

Digital Humanism Summit on
AI and Democratic Sustainability

Vienna
July 3-4, 2023

Hello, world!

Research Scientist at the Wikimedia Foundation focused on on content integrity and resilience to disinformation in Wikipedia and other free knowledge projects.

Adjunct Professor at Universitat Pompeu Fabra for the master course in Web Intelligence.

Collaborator of public institutions for projects on data, technology, and citizen participation. Member of Decidim, the web platform for participatory democracy.

Former researcher at Barcelona Media, Eurecat - Technology Centre of Catalonia, and the Oxford Internet Institute – University of Oxford for projects on computational social science, social media, digital humanities, and big data.

Hello, world!

Research Scientist at the Wikimedia Foundation focused on on content integrity and resilience to disinformation in Wikipedia and other free knowledge projects.

Adjunct Professor at Universitat Pompeu Fabra for the master course in Web Intelligence.

Collaborator of public institutions for projects on data, technology, and citizen participation. Member of Decidim, the web platform for participatory democracy.

Former researcher at Barcelona Media, Eurecat - Technology Centre of Catalonia, and the Oxford Internet Institute – University of Oxford for projects on computational social science, social media, digital humanities, and big data.

Artificial Intelligence: Wikimedia Research ecosystem

Wikimedia: an ecosystem of free knowledge

WMF research team's programs

Address Knowledge Gaps

Improve Knowledge Integrity

Grow the Research Community

White Papers: https://meta.wikimedia.org/wiki/Research:2030

WMF research team's programs

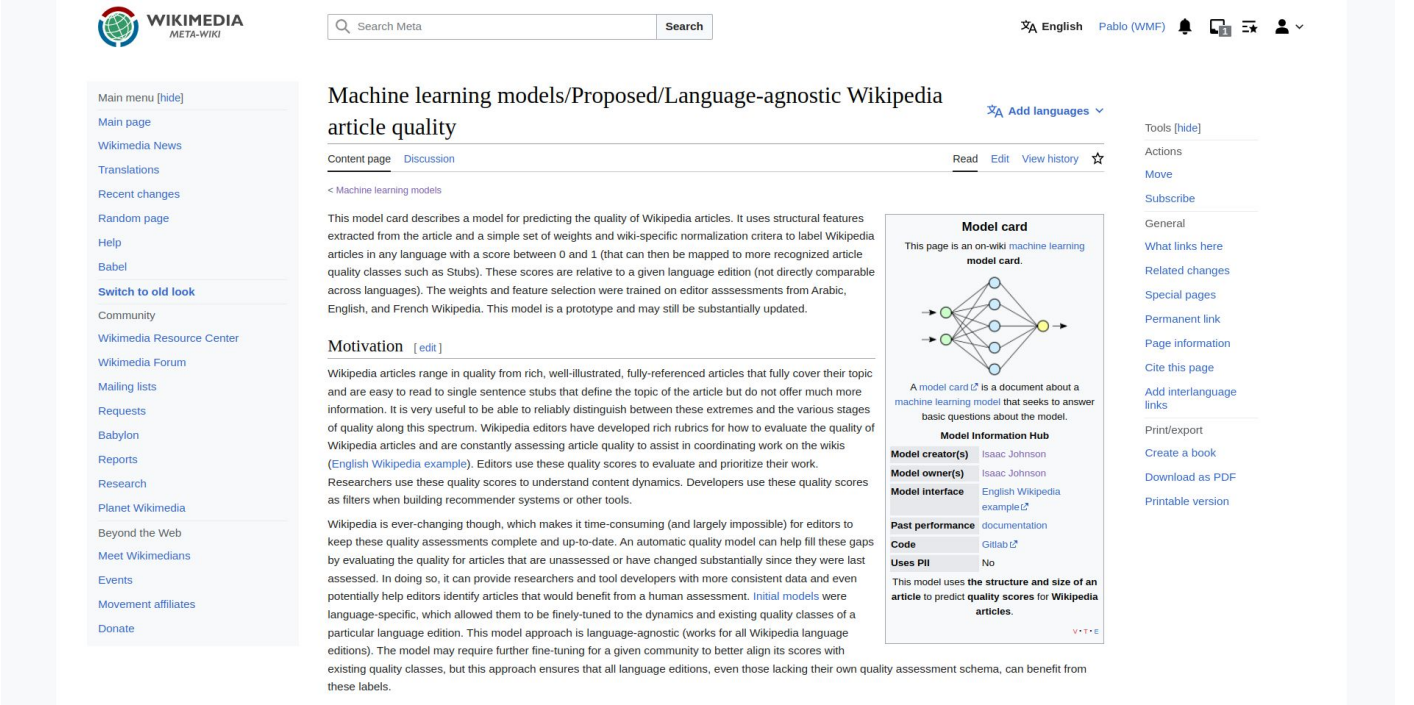
Address Knowledge Gaps

Improve Knowledge Integrity

Grow the Research Community

White Papers: https://meta.wikimedia.org/wiki/Research:2030

Addressing knowledge gaps with AI/ML



WIKIMEDIA META-WIKI

Search Meta | Search | English | Pablo (WMF)

Main menu [hide]
Main page
Wikimedia News
Translations
Recent changes
Random page
Help
Babel
Switch to old look
Community
Wikimedia Resource Center
Wikimedia Forum
Mailing lists
Requests
Babylon
Reports
Research
Planet Wikimedia
Beyond the Web
Meet Wikimedians
Events
Movement affiliates
Donate

Machine learning models/Proposed/Language-agnostic Wikipedia article quality

Add languages

Content page | Discussion | Read | Edit | View history

< Machine learning models

This model card describes a model for predicting the quality of Wikipedia articles. It uses structural features extracted from the article and a simple set of weights and wiki-specific normalization critera to label Wikipedia articles in any language with a score between 0 and 1 (that can then be mapped to more recognized article quality classes such as Stubs). These scores are relative to a given language edition (not directly comparable across languages). The weights and feature selection were trained on editor asssessments from Arabic, English, and French Wikipedia. This model is a prototype and may still be substantially updated.

Motivation [edit]

Wikipedia articles range in quality from rich, well-illustrated, fully-referenced articles that fully cover their topic and are easy to read to single sentence stubs that define the topic of the article but do not offer much more information. It is very useful to be able to reliably distinguish between these extremes and the various stages of quality along this spectrum. Wikipedia editors have developed rich rubrics for how to evaluate the quality of Wikipedia articles and are constantly assessing article quality to assist in coordinating work on the wikis (English Wikipedia example). Editors use these quality scores to evaluate and prioritize their work. Researchers use these quality scores to understand content dynamics. Developers use these quality scores as filters when building recommender systems or other tools.

Wikipedia is ever-changing though, which makes it time-consuming (and largely impossible) for editors to keep these quality assessments complete and up-to-date. An automatic quality model can help fill these gaps by evaluating the quality for articles that are unassessed or have changed substantially since they were last assessed. In doing so, it can provide researchers and tool developers with more consistent data and even potentially help editors identify articles that would benefit from a human assessment. Initial models were language-specific, which allowed them to be finely-tuned to the dynamics and existing quality classes of a particular language edition. This model approach is language-agnostic (works for all Wikipedia language editions). The model may require further fine-tuning for a given community to better align its scores with existing quality classes, but this approach ensures that all language editions, even those lacking their own quality assessment schema, can benefit from these labels.

Model card

This page is an on-wiki machine learning **model card**.

A model card is a document about a machine learning model that seeks to answer basic questions about the model.

Model Information Hub

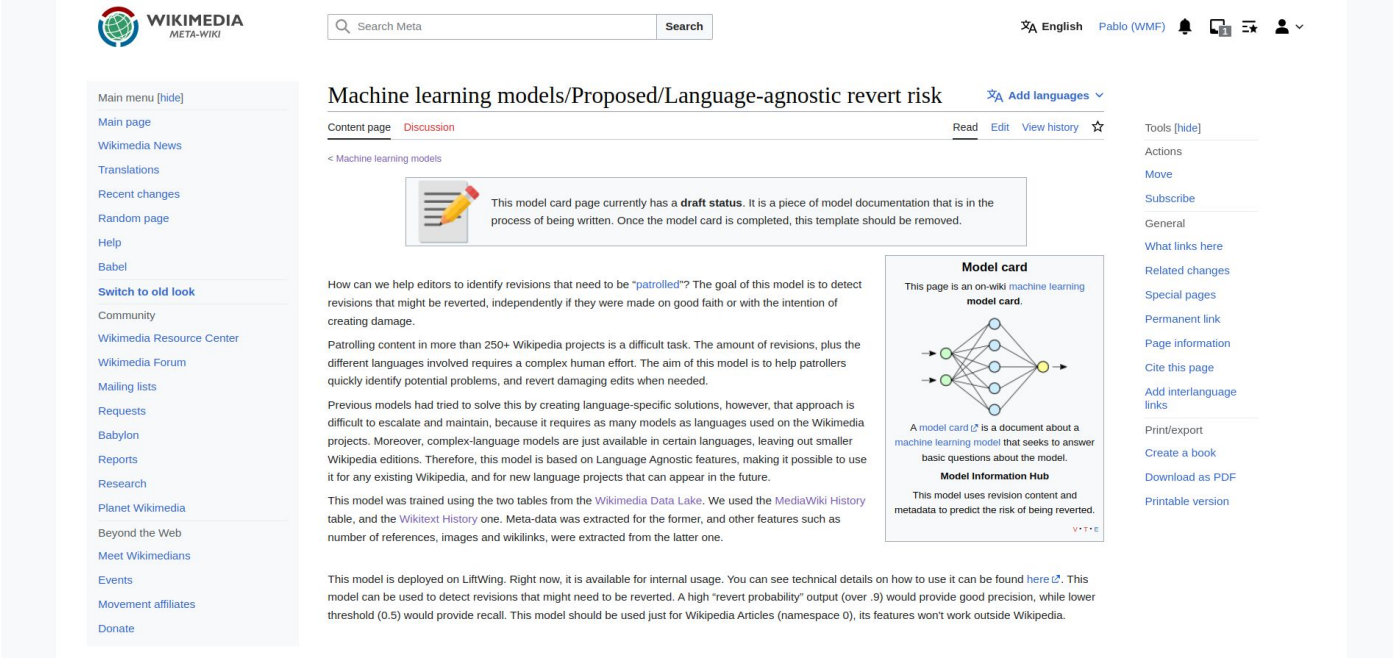
Model creator(s)	Isaac Johnson
Model owner(s)	Isaac Johnson
Model interface	English Wikipedia example
Past performance	documentation
Code	Gitlab
Uses PII	No

This model uses **the structure and size of an article** to predict **quality scores** for **Wikipedia articles**.

Tools [hide]
Actions
Move
Subscribe
General
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Add interlanguage links
Print/export
Create a book
Download as PDF
Printable version

https://meta.wikimedia.org/wiki/Machine_learning_models/Proposed/Language-agnostic_Wikipedia_article_quality

Improving knowledge integrity with AI/ML



https://meta.wikimedia.org/wiki/Machine_learning_models/Proposed/Language-agnostic_revert_risk

Our principles for AI/ML models

Open

Training and inference code are open and public

Reliable and Scalable

In collaboration with the ML Platform Team

Multilingual

Preferring Language-Agnostic approaches, to give the same opportunities to all our communities

Explainable

Explainability is as important as accuracy

Community-centered

Communities are encouraged to provide feedback or report biases, to continuously improve models

WIKIMEDIA FOUNDATION

Artificial Intelligence: Decidim ecosystem

Free open-source participatory democracy for cities and organizations

decidim features

proposals
your ideas in detail

The participatory texts component can be used to convert lengthy text documents into various proposals or results and, vice versa, to compose and display a unified text based on a collection of proposals or results.

results
no proposal without an answer

The results component is used to turn proposals into results and give official responses concerning their acceptance or rejection, merging various proposals into a single result.

meetings
to meet and not miss anything

The meeting component offers organizations and participants the opportunity to convene meetings, determine their location and time, register and limit attendees, define the structure and content of the meeting as well as publishing the minutes, and the resulting proposals.

voting
decide your vote

The voting component offers organizations the possibility of activating different voting or support systems around proposals: unlimited, limited to a given threshold, weighted, cost-based, etc.

accountability
transparency from beginning to end

The accountability component offers the possibility of subdividing results into projects, defining and applying progress statuses around their implementation, as well as displaying the extent of the results' implementation grouped by categories and scopes.

participatory texts
to analyze, synthesize, and build in common

The participatory texts component can be used to convert lengthy text documents into various proposals or results and, vice versa, to compose and display a unified text based on a collection of proposals or results.

conference
where big events are premiered

The conference component allows an organization to create a website for a big event by joining up a series predefined meetings (chats, workshops etc.), putting together a unified program and managing attendees.

surveys
because the opinion of your community matters

The surveys component can be used to design and publish surveys and to display and download their results.

sortition
equality and the justice of randomness

The sortition component allows to select a number of proposals (e.g. candidates for a jury) with random, yet reproducible, procedures that guarantees non-biased and uniform distributions.

comments
building a collective intelligence

The comments component enables users to add comments, to identify the comment as being in favor, against or neutral in relation to the commented object, to vote comments, respond to them and to receive notifications about responses.

pages and blogs
your community, informed and up to date

The pages component is used to create informative pages with rich text formatting, embedded pictures and videos. The blog component makes possible the creation of posts or news, and to navigate them chronologically.

notifications
personalized information on interesting contents

Decidim enables you to track any space or component to receive updates every time they happen.

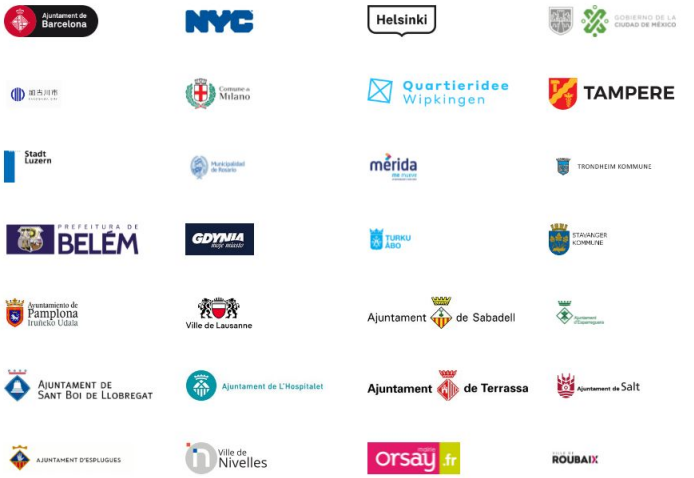
newsletter
simple but powerful

The newsletter component makes possible to send emails to everyone registered in the platform or, more selectively, to those who participate in a specific space.

Over 400 organizations already use Decidim

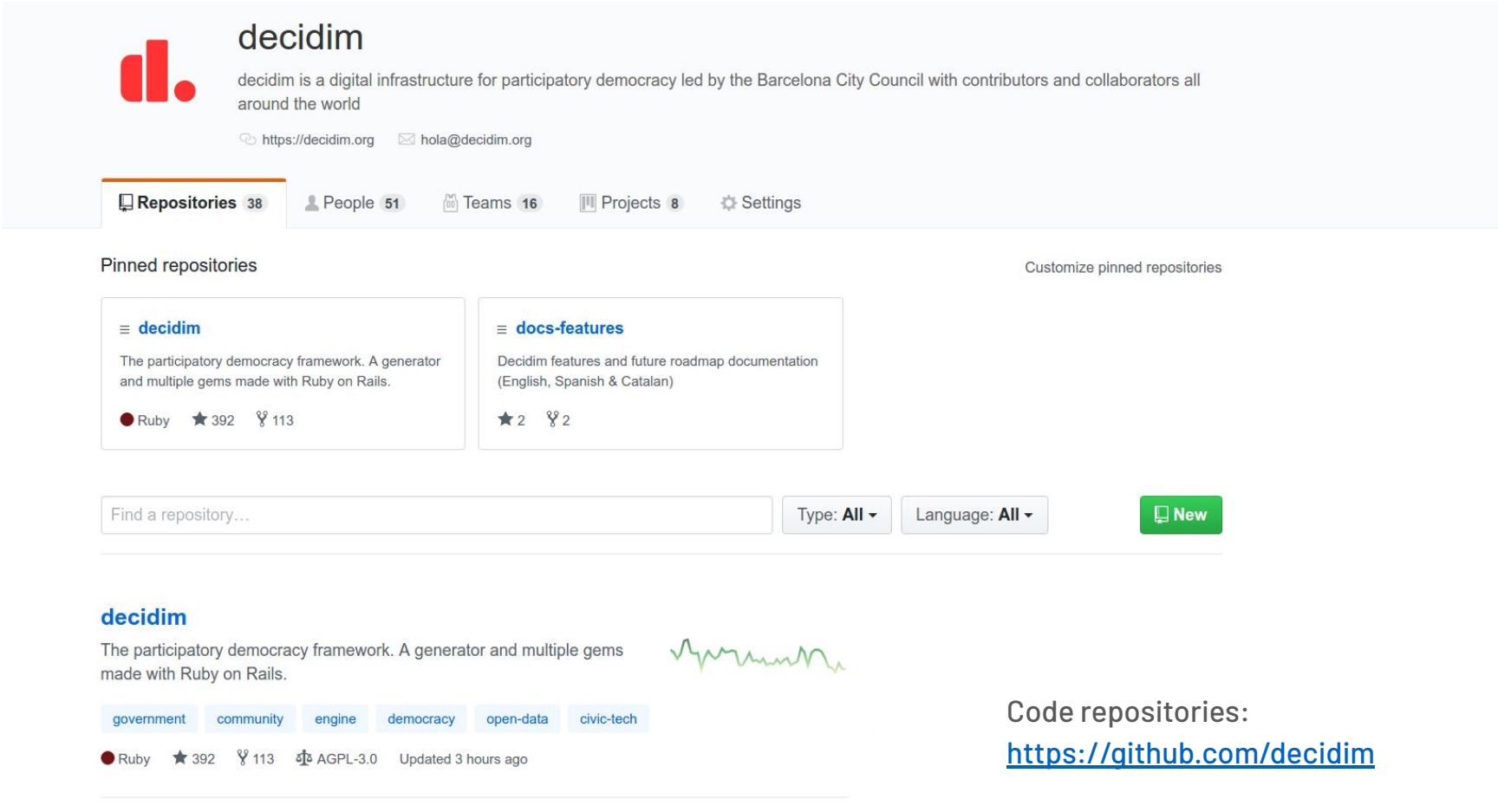
30 countries | 240 public institutions | 180 social organizations

cities



regions

organizations

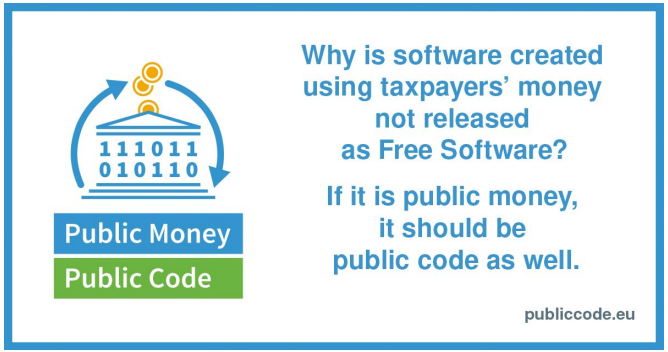


Code repositories:
https://github.com/decidim

/Principles

Software license

GNU Affero General Public License v3.0



Social Contract

Democratic guarantees code

- Open to collaboration
- Transparency, traceability, integrity
- Democratic quality guarantees
- Privacy and security

/Principles

Governance

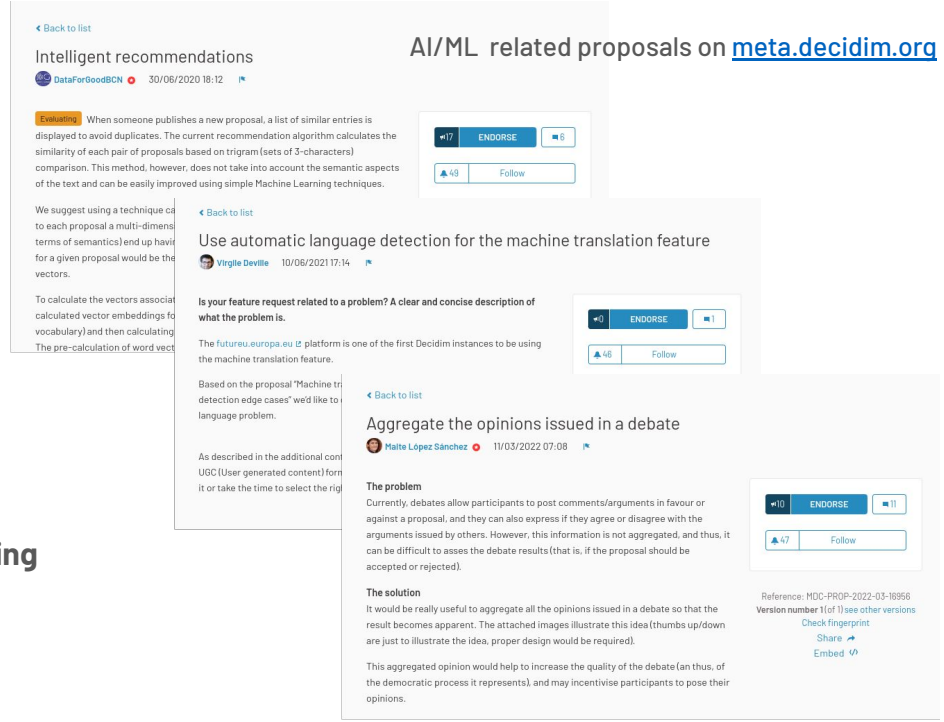
Open community: participation

- Proposal of new features
- Bug reporting
- Documentation
- Language translation
- Community and research meetings

Decidim Free Software Association: decision-making

- Coordination Committee
- General Assembly

AI/ML related proposals on meta.decidim.org



DECIDIM FEST

18 - 20 oct.
2023
canòdrom
bcn

Democracy, technology and ~~artificial~~ collective intelligence

Open Call

Ajuntament de Barcelona

Thanks

Contact: paragon@wikimedia.org

DIGITAL
HUMANISM
SUMMIT

A.I.
AND
DEMOCRATIC
SUSTAINABILITY