# [TITLE]

Adam Wight
awight@wikimedia.org
Wikimedia Foundation

## 1 INTRODUCTION

Artificial intelligence is indispensable to the existence of Wikipedia in its current form. As the use of AI expands in order to help human knowledge workers keep up with the flood of changes, we face the challenge of moderating AI so that it is a positive force on human relationships.

AI is unable to do more than ape the human behaviors we train it on. This means that most effective way to mitigate biases, e.g. favoring English language references, or misogyny in curation, is if we can achieve these same improvements in the culture of the group whose data is being used to train the AI. The same interventions which will improve group social health, such as providing healthy and usable communications channels, and encouraging interaction between subgroups, will also improve the AI's (impact on) social health.

We introduce an auditing system "Judgement and Dialogue Engine", or JADE, which humanizes false positive reporting, and explain why we believe that something like this is necessary for any AI which might have impact on the public. JADE will provide a platform for discussion between auditors, a quantitative mechanism for sharing opinions, and in a manner of speaking, a feedback loop between JADE auditors and the AI will provide something like a dialogue between human and machine, which allows the machine to evolve its models.

The Wikimedia Foundation is in a good position to pioneer such a system, because of its unusually low dependence on capitalist funding, and existing expectations of transparency and participation among its contributor base. Ordinarily, for-profit companies will optimize their AIs in ways that increase profits. As this might be received critically by their customer base, and leaked details might allow for gaming the system, the exact algorithms and tuning parameters necessary to understand their behavior are almost always be a tightly-held trade secret. There are known techniques which allow us to probe these "black-box" algorithm, but the most effective methods are currently illegal to use in the United States.[1] JADE could be applied to these commercial systems, in either its full capacity as a feedback cycle with the cooperation of the algorithm owners, or as an external auditing tool that can be used collaboratively by investigators.
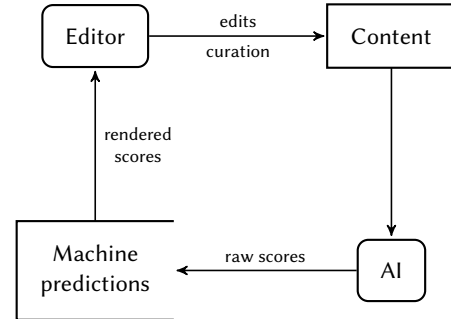
---

[1][3]
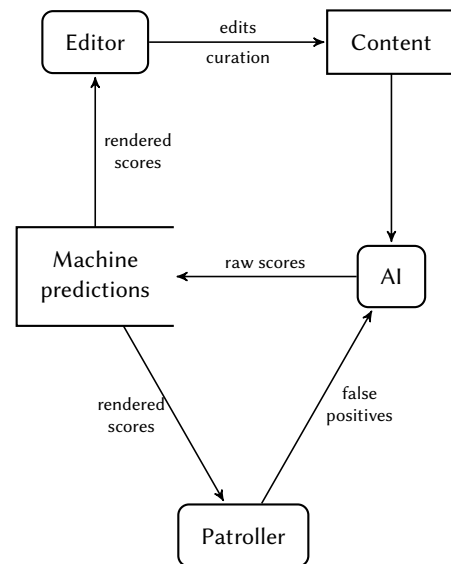
**Figure 1: Unmoderated AI feedback**



**Figure 2: AI feedback with auditing**

## 2 RESOURCE LIFECYCLES

This section is intended to give an outline of our main resource streams, their flow, and who holds the control. This is the social and material context in and upon which the AIs operate.

### 2.1 Money

This story begins with money, only to say that it's not a dominant force in the Wikimedia movement. Money flows are an order of magnitude less than estimates of the worth of knowledge and labor involved.

*2.1.1 Where does the money come from?* The Wikimedia Foundation's revenue comes from donations and grants, and the lion's share from individual donations under $30. The Foundation is prohibited from using any of this money to directly produce wiki content.

- Individual donors giving $100 or less: $65M or about 75% of the WMF budget. [2] [3]
- Unrestricted large grants of $100 or more: $22M or about 25% of the WMF budget.
- Restricted, large grants: Assets of $3.5M [4]

Similar flows of money are donated to Wikimedia chapters, although the mix of donations, grants, and membership dues varies widely.

Another, more troubling, flow is the money being paid to some editors. These payments are unregulated and unaccounted for.

- Shadow editor employers: Unknown. [5]
- Declared paid editor employers: Unknown, see above.

*2.1.2 Where's the money?* The Wikimedia Foundation holds $113M in net assets.[3]

Wikimedia chapters have independent finances. To take two examples of large chapters, the German chapter has roughly $6M in assets, and $6M in revenue. The Swiss chapter has $1.8M in assets and $6M in revenue.

*2.1.3 Who controls the money?* The Wikimedia Foundation's Board has the final say over how their donor money will be spent. This Board consists of 10 members, 5 are self-appointed, 2 members are chosen by chapters, and 3 are nominally elected by a vote. In the 2017 Board election, 874 votes were tallied, or 1.2% of active editors. Internally, the Wikimedia Foundation is organized into a corporate hierarchy in which the executives are the effective body making practical decisions about how to implement mandates given by the Board.

A separate, elected entity called the Funds Dissemination Committee makes recommendations covering roughly $1M of grants.[6]

Wikimedia chapters are legally independent and responsible for their own budget. Most have a board and and an executive staff.

Paid editor employers make private decisions about how to allocate their money, which we can only speculate about by making an analysis of paid edits.

*2.1.4 Where does the money go?* The Wikimedia Foundation's major spending is $33.7M in payroll, and $11.2M in grants to aligned projects. $24.2M more is spent on a mixed bag of services and other supporting costs.

Paid editor funds are like payroll, in that they don't do any further work in the wiki ecosystem once paid, other than making paid editing a sustainable livelihood.

## 2.2 Knowledge

Knowledge is the bread and butter of the Wikipedia movement. This is its currency and the most unique aspect of its worth.

*2.2.1 What is knowledge?* Good question. For our purposes, it has two forms, one is its written form and the other when it's retained in a living human. Knowledge can be true or false. It can be useful or irrelevant to the learner, depending on context. Knowledges can be contradictory and yet coexist. Knowledge has value. Its use value can be decoupled from its exchange value.

The discussions on "talk" pages are the primary means of coordinating and communicating content creation, but we'll exclude from a tally of knowledge, as these are an intermediate product of knowledge production.

*2.2.2 How much knowledge is there?* The corpus of all wikis consists of 351M articles. One estimate is that the replacement cost of Wikipedia would be $6.6B, and the consumer benefit is in the hundreds of billions of dollars.[7]

Content is not equally distributed among wikis, for example 13% of the pages are in English Wikipedia, and the top 6 wikis account for half of all articles. This shows that the project is far from complete, if all 288 languages of Wikipedia were to catch up with English content, it would require a 36-fold increase. It also shows that we have biases favoring English content.

*2.2.3 Where does knowledge come from?* Wikipedia's guidelines suggest that its knowledge should mostly come from secondary sources.[8] Heather Ford points out that these rules, if followed, would create a bias towards scholarly works which only cover a tiny fraction of subjects, but that the actual sources used are only estimated to be 53% secondary, with tertiary sources accounting for 13% of citations and primary sources 34%.[9]

More troubling, the geographic biases in sources are extreme. 56% of the sampled sources were from the U.S., and 13% from the U.K.

[Graham: editing going to colonial language wikis]

*2.2.4 Where is knowledge housed?* Our corpus of knowledge is the written form stored on publicly available servers. The Wikimedia Foundation hosts the dominant, editable interface to our corpus, but the raw data and forks are available, in the ballpark of [hundreds] on the Internet, and [thousands] of offline editions.

There is tremendous knowledge in the intersubjective minds of our current and future editors, and this guides their editing. According to wiki rules, editor's personal experience isn't supposed to be directly transcribed into the corpus, it can only mediate work done with other sources of knowledge.

*2.2.5 Who produces knowledge?* Editors and other wiki contributors are the only producers of knowledge.

*2.2.6 How does knowledge flow?* Content is added, edited, and consumed mostly using the MediaWiki Web interface. Roughly 5M new pages and 5M media files are added each year, and 220,000 articles were added to English Wikipedia during 2017.

*2.2.7 How is knowledge curated?* Not all information added to wikis will be retained. 1,500 pages are deleted every day, and ca. 8,000 edits are reverted.[10] Most wikis have no mandatory review for edits, and as soon as a change is made, it's visible to the public. A handful of automated bots follow these changes and can undo the worst vandalism almost instantaneously, then human patrollers review the most recent changes that remain, and may fix or undo the change, or mark it as good. These tools will be discussed below.

Changes that create a new article go through a more rigorous review, and the "new pages patrol" backlog has recently been the

---

subject of an experiment in restricting the "supply" of incoming articles.[11]

*2.2.8 Where does knowledge go?* Readers find and consume our written corpus. Doing so changes the reader's internal state, hopefully in wonderful ways.

Articles must be discovered in order to be read. If an article is orphaned or has few incoming links, it's unlikely to be read.

Countless AIs and other downstream software devour the corpus in real-time, as it evolves. There is no control or accounting for where this is going. Our licenses don't require anything from these consumers.

## 2.3 Labor

Most editing is done by unpaid editors, whose work is worth an estimated \$490M per year.[12]

*2.3.1 Who are the workers?* The number of active editors on English Wikipedia, those who have made 5 or more edits in the last month, is around 73,000. There are [XX] active editors across all wiki projects.

Currently, 2,076 editors, or 2.8% of the active editors on English Wikipedia disclose that they are paid for their work, using the recommended tag.[13]

Estimating by subtracting the number of paid editors, and extrapolating to all wikis, 97.2% of the active editors are assumed to be unpaid, or roughly 65,000 active editors across all wiki projects.

There's no authoritative census of editor demographics, this information is rarely self-reported, and surveys vary widely.[14] We can say with some certainty that women are underrepresented among editors, one survey estimated that 12.6% of editors identify as female.

*2.3.2 Who's the boss?* Unpaid workers are independent.

Declared and shadow paid editors are on either short- or long-term contracts with their clients, who might be individuals or organizations.

*2.3.3 Collaborative relationships between contributors.* Each article has an associated page for discussion, called a "talk page". The talk pages are a major part of editor culture, and are the preferred way to communicate and coordinate edits.[15] These pages exist to discuss improvements to articles, and it's discouraged to use them for other topics. Each talk page has a natural relationship to exactly one article, for example "Talk:Earth".

Another class of pages are the noticeboards, which each have a specific function such as requesting help with a dispute, or coordinating a work backlog.[16]

Interestingly, there is strong disagreement about the potential for more general social interactions on Wikipedia. Compare an essay encouraging community-building[17] (69 edits by 34 editors, 747 incoming links) to the essay discouraging social networking[18] (17 edits by 12 editors, 39 incoming links). Regardless of the lack of editor consensus on this issue, the MediaWiki technology itself is unadaptable to most social network behaviors. You can't "follow", "friend", or have private conversations. Spaces for discussion are mostly instrumental, intended for a single purpose.

If no space exists for other types of collaboration, it can be created in an ad-hoc fashion. For example, many WikiProjects have been created for communities of shared interest.[19] These are where you would coordinate work across articles in some subset of medical knowledge, for example. The ad-hoc format means that any correlation between a discussion and content must be created explicitly, often by pasting a bullet list of articles.

*2.3.4 Power relationships between contributors.* There is an explicit administrative hierarchy among volunteers, in which some individuals and bodies have the privileges to overrule decisions, and restricted technical powers such as the ability to view private information about editors.[20] Contributors with these extra powers are not allowed to directly manage or command others, and the powers are granted through open nominations and consensus or vote.[21] Still, this elevated power is a site of contest and is at odds with the generally wiki culture of equality and democracy.

Users can be blocked in cases of extreme misbehavior, normally something that's done to prevent ongoing damage to the content. Blocking is just one tool among the dispute resolution processes.[22]

Editors will sometimes be granted respect on account of a high edit count, which serves as a "quick and crude aid" to gauge another person's experience within the project.

*2.3.5 Alienation.* Alienation is the lack of community between editors.[connectivity studies?]

On a larger scale, isolation between wiki language communities is another form of isolation.[self-focus]

*2.3.6 What are the tools of production?* Editors own their own computers. The Wikimedia Foundation owns servers and storage attached to their flagship domain names. The Internet does the distribution. Readers will need a networked device to consume.

Some editors rely on bot assistance

*2.3.7 Where is labor performed?* Editing is mostly done in isolation. Occasionally, groups of editors get together to discuss or edit. Paid editors may work from an office.

*2.3.8 Who owns the tools of production?*

*2.3.9 What are the preconditions for knowledge production?* At the moment, editors must have unhindered Internet access. They must be literate, and if unpaid must have the free time to edit. They must have tech skills and be motivated to participate. If editors lose any of these, for example through time constraints, or discouraging social interactions, they will leave the project.

*2.3.10 What is being produced?* [don't repeat knowledge section] [how much is produced?]

---

[11] https://en.wikipedia.org/wiki/Wikipedia:New_pages_patrol, https://en.wikipedia.org/w/index.php?title=WP:ACTRIAL
[12] [2]
[13] https://tools.wmflabs.org/templatecount/index.php?lang=en&namespace=10&name=Connected+contributor+%28paid%29
[14] https://en.wikipedia.org/wiki/Wikipedia:Wikipedians#Demographics
[15] https://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines
[16] https://en.wikipedia.org/wiki/Wikipedia:Noticeboards
[17] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_a_community
[18] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_not_a_social_networking_site

[19] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Directory
[20] https://en.wikipedia.org/wiki/Wikipedia:Administration#Editors
[21] https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship, https://en.wikipedia.org/wiki/Wikipedia:What_adminship_is_not
[22] https://en.wikipedia.org/wiki/Wikipedia:Dispute_resolution

## 3 ARTIFICIAL INTELLIGENCE: TO THE RESCUE?

## ACKNOWLEDGMENTS

## REFERENCES

[1] Ford, H., Sen, S., Musicant, D. R., and Miller, N. Getting to the source: Where does wikipedia get its information from? In *Proceedings of the 9th International Symposium on Open Collaboration* (New York, NY, USA, 2013), WikiSym '13, ACM, pp. 9:1–9:10.

[2] Lund, A., and Venäläinen, J. Monetary materialities of peer-produced knowledge: The case of wikipedia and its tensions with paid labour. *TripleC 14* (02 2016), 78–98.

[3] Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry* (2014).

## GLOSSARY

**Wikimedia chapters** Independent organizations founded to support and promote the Wikimedia projects in a specified geographical region, in most cases, a country.

**Wikimedia Foundation** A non-profit organization registered in the USA, which provides hosting and software support to run the Wikimedia sites.

**Wikimedia movement** The totality of people, activities, organizations, and values which revolve around Wikimedia sites and projects.

**Wikimedia sites** Wikis which are officially supported by the Wikimedia Foundation and the Wikimedia movement. Wikipedia is the best known project, see https://wikimediafoundation.org/wiki/Our_projects for the others. Most projects exist independently in each language.