Ms. Irene Khan
Special Rapporteur on the Right to Freedom of Opinion and Expression
Office of the United Nations High Commissioner for Human Rights
Geneva, Switzerland

San Francisco, CA - February 23, 2021

Wikimedia's contribution to the the Special Rapporteur's thematic report on disinformation

The Wikimedia Foundation is grateful for the opportunity to submit input for the Special Rapporteur's annual thematic report, focusing on the issue of disinformation and freedom of opinion and expression. As the non-profit organization that hosts and supports a number of free, online, collectively-produced resources, including Wikipedia, we welcome this focus on disinformation. Preventing and combating disinformation is vital to preserving and protecting a healthy public discourse, transparency and accountability, strong civil society, as well as trust in institutions and the internet.

We also want to underscore the urgency of this issue, as the devastating potential of targeted disinformation campaigns to impact elections, public health, and the safety of marginalized communities has been evident for several years. What can begin as words online, has the ability to turn into irreparable harm in the real world as people use that information to make choices about their health, their identity, and their actions. This is especially true on Wikipedia, where topical articles, like those related to the Covid-19 pandemic, receive millions of pageviews and often see spikes in traffic when important new developments occur and news stories are reported.

In addition, disinformation campaigns can have a lasting effect that undermines confidence in the foundational institutions of knowledge such as journalism, academia and research, and

public-interest institutions including government. This affects collective knowledge projects like Wikipedia, which require citations to the very sources that disinformation destabilizes.

Wikipedia and reliable, trusted knowledge

As the world's largest online encyclopedia, and one of the most consulted knowledge resources online [with over 20 billion views of articles every month](), Wikipedia [exists to]() provide people with reliable information about the topics, moments and people who shape our world. While anyone can contribute to Wikipedia, that simple description omits the many systems, structures, policies, and practices that are supported by tens of thousands of volunteers, who curate, refine, monitor, and protect the information on the website every day in around 300 languages. They do this work because of a shared commitment to knowledge, and without financial compensation.

Disinformation is not new to the volunteers who edit and maintain Wikipedia, who have been combating threats to knowledge integrity for as long as the website has existed. This community and the Wikimedia Foundation are increasingly concerned about coordinated campaigns from state actors, special interest groups, or self-organized groups online. In addition, disinformation not directly targeted at Wikipedia tends to eventually arrive there as well, when those convinced by disinformation campaigns edit articles based on that false information.

Wikipedia is also fundamentally different from other online platforms because nearly all participation is driven by the needs of building an encyclopedia that aims to encompass the sum of all human knowledge. Rather than breaking the news, Wikipedia articles are a record of how the news has broken. Instead of focusing on being the first to break a story, community contributors focus on accuracy. Wikipedia articles give context to an issue with the most accurate, verifiable information, and Wikipedia's content is central to the broader knowledge ecosystem. Content on the platform is freely licensed in a way that allows for companies and individuals to use, edit, and remix it onto their own platforms. Because of this, we at the Foundation, and across the broader free knowledge landscape, have to recognize that disinformation on one platform can have far-reaching effects across the web. Despite there being no way to amplify content on Wikipedia itself, e.g. through a broadcast button, content on the Wikimedia projects is often re-used on other platforms, from search results, to voice assistants; so if something is wrong on Wikipedia, it can be wrong nearly everywhere else too.

Several key content policies of Wikipedia underscore the commitment its contributors and the projects at large make to reliable, trusted knowledge. These policies include:

Verifiability

The policy of verifiability means that every fact on Wikipedia must be cited to a reliable source. This is perhaps the most important policy for combating disinformation on the Wikimedia projects, but also one that is most threatened by disinformation and the loss of trust in media. Sources can be anything deemed a reliable source, ranging from a newspaper article to a university textbook, and it is up to individual editors to identify whether a source is considered reliable. To assess reliability, editors look at the characteristics of a source. In what context was this published? Is this particular source known for being fact-checked or issuing corrections? How old is the information and could it be outdated? Editors seek sources that have processes with integrity; the discussions about which sources to use on Wikipedia are transparent and the decisions are public on the site for all to see.

Ultimately, this has led to an editor-maintained list of both perennial sources, which are generally considered reliable, and deprecated sources, which are no longer considered reliable for use on Wikipedia. This practice is non-exhaustive and relatively new as of the last few years, but has followed Wikipedia's norms around both transparency and consensus building. Already, some sources like The Sun (UK) and Breitbart News (US) have been labeled defacto unreliable except in very specific circumstances. In this way, we are seeing the Wikipedia community organically reacting to an increasingly complicated news landscape where disinformation often looks more like traditional news sources and is increasingly endorsed by government officials and others in power.

No original research

An important accompanying policy to verifiability is the prohibition of original research on Wikipedia. The phrase "original research" is used on Wikipedia to refer to material—such as facts, allegations, and ideas—for which no reliable, published sources exist. While this policy complements and enhances the verifiability policy, it also serves as a powerful policy against disinformation and conspiracy on its own. The prohibition on original research prevents the intentional misrepresentation of facts contained in otherwise reliable sources, disallowing analyses or conclusions based on multiple sources if those sources do not directly support that conclusion. Additionally, primary accounts are also generally treated as original research without additional sources to verify those claims, an increasingly important and complicated

policy as breaking news events are documented instantaneously online through things like personal accounts, videos, and photographs. As a platform dedicated to a best possible reflection of the world's knowledge, this policy helps not only to maintain trust in the platform over time but also to ensure that coverage of topics on Wikipedia remains as neutral as possible.

Neutrality

Everything on Wikipedia has to be written from a neutral point of view. What this policy means in practice is that articles should represent fairly, proportionately, and, as far as possible, without editorial bias, all the significant views that have been published by reliable sources on a topic. Wikipedia is where people go to see an overview of a subject, and editors cannot selectively pick sources or facts for articles.

That does not mean, however, that fringe views are given equal weight on Wikipedia. In fact, the neutrality policy specifically warns against giving "equal validity" to extraordinary claims or conspiracies which may give the impression that such claims are just as valid as those cited widely to reliable sources. For example, in Wikipedia's article about climate change, only a small subsection of the much larger article is dedicated to climate change denial, and that information is contextualized by describing how it diverges from broad scientific consensus.

Wikipedia contributors have a shared goal, the creation of an encyclopedia, and because of that the project is one of the only places on the internet where people's contributions tend to become more neutral the more they participate. In 2016, researchers Shane Greenstein and Feng Zhu at the Harvard Business Review published a study with findings supporting Wikipedia as a grounds for more neutral interactions and outcomes. They found that editors' contributions and discussions tended to become more neutral over time as controversial issues were discussed with an aim toward improving the underlying Wikipedia article. We believe that the neutral, collaborative nature of the Wikimedia projects, compared to an open forum such as social media, is a key aspect of preventing the spread of disinformation on the projects.

Other policies

In addition to these broad content policies, Wikipedia editors have created more granular rules and systems to ensure that the spirit of these policies are upheld. This includes content guidelines like special rules for biographies and articles about medicine. Many of these guidelines have evolved with the needs of the community, and volunteer editors are the first to catch and enforce violations of these guidelines and the broader site policies.

Part of what makes these principles and tactics of moderation work so well is that the site is radically transparent, meaning that it essentially operates as a large peer review system where anyone can review content and decisions. Everything from the way an article has evolved over time, to the discussion page for article changes, to administrator noticeboards — where more experienced volunteers discuss potential issues on the site — is all publicly viewable. Taken together, this creates a sense of accountability and service to the public in Wikipedia volunteers who know that this information has an impact on how people see, experience, and learn about the world.

Defense against disinformation

In a way, disinformation and misinformation are just violations of Wikipedia's policies on content. To defend against them, by monitoring new information that is added to Wikipedia and applying and enforcing the policies described above, processes and systems have been developed and refined by the communities of contributors over many years. While Wikipedia can be edited by anyone, there are strict governance structures and hierarchies that allow the community to curate and moderate information collaboratively. Trusted, long-time editors are voted into positions of increased access to moderation actions, including freezing certain articles from being edited, preventing people who repeatedly violate Wikipedia policies from editing, or overturning specific content decisions. This governance system works independently from the Wikimedia Foundation as the host of the website and is based on people from around the world making decisions together through votes and consensus in their language project communities.

Human Review

Quality control on Wikipedia relies mostly on human review of information, but is aided by technology ([such as sorting tools or bots](#)) that can fulfill easy -- but time-consuming -- tasks like flagging of obvious copyright violations for review or detection of vandalism (i.e. the defacing of Wikipedia articles), and fixing internal links or external references. Such automatization of repetitive tasks that are tedious to do manually, frees up volunteers' time to thoroughly review changes to Wikipedia's content that are more difficult to assess with a view to their compliance with policies and their veracity while also considering specific context.

This human review, in collaboration, depends heavily on a rough consensus and mutual understanding of the objectives of Wikipedia and a common purpose, which Wikipedians have

built their community and enormous online encyclopedia around. The Wikimedia Foundation supports the volunteers through the development of technical tools that aid to quickly detect low quality information that is added to Wikipedia. Even so, content moderation on Wikipedia is a lot of work for the editors, especially when someone engages in disruptive behavior or tries to push a certain perspective beyond a generally accepted consensus. Wikipedia is not perfect and will always be a work-in-progress; it is a living, dynamic body of information and knowledge. The community has full editorial agency over the websites and problems are only escalated to the Foundation, as the host and "support unit", when community mechanisms and governance fail to effectively address them.

Coordination

In the meantime, Wikipedia has become the main reference work for millions of people around the world, and home assistant services and other re-users of freely licensed content rely on it to serve consumers and train algorithms used across industries, which makes it a potential target for disinformation campaigns. A well-funded and coordinated campaign could infiltrate the website's articles on specific topics to misconstrue facts to a certain degree and for a limited time, but the Wikipedia volunteer community and the Foundation have been working together to take steps to improve Wikipedia's defenses and reduce response times.

The most recent example for this collaboration is a joint effort around the 2020 US Presidential elections for which a [task force at the Foundation was formed](#) to coordinate proactive and reactive measures around disinformation that could affect the outcome of the vote and liaise closely with members of the community who wanted to help. A couple years prior to this, the community proactively began monitoring key political articles for edits that were malicious or not in good faith, including personal articles on nominees, important political parties, and key political races. When appropriate, these editors can apply editing restrictions to these potential targets, i.e. preventing newly created accounts from contributing to certain articles or freezing those pages for edits from anyone.

The combination of preventive steps, heightened attention by the community and the public, and default transparency of any manipulation attempts across Wikipedia's content limits attack vectors and reduces the impact of successful attacks through near-instant, effective response.

Contributing up-to-date knowledge that is freely accessible

Wikipedia volunteer communities have stepped forward and further organized themselves to protect and enhance articles during the pandemic as well. Thanks to the hard work of

volunteers around the world, more than 6,900 Wikipedia articles about the pandemic have been created in 188 languages. These articles have been visited more than 579 million times. The main article on COVID-19 in English has been translated into 141 different languages, and has been viewed over 82 million times. Every day, the [WikiProject COVID-19 group](#) reviews, prioritizes, and translates content to ensure that these articles remain accurate, and resistant to undue influence. Such regularly updated, easily accessible knowledge about the pandemic can help address the harm by misinformation around COVID-19 elsewhere on the web, especially as many reliable news sites change their revenue models and high quality information is often only accessible via subscriptions, i.e. disappears behind paywalls.

Updating Wikipedia policies

In addition to content-focused approaches to disinformation, the Wikimedia Foundation has taken steps to enhance the policy structures that support collaboration on the projects to make them safer, and more resilient to bad actors. Following a 19-month consultation process, the Wikimedia Foundation recently ratified a new [Universal Code of Conduct](#) to support these goals. This Code was drafted by a joint committee of volunteers and Foundation staff in collaboration with Wikimedia community members around the world, in order to create a more welcoming, safe, and inclusive environment for all contributors, and a more open and thriving movement for free knowledge.

In addition to policies that will promote mutual respect and civility, and discourage harassment and aggression on the platforms, the Code of Conduct also targets disinformation, including the deliberate addition of biased, false, or inaccurate content; and systemic manipulation to favor specific interpretations of facts, or deliberately false rendering of sources. Such additional clear rules will give individual Wikipedia volunteers increased clarity and guidance on how to deal with misleading information and enable them to make better and quicker decisions to remove content that doesn't meet Wikipedia's policies.

Recommendations

In the Wikimedia Foundation's view, there are no simple solutions to the problem of disinformation or the harms caused by it. There is also no way to remove disinformation from the internet or the world entirely. Any approach to risk mitigation in this context will have to be multidimensional and consider all stages of the information lifecycle, from collecting/gathering to writing/production, to sharing/distribution, to reading/consumption, to

further sharing/distribution or back to collecting/gathering. The Wikimedia projects successful model to address various issues of information quality, not just disinformation, makes us confident to articulate the following recommendations:

Transparency

Transparency is both an enabler of collaboration and a necessary requirement for trust online. Version history, clear traceability of information, and openly visible decision-making processes are necessary to allow people from various backgrounds to verify the veracity of information and comprehend why they are seeing certain kinds of information but not other kinds. While the Wikipedia model cannot be replicated in all contexts or for all platforms, we recommend that transparency should be prioritized as a design standard for internet architecture and platforms' practices whenever possible.

Clear Policies

A diverse and detailed set of clear policies enables the community of volunteers to engage in content moderation and make informed and good decisions about information quality. These policies also give users and volunteers a clear view into how and why content moderation decisions are made, and how to potentially appeal these decisions. The policies are supported by the editors of Wikipedia who have developed them collectively. They are concisely articulated to serve the purpose of an online encyclopedia. Clear policies help moderators at online platforms, be they hired professionals or volunteers from a community, to make justifiable and defensible decisions, even in complex situations that require a certain level of discretion.

Skills and tools

Editing on Wikipedia and engaging in content moderation discussions requires an increased level of digital (media) literacy. Identifying disinformation among a variety of sources is a skill that is useful in any online setting. Contributing to Wikipedia is a good way to hone that skill, which is transferable to navigating other platforms or to the consumption of other sources of online content too. Digital literacy is an important part of education and it should go beyond mere critical reading to include the use of tools that aid during the various phases of the information lifecycle. Teaching these skills should not just be the responsibility of states or

civil society: platforms should include it in the on-boarding process of setting up accounts or tutorials that they offer to new users.

Access to verified information

Wikipedia's value to its readers grows with every piece of verified and relevant information that is included in the website. However, many people around the world are prevented from contributing to the project because they simply do not have access to verifiable information and sources in their language. As a general matter, disinformation is more easily accessible than well researched reports or thoroughly investigated stories which require funding and therefore often disappear behind paywalls. In the scientific context, open access models address this issue and make research output more accessible to anyone and therefore verifiable. Governments and funders should support open access models and make publicly funded information accessible through free licenses. Innovative revenue models for journalism that both pay for good reporting and make it easily accessible should also be explored.

Regulation

Wikipedia's model of collaborative content moderation [has been shown to be effective](). The Wikimedia Foundation is only rarely forced to step in for the safety of users or the integrity of the website. This system of [subsidiarity]() functions because the law leaves room for community deliberation. And while there is a legitimate interest in protecting public discourse and democratic institutions from the harms of disinformation, increasing platform liability for disinformation uploaded by their users, could severely limit the effectiveness of the recommendations above. Wikipedia is a successful model in the fight against disinformation because volunteers are empowered to work transparently, jointly develop policies for content, and apply their skills to distinguish between information that is verifiable and such that is not. Further, as the [Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda]() makes the challenges of regulation in this space abundantly clear, restrictions on platforms to stem the flow of disinformation could also impact legitimate speech, and cause additional harms. Regulation of the internet ecosystem with the aim of curbing disinformation should consider the above recommendations around transparency, policies, skills, and access to verifiable information to ensure systems like Wikipedia's governance structures and content moderation can continue to work and be applied elsewhere.

WIKIMEDIA

FOUNDATION

Conclusion

In twenty years, Wikipedia has gone from an unproven, but ambitious project to democratize knowledge to a well [respected reference work on the internet](). While the project's model and policies aren't easily applied in other contexts or on other websites, the focus on verifiable information as a core principle has allowed Wikipedia to both include diverging perspectives and remain a trusted source of knowledge.

Together, Wikipedia contributors and the Wikimedia Foundation have developed policies, tools, safety measures to stop the spread of disinformation online, and we believe that factors such as transparency, clear policies, and access to quality information will be important to future disinformation prevention globally.

We appreciate the opportunity to submit our comments for your consideration, and hope to have further opportunities to engage with you and other organizations who share our goal that every human can freely share in the sum of all knowledge. Implicit in that desire, is a commitment to ensure that what is shared is free from undue influence and the caustic effects of disinformation.

We applaud your focus on this area, and welcome further discussions and continued engagement.

Sincerely,

Amanda Keton, General Counsel
Wikimedia Foundation