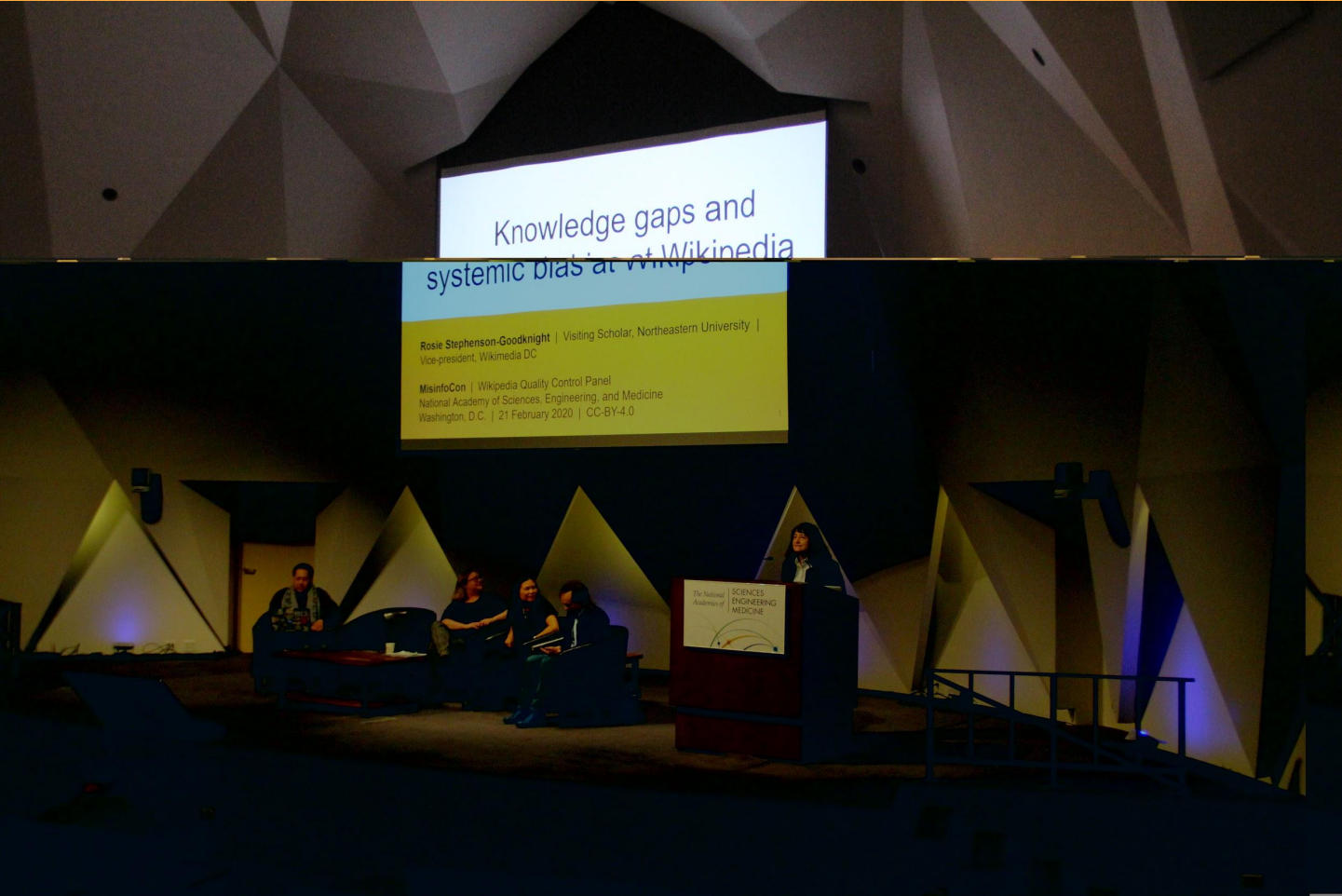


# Knowledge gaps and systemic bias at Wikipedia

**Rosie Stephenson-Goodknight** | Visiting Scholar, Northeastern University |  
Vice-president, Wikimedia DC

**MisinfoCon** | Wikipedia Quality Control Panel  
National Academy of Sciences, Engineering, and Medicine  
Washington, D.C. | 21 February 2020 | CC-BY-4.0



# Knowledge gaps and systemic bias at Wikipedia

Rosie Stephenson-Goodknight | Visiting Scholar, Northeastern University |  
Vice-president, Wikimedia DC

MisinfoCon | Wikipedia Quality Control Panel  
National Academy of Sciences, Engineering, and Medicine  
Washington, D.C. | 21 February 2020 | CC-BY-4.0



# Knowledge gaps and systemic bias at Wikipedia: the reader



# Who is accessing Wikipedia?

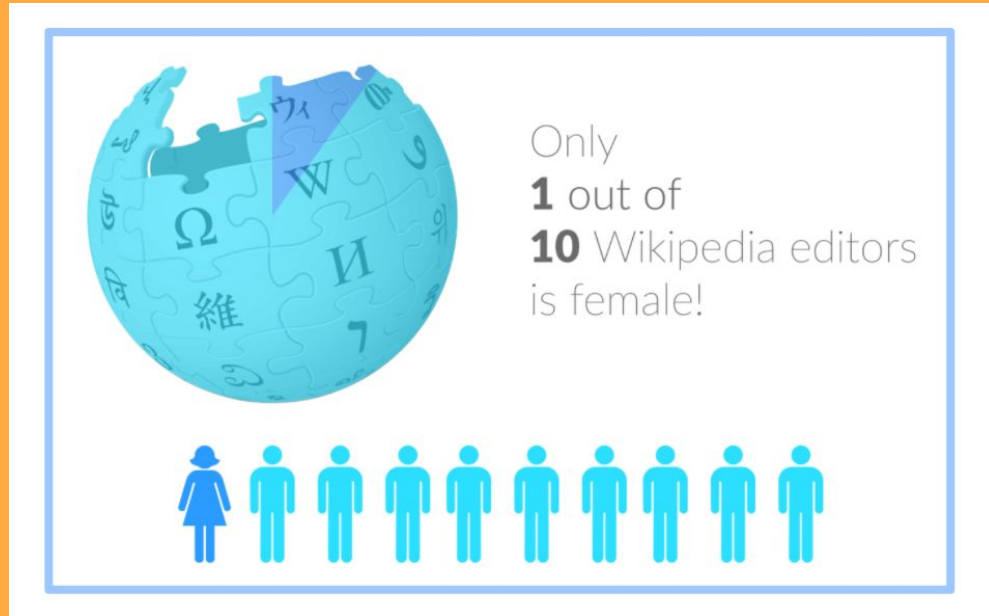
- January 2020:
  - Approximately 7.8 billion pageviews
  - Approximately 830 million unique devices accessed the English Wikipedia

# Who is reading Wikipedia?

## WMF Research: Characterizing Wikipedia Reader Behaviour

- March 2019, a small-scale pilot of the WMF survey was run on English Wikipedia.
- Survey respondents skewed heavily young and male.
- Including those who were under the age of 18, 70% of respondents were under the age of 30.
- Overall, 76% identified as male.
  - Gender percentage was consistent across age groups.
  - United States: 67% male
  - United Kingdom: 75% male
  - India: 83% male

# Knowledge gaps and systemic bias at Wikipedia: the editor



# Who is editing Wikipedia?

- Active users/editors (at least 1 edit/month): 138,492 in January 2020
- Very active users/editors (at least 100 edits/month): 3,541 Wikipedia editors; “and very few of them are female”. (*New Statesman*, 8 January 2019)
  - 77% of Wikipedia articles are written by 1% of Wikipedia editors (Daniel Oberhaus, *Motherboard*, *Tech by Vice*, 7 November 2017)
- 2018 Wikimedia Community Engagement Insights survey (<https://w.wiki/GUp>)
  - Median age: between 35 and 44 years.
  - 81% of the survey responders were from the Global North.
  - 85% of the responders have post-secondary education.

# Editor gender

<u>Year</u>	<u>Survey</u>	<u>% of women</u>	<u>Comments</u>								
2010	United Nations University study	12.6%									
2011	Wikipedia Editor Survey	8.5%	The Wikimedia Foundation set a goal of 25% women editors by 2015								
2015			No survey conducted								
2018	Wikimedia Community Engagement Insights survey <a href="https://w.wiki/GUp">https://w.wiki/GUp</a>	9%	<p>ED15: Gender across Wikimedia project contributors in 2018, weighted</p> <table border="1"><caption>Gender across Wikimedia project contributors in 2018, weighted</caption><thead><tr><th>Gender</th><th>Percentage</th></tr></thead><tbody><tr><td>Male</td><td>90.0%</td></tr><tr><td>Female</td><td>9.0%</td></tr><tr><td>Other</td><td>1.0%</td></tr></tbody></table>	Gender	Percentage	Male	90.0%	Female	9.0%	Other	1.0%
Gender	Percentage										
Male	90.0%										
Female	9.0%										
Other	1.0%										



# Knowledge gaps and systemic bias at Wikipedia: the content

## Gender Gap by occupation

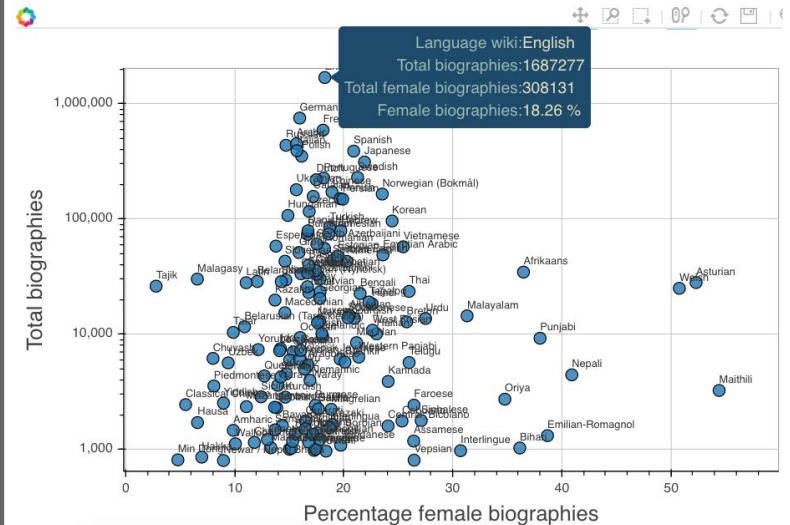
Occupation ▲	Total ▼	Total with gender ▼	Females ▼	% Fem. ▼	Gap	Males ▼	% Mal. ▼	Others ▼	% Oth. ▼
agent	3,311,455	3,209,217	573,452	17.869 %		2,634,587	82.094 %	1,178	0.037 %
worker	1,642,656	1,577,779	215,334	13.648 %		1,361,943	86.320 %	502	0.032 %
creator	1,568,603	1,520,948	365,985	24.063 %		1,154,148	75.883 %	815	0.054 %
professional	1,119,574	1,073,605	150,530	14.021 %		922,753	85.949 %	322	0.030 %
sports figure	807,432	798,712	117,139	14.666 %		681,493	85.324 %	80	0.010 %
artist	795,516	776,856	219,995	28.319 %		556,321	71.612 %	540	0.070 %
athlete	784,450	776,549	115,705	14.900 %		660,766	85.090 %	78	0.010 %
author	612,323	595,067	120,367	20.227 %		474,390	79.720 %	310	0.052 %
competitive player	536,164	530,903	52,640	9.915 %		478,234	90.079 %	29	0.005 %
position	515,678	500,024	76,387	15.277 %		423,523	84.701 %	114	0.023 %
politician	513,578	483,366	56,182	11.623 %		427,072	88.354 %	112	0.023 %
erudite	507,992	491,360	78,387	15.953 %		412,860	84.024 %	113	0.023 %
writer	447,502	434,403	93,448	21.512 %		340,702	78.430 %	253	0.058 %
researcher	444,189	429,036	65,341	15.230 %		363,604	84.749 %	91	0.021 %
scientist	294,872	283,041	41,699	14.732 %		241,280	85.246 %	62	0.022 %
musician	256,387	250,804	61,144	24.379 %		189,496	75.555 %	164	0.065 %
association football player	251,162	249,117	11,388	4.571 %		237,725	95.427 %	4	0.002 %
actor	245,492	241,279	108,133	44.817 %		132,879	55.073 %	267	0.111 %

# Percentages of women's biographies, a comparison

- “*Diccionario biográfico, geográfico e histórico de Venezuela*”, -Ramón Armando Rodríguez (1957)
- **3.6%**

- English Wikipedia
- Oct 2014: **15.53%**
- Feb 2020: **18.26%**

All time, as of 17 Feb '20



## Donna Strickland [edit]

ORES predicted quality:  GA (4.22)

A *B-class* article from Wikipedia, the free encyclopedia  
Canadian physicist, engineer, and Nobel laureate (Edit)

**Donna Theo Strickland**, CC (born 27 May 1959)<sup>[1][2][3]</sup> is a Canadian **optical physicist** and pioneer in the field of **pulsed lasers**. She was awarded the **Nobel Prize in Physics** in 2018, together with **Gérard Mourou**, for the invention of **chirped pulse amplification**.<sup>[4]</sup> She is a professor at the **University of Waterloo**.<sup>[5]</sup>

She served as fellow, vice president, and president of **The Optical Society**, and is currently chair of their Presidential Advisory Committee. In 2018, she was named as one of **BBC's 100 Women**.<sup>[6]</sup>

### Contents [hide]

- Early life and education
- Career
- Awards and recognition
  - Nobel Prize
  - Order of Canada
- Selected publications
- Personal life
- See also
- Notes
- References
- External links

## Early life and education [edit]

Strickland was born on 27 May 1959, in **Guelph**, Ontario, Canada to Edith J. (née Ranney), an English teacher,<sup>[7]</sup> and Lloyd Strickland, an electrical engineer.<sup>[1]</sup> After graduating from Guelph Collegiate Vocational Institute, she decided to attend **McMaster University** because its **engineering physics** program included lasers and **electro-optics**, areas of particular interest.<sup>[7]</sup> At McMaster, she was one of three women in a class of twenty-five. Strickland graduated with a **B.Eng.** degree in engineering physics in 1981.<sup>[8]</sup>

Strickland studied for her graduate degree in **The Institute of Optics**,<sup>[9]</sup> receiving a Ph.D. degree from the University of Rochester in 1984.<sup>[10]</sup>

## Donna Strickland



Donna Theo Strickland  
27 May 1959 (age 60)  
Guelph, Ontario, Canada

**Education**  
**McMaster University** (BEng)  
**University of Rochester** (MS, Ph.D)

**Known for**  
Intense laser-matter interactions  
**Nonlinear optics**  
Short-pulse intense laser systems  
Chirped pulse amplification  
**Ultrafast optics**

**Spouse(s)**  
Doug Dykaar

**Children**  
2

**Awards**  
**Alfred P. Sloan Research Fellowship** (1998)  
Fellow of the Optical Society (2008)  
**Nobel Prize in Physics** (2018)

### Scientific career

**Fields**  
Physics  
Optics  
**Lasers**

**Institutions**  
**University of Waterloo**

**Thesis**  
*Development of an ultra-bright laser and an application to multi-photon ionization* (1988)

## Katie Bouman [edit]

ORES predicted quality:  GA (4.31)

U.S. computer scientist. A *C-class* article from Wikipedia, the free encyclopedia  
U.S. computer scientist (Edit)

*Not to be confused with **Katie Bowmans***

**Katherine Louise Bouman** (/ˈbaʊmən/<sup>[1]</sup> born 1989/1990<sup>[2]</sup>) is an American computer scientist working in the field of computer imagery.

She led the development of an algorithm for imaging black holes, known as **Continuous High-resolution Image Reconstruction using Patch priors** (CHIRP), and was a member of the **Event Horizon Telescope** team that captured the first image of a black hole.<sup>[3][4]</sup>

As of June 2019, she is an assistant professor of computing and mathematical sciences at the **California Institute of Technology**.<sup>[5][6]</sup> [7][8]

### Contents [hide]

- Early life and education
- Research and career
- References
- External links

## Early life and education [edit]

Bouman grew up in **West Lafayette**, Indiana, and graduated from **West Lafayette Junior-Senior High School** in 2007. Her father, **Charles Bouman**, is a professor of **electrical and computer engineering** and biomedical engineering at **Purdue University**.<sup>[9]</sup> As a high school student, she conducted imaging research at **Purdue University**.<sup>[9]</sup> She first learned about the **Event Horizon Telescope** in school in 2007.<sup>[10]</sup>

Bouman studied electrical engineering at the **University of Michigan** and graduated **summa cum laude** in 2011. She earned her master's degree (2013) and doctoral degree (2017) in electrical engineering and computer science from the **Massachusetts Institute of Technology** (MIT).<sup>[11]</sup>

At MIT, she was a member of the **Haystack Observatory**.<sup>[12][13]</sup> She was supported by a **National Science Foundation** Graduate Fellowship. Her master's thesis, *Estimating Material Properties of Fabric through the Observation of Motion*,<sup>[14]</sup> was awarded the **Ernest Guillermo Award for best Master's Thesis** in electrical

## Katie Bouman



Katie Bouman speaks about the Event Horizon Telescope in 2019

Katherine Louise Bouman  
1989/1990 (age 29–30)

**Education**  
**Massachusetts Institute of Technology**  
**University of Michigan**

**Known for**  
**CHIRP algorithm**

### Scientific career

Computer vision · machine learning

**Institutions**  
**Massachusetts Institute of Technology**  
**Harvard University**

**Thesis**  
*Extreme Imaging via Physical Model Inversion: Seeing Around Corners and Imaging Black Holes* @

**Doctoral advisor**  
William T. Freeman

**Website**  
www.cms.caltech.edu/people/kbouman @

### External video

How to take a picture of a black hole @, Katie Bouman,

# Research:Content gaps on Wikipedia/Matrix Sample content gap matrix - Gender gap

	Selection	Extent	Framing
<b>Internal comparison</b>	There are more biographies of men than women on English Wikipedia (Klein et al. 2016)	Biographies of women scientists are lower quality, on average, than Wikipedia articles as a whole (Halfaker 2017)	The lead sections of biographies of women use terms like “divorce”, “family” and “spouse” more than biographies of men (Wagner et al. 2016)
<b>External comparison</b>	Women scholars with high h-index are less likely to have Wikipedia articles compared to men scholars with comparable h-index (Schellekens et al. 2019)	Wikipedia biographies of notable men and women are both longer and more equal in length, on average, than biographies of those same people in <i>Encyclopedia Britannica</i> . (Reagle and Rhou 2011)*	Articles about professions on German Wikipedia are substantially more likely to include pictures of men than women, regardless of the relative proportion of men and women in that profession according to labor statistics. (Zagovora et al. 2017)
<b>Reader needs</b>	<i>Hypothetical: Wikipedia articles are less likely to show up in the first page of results of Google searches by women vs. searches by men.</i>	Top hits from Wikipedia’s internal search for topic keywords related to women’s interest (sourced from a magazine corpus) were more likely to be redirects than keywords sourced from mens’ magazines (Menking et al. 2017)	<i>Hypothetical: Women seeking health information on Wikipedia are less satisfied with the way medical information is presented than are men seeking health information on Wikipedia.</i>



# Research:Content gaps on Wikipedia/Matrix Sample content gap matrix - Spanish Civil War

	Selection	Extent	Framing
<b>Internal comparison</b>	Spanish Wikipedia contains more articles about the Spanish Civil War than English Wikipedia.	Articles about the Spanish Civil War on Spanish Wikipedia contain more images than articles about the Spanish Civil War on English Wikipedia.	Articles about Francisco Franco on Spanish Wikipedia present his rise to power in a more sympathetic way than similar articles on English Wikipedia.
<b>External comparison</b>	Wikipedia contains fewer articles about important events in the Spanish Civil War than Salvado's <i>Historical Dictionary of the Spanish Civil War</i> .	Wikipedia articles about important events in the Spanish Civil War are shorter on average than sections devoted to these events in Salvado's <i>Historical Dictionary of the Spanish Civil War</i> .	Spanish Wikipedia articles about the Spanish Civil War uses more patriotic images, and fewer images of violence and devastation, than a recent museum exhibit about the war.
<b>Reader needs</b>	Articles about the Spanish Civil War that receive a high volume of page views on Spanish Wikipedia are under-represented on other Wikipedia languages.	Articles about the Spanish Civil War in English Wikipedia overwhelmingly cite Spanish-language sources.	Articles about the Spanish Civil War are not heavily linked to articles about the Catalan Independence Movement.



# Appendix

- MisinfoCon@NASEM:  
<https://misinfocon.com/misinfocon-nasem-is-here-misinfocon-7-0-is-focused-on-health-and-science-misinformation-71baa93d8717>
- MisinfoCon panel page:  
[https://en.wikipedia.org/wiki/Wikipedia:National\\_Academy\\_of\\_Sciences,\\_Engineering,\\_and\\_Medicine/quality\\_2020](https://en.wikipedia.org/wiki/Wikipedia:National_Academy_of_Sciences,_Engineering,_and_Medicine/quality_2020)
- Other:
  - [https://en.wikipedia.org/wiki/Wikipedia:Village\\_pump\\_\(miscellaneous\)#Percentage\\_of\\_editors\\_vs.\\_visitors/readers\\_in\\_English\\_Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:Village_pump_(miscellaneous)#Percentage_of_editors_vs._visitors/readers_in_English_Wikipedia)
  - [https://meta.wikimedia.org/wiki/Research:Characterizing\\_Wikipedia\\_Reader\\_Behaviour/Demographics\\_and\\_Wikipedia\\_use\\_cases](https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour/Demographics_and_Wikipedia_use_cases)
  - [https://en.wikipedia.org/wiki/Wikipedia:Village\\_pump\\_\(miscellaneous\)#Percentage\\_of\\_editors\\_vs.\\_visitors/readers\\_in\\_English\\_Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:Village_pump_(miscellaneous)#Percentage_of_editors_vs._visitors/readers_in_English_Wikipedia)
  - [https://meta.wikimedia.org/wiki/Gender\\_Diversity\\_Mapping](https://meta.wikimedia.org/wiki/Gender_Diversity_Mapping)
  - [https://meta.wikimedia.org/wiki/Gender\\_equity\\_report\\_2018](https://meta.wikimedia.org/wiki/Gender_equity_report_2018)
  - [https://medium.com/@marianarra\\_/women-in-the-wikimedia-movement-roles-culture-and-opportunities-c1092b0b924](https://medium.com/@marianarra_/women-in-the-wikimedia-movement-roles-culture-and-opportunities-c1092b0b924)
  - [https://upload.wikimedia.org/wikipedia/commons/c/c0/WMF\\_StrategicPlan2011\\_spreads.pdf](https://upload.wikimedia.org/wikipedia/commons/c/c0/WMF_StrategicPlan2011_spreads.pdf)
  - [https://meta.wikimedia.org/wiki/Community\\_Insights/2018\\_Report#Diversity\\_of\\_contributors\\_on\\_the\\_Wikimedia\\_projects\\_seems\\_to\\_remain\\_unchanged](https://meta.wikimedia.org/wiki/Community_Insights/2018_Report#Diversity_of_contributors_on_the_Wikimedia_projects_seems_to_remain_unchanged).
  - <http://whgi.wmflabs.org/gender-by-language.html>
  - <https://www.denelezh.org/gender-gap/>
  - [https://meta.wikimedia.org/wiki/Research:Content\\_gaps\\_on\\_Wikipedia/Matrix](https://meta.wikimedia.org/wiki/Research:Content_gaps_on_Wikipedia/Matrix)

# WikiCommons

- [https://commons.wikimedia.org/wiki/File:Word\\_cloud\\_Wikimedia\\_Recommendations\\_Feb\\_2020.JPG](https://commons.wikimedia.org/wiki/File:Word_cloud_Wikimedia_Recommendations_Feb_2020.JPG) (Martina Nolte, PD-text)
- [https://commons.wikimedia.org/wiki/File:Wikipedia\\_Offline\\_-\\_Education\\_Tool\\_March92012.pdf](https://commons.wikimedia.org/wiki/File:Wikipedia_Offline_-_Education_Tool_March92012.pdf) (Jwild, CC-SA-3.0)
- [https://commons.wikimedia.org/wiki/File:Tulu\\_Wikipedia\\_editathon\\_Mangaluru\\_Dec\\_07-08\\_2019\\_26.jpg](https://commons.wikimedia.org/wiki/File:Tulu_Wikipedia_editathon_Mangaluru_Dec_07-08_2019_26.jpg) (Pavanaja, CC-SA-4.0)
- [https://commons.wikimedia.org/wiki/File:Climate\\_at\\_the\\_National\\_Academies\\_Wikipedia\\_Edit-a-thon\\_3057.jpg](https://commons.wikimedia.org/wiki/File:Climate_at_the_National_Academies_Wikipedia_Edit-a-thon_3057.jpg) (Avery Jensen, CC-SA-4.0)