

# The AI Ethics Revolution

## A Brief History

Milestones – 2011 to 2023



**Edited by Mia Shah-Dand, Founder, Women in AI Ethics™**

**Originally published on April 4, 2023**

***[Last updated January 1, 2024]***

**Photo by [Miguel Bruna](#) on [Unsplash](#)**



## Introduction

The myth that “genius” men in Silicon Valley are solely responsible for advancements in AI has been repeatedly debunked as the contributions of hidden figures in the world of computing became more visible. It is now widely acknowledged that [Ada Lovelace wrote the first computer program](#) in the 1800s, [Joan Clarke](#) used her cryptology skills to help western allies during World War 2, and “(human) computers” like [Katherine Johnson](#) overcame racial segregation and used their mathematical genius to send the first American into space in the 1900s.

As the techno-elites in Silicon Valley pontificate about the future existential risks of AI, many pioneering women, non-binary people, and others from marginalized communities around the world are fighting an uphill battle to keep humanity safe from the harms of recklessly developed and deployed AI, while also ensuring that its benefits are distributed equitably.

Women and non-binary people have made consequential contributions to the AI Ethics space long before it was officially recognized as an emerging yet critical field of study. The first “[100 Brilliant Women in AI Ethics™](#)” list was published in 2018 and since then, Women in AI Ethics™ has published a list every year to recognize experts toiling away in obscurity.

Given the recent developments, it is even more critical to elevate voices being overshadowed by the AI hype boosted by ratings-hungry media. This timeline of women’s contributions to AI Ethics is kept updated with submissions from our community.

Please [use this form](#) to submit other milestones so we can collectively ensure that we continue to elevate significant work by underrepresented experts in this space.

In solidarity,

*Mia Shah-Dand*

*Founder, Women in AI Ethics™*

## AI Ethics Milestones - 2011 to 2023

### 2011

Karèn Fort and colleagues wrote an important paper on the risks of exploitation in crowd work “[Last Words: Amazon Mechanical Turk: Gold Mine or Coal Mine?](#)” and a position paper on Amazon Mechanical Turk platforms in French “[Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé](#)” to inform researchers, so that they make their choice with full knowledge of the facts, and propose practical and organizational solutions to improve the development of new linguistic resources by limiting the risks of ethical and legal abuses.

### 2014

The Francophone Natural Language Processing (NLP) research community were the first to host events such as this [workshop on ethics and NLP](#).

Cynthia Dwork co-authored the paper, “[The Algorithmic Foundations of Differential Privacy](#)”, to address the need for a robust, meaningful, and mathematically rigorous definition of privacy, together with a computationally rich class of algorithms as technology enabled ever more powerful collection of electronic data about individuals.

[Danielle Citron](#) co-authored “[The Scored Society: Due Process for Automated Predictions](#)”, a research paper that made the case for formalized safeguards to ensure fairness and accuracy of predictive algorithms.

### 2016

[Julia Angwin and ProPublica team](#) published their study on risk scores, as part of a larger examination of the powerful, largely hidden effect of algorithms in American life, and found significant racial disparities in sentencing recommendations.

[Cathy O’Neil’s book “Weapons of Math Destruction”](#) was instrumental in raising public awareness about the societal harms of algorithms and introducing it into mainstream conversations. In 2019, O’Neil’s [book won the Euler Book Prize](#), awarded annually to the author of an outstanding book about mathematics.

The [Gender Shades Project](#) began in 2016 as the focus of Joy Buolamwini’s [MIT master’s thesis](#) inspired by [her struggles with face detection systems](#).

### 2017

Shannon Spruit, Margaret Mitchell, Emily M. Bender, Hanna Wallach, and their colleagues [hosted the first ACL Workshop on Ethics in Natural Language Processing](#).

Historian Mar Hicks published “[Programmed Inequality](#)”, a documented history of how Britain lost its early dominance in computing by systematically discriminating against its most qualified workers: women.

[AI Now 2017 Report](#) published by Meredith Whittaker and Kate Crawford identified emerging challenges, made recommendations to ensure that the benefits of AI will be shared broadly, and that risks can be identified and mitigated.

Lina Khan’s article, [Amazon’s Antitrust Paradox](#) was published in the Yale Law Journal, where she noted, “Though relegated to technocrats for decades, antitrust and competition policy have once again become topics of public concern.”

## 2018

On January 15, Joy Buolamwini and Timnit Gebru presented “[Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification](#)” paper at the Conference on Fairness, Accountability, and Transparency.

In her extensively researched book, “[Algorithms of Oppression](#)”, [Safiya Umoja Noble](#) challenged the idea that search engines like Google offer an equal playing field for all forms of ideas, identities, and activities.

[Lucy Suchman and Lilly Irani](#) were among hundreds of Google employees and academics who penned an [open letter](#) calling upon Google to end its work on Project Maven, a controversial US military effort to develop AI to analyze surveillance video, and to support an international treaty prohibiting autonomous weapons systems. As of June 25, the letter received 1179 signatures. [Google later revised](#) its stance on Project Maven and announced [new AI principles](#) to govern the company’s future involvement in military projects.

On October 5, [Margaret Mitchell](#), [Inioluwa Deborah Raji](#), [Timnit Gebru](#), and other researchers introduced, “[Model Cards for Model Reporting](#)”, frameworks to document trained machine learning models as a step towards responsible democratization of machine learning and related AI technology by increasing transparency into how well AI technology works.

[Carole Cadwalladr](#), investigative journalist and features writer exposed the Facebook–Cambridge Analytica data scandal, where personal data belonging to millions of Facebook users was collected without their consent by the British consulting firm predominantly to be used for political advertising. Before Cambridge Analytica closed operations in 2018, the [company took legal action against The Observer](#) to stop them from publishing claims made in Cadwalladr’s articles.

[Virginia Eubanks’ book, “Automating Inequality”](#), investigated impacts of data mining, policy algorithms, and predictive risk models on poor and working class to find that algorithms used in public services often failed the most vulnerable in America.

The first [“100 Brilliant Women in AI Ethics™”](#) was published by Mia Shah-Dand to recognize and highlight the contributions of women to this nascent critical field.

## 2019

[Joy Buolamwini testified](#) in front of U.S. Congress on the limitations of facial recognition technology and urged Congress to adopt a moratorium prohibiting law enforcement use of face recognition or other facial analysis technologies.” She co-authored [a follow up paper](#) with Inioluwa Deborah Raji to analyze the impact of publicly naming and disclosing performance results of biased AI systems.

Walkout organizers Meredith Whittaker and Claire Stapleton revealed that they were [demoted or reassigned as retaliation](#) for their role in worker protests at Google in 2018.

Created in 2007 by researchers at Stanford and Princeton, [Imagenet](#) is considered the gold standard for image databases. Kate Crawford co-led the Imagenet Roulette project, which [uncovered racist and sexist content](#) and led to [removal of more than half a million images](#) from the database.

Mary L. Gray co-wrote [“Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass”](#), which reveals that much of the “intelligence” in tech is the result of exploitation of a vast, invisible human labor force.

Eight years of research by Sarah T. Robertson culminated in [“Behind the Screen: Content Moderation in the Shadows of Social Media”](#), which extensively documents the emotional toll of content moderation on a barely-paid invisible workforce.

Shoshana Zuboff’s [“The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power”](#) suggests that Google and Amazon’s business models represent a new form of “surveillance capitalism.”

In her book, [“Race After Technology: Abolitionist Tools for the New Jim Code”](#), Ruha Benjamin exposes how a range of discriminatory designs encode inequity by explicitly amplifying racial hierarchies.

## 2020

Google fired Timnit Gebru because of the highly acclaimed paper (published in March 2021), “[On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?](#)”, which she co-authored with Margaret Mitchell, Emily Bender, and Angelina McMillan-Major. The paper highlighted possible risks associated with large machine learning models and suggested exploration of solutions to mitigate those risks. After her firing, Dr. Gebru was the [target of a long vicious harassment campaign](#), bolstered by anonymous accounts on Twitter (later renamed X).

Abeba Birhane and Vinay Prabhu, published their research paper “[Large image datasets: A pyrrhic win for computer vision?](#)” that found massive datasets used to develop thousands of AI algorithms and systems contained racist and misogynistic labels and slurs as well as offensive images. [MIT](#) as well as others [including Microsoft took down databases](#) meant for researchers, but their [data still lingers online](#).

After initially disputing the findings of researchers, [Amazon halted police use of its facial recognition technology](#). That same year, IBM’s first CEO of color, Arvind Krishna announced the [company was getting out of the facial recognition business](#).

Facebook announced the [launch of its very first Oversight Board](#), 2 years after the Cambridge Analytica scandal. In response, a group of industry experts including Shoshana Zuboff, Carole Cadwalladr, Safiya Noble, Ruha Benjamin, and others launched “[Real Facebook Oversight Board](#)” to analyze and critique Facebook’s content moderation decisions, policies and other platform issues.

Sasha Costanza-Chock published “[Design Justice: Community-Led Practices to Build the Worlds We Need](#)”, a powerful exploration of how design might be led by marginalized communities, dismantle structural inequality, and advance collective liberation and ecological survival.

Sasha (Alexandra) Luccioni developed “Machine Learning Emissions Calculator”, a tool for estimating the carbon impact of machine learning processes, and presented it along with related issues and challenges in this article, “[Estimating the Carbon Emissions of Artificial Intelligence](#).”

## 2021

Google [fired its ethical AI team co-lead Margaret Mitchell](#) who along with Timnit Gebru had called for more diversity among Google’s research staff and expressed concern that the company was starting to censor research critical of its products.

On the one-year anniversary of her ouster, Gebru [launched Distributed Artificial Intelligence Research institute \(DAIR\)](#) — an independent, community-rooted institute set to counter Big Tech’s pervasive influence on research, development, and deployment of AI.

Emily Denton, Alex Hanna, and their fellow researchers published “[On the genealogy of machine learning datasets: A critical history of ImageNet](#)”, to propose reflexive interventions during dataset development, the initial ideation/design stage, the creation and collection stages, and subsequent maintenance and storage stages.

[Lina Khan was appointed as Federal Trade Commission \(FTC\) Chair](#) and announced [several new additions](#) to the FTC’s Office of Policy Planning including Meredith Whittaker, Amba Kak, and Sarah Myers West as part of an informal AI Strategy Group.

The [FTC issued a report to Congress](#) warning about using AI to combat online problems and urging policymakers to exercise “great caution” about relying on it as a policy solution.

Whistleblower Frances Haugen, a former data scientist at Facebook, [testified at a U.S. Senate Hearing](#) that “Facebook’s algorithm amplified misinformation” and “it consistently chose to maximize its growth rather than implement safeguards on its platforms.”

California governor signed the “[Silenced No More Act](#)”, authored by state senator Connie Leyva and initiated by [Pinterest whistleblower, Ifeoma Ozoma](#) to protect workers who speak out about harassment and discrimination even if they’ve signed a non-disclosure agreement (NDA). Ozoma also launched a free online resource guide for tech whistleblowers, called “[The Tech Worker Handbook](#).”

## **2022**

After a year of engagement with the public and discussions with AI experts, Dr. Alondra Nelson and President Biden’s White House released [the Blueprint for an AI Bill of Rights](#).

Cori Crider and her team at Foxglove, an NGO based in London, announced their [support for a major new](#) case demanding fundamental change to Facebook’s algorithm, prioritizing the safety of users who live in Eastern and Southern Africa. If successful, this would be the first time a case has forced changes to Facebook’s algorithm. Crider’s group has also [worked with Facebook content moderators](#) since 2019 to help end their exploitation and [put together a briefing](#) to debunk Amazon’s myths about its unsafe warehouses and management by algorithm.

## 2023

Hilke Schellmann co-led [a Guardian investigation of AI algorithms](#) used by social media platforms and found many of these algorithms have a gender bias, and may have been censoring and suppressing the reach of photos of women's bodies.

Authors of Stochastic Parrots [responded to an "AI pause" open letter](#) stating that, "The harms from so-called AI are real and present and follow from the acts of people and corporations deploying automated systems. Regulatory efforts should focus on transparency, accountability and preventing exploitative labor practices."

[Under President Fran Drescher's leadership](#), the American actors' union SAG-AFTRA (Screen Actors Guild – American Federation of Television and Radio Artists) walked out in mid-July over a labor dispute with the Alliance of Motion Picture and Television Producers. When the talks broke down with the trade association, Drescher is quoted as saying, "I cannot believe it, quite frankly, how far apart we are on so many things. How they plead poverty. That they're losing money left and right when giving hundreds of millions of dollars to their CEOs. It is disgusting. Shame on them."

[On July 12, 2023, Artist Karla Ortiz](#) testified before the US Senate Judiciary Committee, sharing concerns about AI and intellectual property and stated, "The unfettered growth of generative AI without appropriate guardrails would lead to an unprecedented escalation and use of these technologies in not only creative spaces, but in businesses across all sectors. This could potentially devastate the contributions of workers and the workforce."

[Jingna Zhang founded Cara](#), a social media and portfolio-sharing platform for artists and art enthusiasts that filters out gen AI images to make it easier for people to find authentic creatives and images. On December 1, 2023, Cara introduced support for Glaze, a tool developed by SAND Lab at the University of Chicago, which artists could use to deter style mimicry by AI image generators.

To be continued...

-----X-----

Please [fill out this form](#) if you would like to add a new milestone to this timeline.