Ethics in AI research papers and articles

By: Kathy Baxter

Photo by Annie Spratt on Unsplash

Last updated: July 15, 2021

This is my obsessively curated list of research papers and articles on ethics in AI that I have been collecting over the years. Let me know if I am missing your favorites.

2021

Peer-reviewed

Racial/Ethnic Disparities in the Performance of Prediction Models for Death by Suicide After Mental Health Visits (7/1/21)

The Limits of Differential Privacy (and Its Misuse in Data Release and Machine Learning) (July 2021)

The Values Encoded in Machine Learning Research (6/29/21)

A Silicon Valley Love Triangle: Hiring Algorithms, Pseudo-Science, and the Quest for Auditability (6/23/21)

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients (6/21/21)

Are we done with ImageNet? (6/12/21)

Intrinsic Bias Metrics Do Not Correlate with Application Bias (6/8/21)

Measuring Model Biases in the Absence of Ground Truth (6/6/21)

Reliability Testing for Natural Language Processing Systems (6/1/21)

Explaining the Principles to Practices Gap in AI (June 2021)

Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities (5/24/21)

Online Selection of Diverse Committees (5/19/21)

Towards accountability in the use of Artificial Intelligence for Public Administrations (5/18/21)

Including Signed Languages in Natural Language Processing (5/11/21)

A Step Toward More Inclusive People Annotations for Fairness (5/5/21)

To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes (May 2021)

[Denmark] "We Would Never Write That Down": Classifications of Unemployed and Data Challenges for AI (April 2021)

Precarity: Modeling the Long Term Effects of Compounded Decisions on Individual Instability (4/24/21)

Context-sensitive evaluation of automatic speech recognition: considering user experience & language variation (4/20/21)

Engineering Bias Out of AI (4/20/21)

Moral zombies: why algorithms are not moral agents (4/16/21)

Autonomous Driving Ethics: from Trolley Problem to Ethics of Risk
(4/12/21)

How medicine discriminates against non-white people and women
(4/10/21)

Moving beyond "algorithmic bias is a data problem" (4/9/21)

Evaluating eligibility criteria of oncology trials using real-world data and
AI (4/7/21)

How artificial intelligence could make clinical trials smarter (4/7/21)

The Myth of Complete AI-Fairness (4/6/21)

How medical AI devices are evaluated: limitations and
recommendations from an analysis of FDA approvals (4/5/21)

SENSEI: SENSITIVE SET INVARIANCE FOR ENFORCING INDIVIDUAL
FAIRNESS (4/1/21)

The (Im)possibility of Fairness: Different Value Systems Require
Different Mechanisms For Fair Decision Making (Apr 2021)

INDIVIDUALLY FAIR GRADIENT BOOSTING (3/31/21)

STATISTICAL INFERENCE FOR INDIVIDUAL FAIRNESS (3/30/21)

Can an AI Algorithm Mitigate Racial Economic Inequality? An Analysis
in the Context of Airbnb (3/22/21)

Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets
(3/22/21)

Lawyers are Dishonest? Quantifying Representational Harms in Commonsense Knowledge Resources (3/21/21)

INDIVIDUALLY FAIR RANKING (3/19/21)

Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? (3/16/21)

Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans (3/15/21)

People Analytics must benefit the people. An ethical analysis of data-driven algorithmic systems in human resources management (3/12/21)

The race to teach sign language to computers (3/6/21)

Measuring Model Biases in the Absence of Ground Truth (3/5/21)

Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law (3/3/21)

Machines, Artificial Intelligence and rising global transphobia (3/1/21)

Narratives and Counternarratives on Data Sharing in Africa (3/1/21)

Artificial intelligence has a problem with grammar (2/27/21)

Sustainable AI: AI for sustainability and the sustainability of AI (2/26/21)

Benchmarking and Survey of Explanation Methods for Black Box Models (2/25/21)

Ethics-Based Auditing to Develop Trustworthy AI (2/19/21)

The Privatization of AI Research(-ers): Causes and Potential Consequences (2/15/21)

Algorithmic injustice: a relational ethics approach (2/12/21)

The Limits of Global Inclusion in AI Development (2/7/21)

Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models (2/4/21)

Insiders and Outsiders in Research on Machine Learning and Society (2/3/21)

About Face: A Survey of Facial Recognition Evaluation (2/1/21)

Adversarial Learning with Cost-Sensitive Classes (1/29/21)

Hiding Behind Machines: When Blame Is Shifted to Artificial Agents (1/27/21)

Re-imagining Algorithmic Fairness in India and Beyond (1/25/21)

Black Feminist Musings on Algorithmic Oppression (1/25/21)

It's Too Easy to Hide Bias in Deep-Learning Systems (1/19/21)

Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases (1/19/21)

Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis (1/18/21)

Persistent Anti-Muslim Bias in Large Language Models (1/14/21)

AI Ethics: A Long History and a Recent Burst of Attention (1/14/21)

Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19 (Jan 2021)

The Role of Arts in Shaping AI Ethics (2021)

Government, NGO, and expert publications

America's global leadership in human-centered AI can't come from industry alone (7/6/21)

Demystifying the Draft EU Artificial Intelligence Act (July 2021)

Research Summary: Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Within the Next Decade (6/29/21)

WHO issues first global report on Artificial Intelligence (AI) in health and six guiding principles for its design and use (6/28/21)

Mobilising the intellectual resources of the arts and humanities (6/25/21)

[China] Lofty principles, conflicting incentives: AI ethics and governance in China (6/24/21)

Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Within the Next Decade (6/16/21)

A Step Toward More Inclusive People Annotations in the Open Images Extended Dataset (6/15/21)

[videos] NIST AI Measurement and Evaluation Workshop (6/15/21)

3 GUIDING PRINCIPLES TO KICK-START YOUR AI INITIATIVE (6/15/21)

A Step Toward More Inclusive People Annotations in the Open Images Extended Dataset (6/15/21)

The role of the arts and humanities in thinking about artificial intelligence (AI) (6/14/21)

10 steps to educate your company on AI fairness (6/9/21)

Karlsruhe win against biometric mass surveillance in Germany (6/8/21)

The current state of affairs and a roadmap for effective carbon-accounting tooling in AI (6/7/21)

Managing the risk in AI: Spotting the "unknown unknowns" (6/5/21)

The Contestation of Tech Ethics: A Sociotechnical Approach to Ethics and Technology in Action (6/3/21)

Facial Recognition Technology: Federal Law Enforcement Agencies Should Better Assess Privacy and Other Risks (6/3/21)

[EU] Guidance on Strengthening the Code of Practice on Disinformation (6/2/21)

Automated Decision-Making Systems in the Public Sector An Impact Assessment Tool for Public Authorities (June 2021)

AI GOVERNANCE IN 2020: A YEAR IN REVIEW: OBSERVATIONS FROM 52 GLOBAL EXPERTS (June 2021)

NIST: A Proposal for Identifying and Managing Bias in Artificial Intelligence (June 2021)

AI for disability inclusion: Enabling change with advanced technology (May 2021)

AI Ethics Has a Surplus Problem (5/26/21)

[Spain] In Catalonia, the RisCanvi algorithm helps decide whether inmates are paroled (5/25/21)

Understanding Contextual Facial Expressions Across the Globe (5/24/21)

People analytics in the workplace – how to effectively enforce labor rights (5/19/21)

Tech for disabled people is booming around the world. So where's the funding? (5/19/21)

Data Capitalism (2021)

The Chief AI Ethics Officer: A Champion or a PR Stunt? (5/16/21)

Image classification algorithms at Apple, Google still push racist tropes (5/14/21)

US House AI Task Force Is the latest authority to address algorithms and racism (5/12/21)

Detecting and mitigating bias in natural language processing (5/10/21)

[Italy] In Italy, general practitioners and some regions adopt COVID-19 vaccine prioritization algorithms (5/19/21)

Can a COVID-19 face mask protect you from facial recognition technology too? (5/10/21)

New research helps make AI fairer in decision-making (5/7/21)

[France] Evaluation Methods for Content Recommendation Algorithms (5/6/21)

Hard choices: AI in health care (2021)

The hidden work created by artificial intelligence programs (5/4/21)

Reclaim Your Face – A European Citizens Initiative to ban biometric mass surveillance (5/4/21)

[Greece] Greek camps for asylum seekers to introduce partly automated surveillance systems (4/27/21)

Will Artificial Intelligence Foster or Hamper the Green New Deal? (4/22/21)

[UK] Home Office algorithm to detect sham marriages may contain built-in discrimination (4/19/21)

Aiming for truth, fairness, and equity in your company's use of AI (4/19/21)

Face Recognition Is Far from the Sci-Fi Super-Tool Its Sellers Claim (4/16/21)

61 MEPs urge the EU to ban biometric mass surveillance (4/16/21)

[France] How French welfare services are creating 'robo-debt' (4/15/21)

[Singapore] Are we stuck with Covid-19 surveillance? – Join the democracy classroom (4/12/21)

The Abuse and Misogynoir Playbook, explained (workshop transcript) (4/8/21)

Emails between NYPD and Clearview AI obtained by FOIL request prompt questions (4/6/21)

Research Summary: Mapping the Ethicality of Algorithmic Pricing (4/5/21)

Identity systems and social protection in Venezuela and Bolivia: GENDER IMPACTS AND OTHER INEQUALITIES (4/5/21)

Research Summary: Algorithmic Impact Assessments – What Impact Do They Have? (4/4/21)

If Your Company Uses AI, It Needs an Internal Review Board (4/1/21)

Error-riddled data sets are warping our sense of how good AI really is
(4/1/21)

[UK] Technology Managing People – the legal implications (Mar 2021)

Six Unexamined Premises Regarding Artificial Intelligence and National
Security (3/31/21)

Are You Overestimating Your Responsible AI Maturity? (3/30/21)
Responsible AI: From principles to practice (3/30/21)

[China] Research Summary: The Chinese Approach to AI: An Analysis
of Policy, Ethics, and Regulation (3/29/21)

The IEEE Trusted Data and Artificial Intelligence Systems (AIS)
Playbook for Financial Services (3/29/21)

Using Clinical Text to Combat Selection Bias in Medical Research
(3/29/21)

From Principles to Practice: An interdisciplinary framework to
operationalize AI ethics (2020)

How AI Can Help Companies Set Prices More Ethically (3/26/21)

[UK] Technology Managing People – the legal implications (3/25/21)

Alation State of Data Culture Report Reveals Barriers in Adopting
Artificial Intelligence (3/24/21)

EUGENICS POWERS IQ AND AI (3/24/21)

A Better Measuring Stick: Algorithmic Approach to Pain Diagnosis Could Eliminate Racial Bias (3/24/21)

Research Summary: Artificial Intelligence and the Privacy Paradox of Opportunity, Big Data and The Digital Universe (3/21/21)

Mapping India's AI Potential (3/21/21)

Hiring by Algorithm: Legal Issues Presented by the Use of Artificial Intelligence in Sourcing and Selection (3/17/21)

The Algorithmic Auditing Trap (3/17/21)

Should AI Models Be Explainable? That depends. (3/16/21)

How AI is Changing Our Understanding of Language (3/16/21)

Who Is Governing AI Matters Just as Much as How It's Designed (3/15/21)

10 Takeaways from the State of AI Ethics in Canada & Spain (3/15/21)

How Can We Better Regulate AI Companies? (3/15/21)

Myanmar: Facial Recognition System Threatens Rights (3/12/21)

Auditing employment algorithms for discrimination (3/12/21)

Fairness On The Ground: Applying Algorithmic Fairness Approaches To Production Systems (3/11/21)

Assessing Regulatory Fairness Through Machine Learning (3/10/21)

[Rome] The Ethics of AI to Ensure Food Security and Development (3/8/21)

[Britain] Malgorithm: How Instagram's algorithm publishes misinformation and hate to millions during a pandemic. (3/8/21)

If It's Free, You're the Product: The New Normal in a Surveillance Economy (3/7/21)

Corsight AI addresses face biometrics ethics in Asia Pacific expansion (3/5/21)

Where are the women? Mapping the gender job gap in AI Policy Briefing – Full Report (Mar 2021)

Stanford HAI: AI Index Report 2021 (Mar 2021)

KPMG: Trust in Artificial Intelligence: A five country study (March 2021)

WEF: A 5-step guide to scale responsible AI (3/1/21)

RESPONSIBLE AI #AIFORALL: Approach Document for India Part 1 – Principles for Responsible AI (Feb 2021)

WEF: Responsible Use of Technology: The Microsoft Case Study (2/25/21)

WEF: 4 lessons on designing responsible, ethical tech: Microsoft case study (2/25/21)

A Better Measuring Stick: Algorithmic Approach to Pain Diagnosis Could Eliminate Racial Bias (2/24/21)

We need to talk about Artificial Intelligence (2/22/21)

The Artificiality of AI – Why are We Letting Machines Manage Employees? (2/22/21)

Research Summary: The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms (2/21/21)

Why Organizations Should Care About Responsible AI & Digital Ethics (2/19/21)

Black voices bring much needed context to our data-driven society (2/18/21)

Privacy enhancing technologies for trustworthy use of data (2/9/21)

The Physician's Conundrum: Assigning Moral Responsibility for Medical Artificial Intelligence and Machine Learning (2/8/21)

Legislation Related to Artificial Intelligence

TRAINING AI SYSTEMS FOR USE IN DANGEROUS SITUATIONS (2/4/21)

As the FDA clears a flood of AI tools, missing data raise troubling questions on safety and fairness (2/3/21)

Governance by Algorithms (Research Summary) (2/1/21)

[Germany] 10 DIFFICULT TERMS TO TRANSLATE REGARDING RACE (Feb 2021)

MAEI The State of AI Ethics Report (Jan 2021)

Addressing Ethical Dilemmas in AI: Listening to Engineers (Jan 2021)

The CREO AI Creativity Report 2021 (Jan 2021)

NYC Summary of Agency Compliance Reporting (Jan 2021)

What algorithm auditing startups need to succeed (1/30/21)

EU Guidelines on Facial Recognition (1/28/21)

FDA Announces Action Plan for Oversight of AI/ML in Medical Devices (1/28/21)

Mapping AI in the Global South (1/26/21)

Beyond Engagement: Aligning Algorithmic Recommendations With Prosocial Goals (1/21/21)

The building blocks of Microsoft's responsible AI program (1/19/21)

Seven Legal Questions for Data Scientists (1/19/21)

To achieve Responsible AI, close the "believability gap" (1/18/21)

Civil society calls for AI red lines in the European Union's Artificial Intelligence proposal (1/12/21)

China's social credit system was due by 2020 but is far from ready (1/12/21)

California Company Settles FTC Allegations It Deceived Consumers about use of Facial Recognition in Photo Storage App (1/11/21)

How to Build a Likable Chatbot (1/11/21)

In Poland, a law made loan algorithms transparent. Implementation is nonexistent. (1/6/21)

Getting Specific About AI Risks (1/6/21)

In Focus: Facial Recognition Tech Stories and Rights Harms Around the World (Jan 2021)

AI Governance in Japan Ver. 1.0 (Jan 2021)

Popular Press

Promoting Trustworthy AI in Government (7/8/21)

OpenAI warns AI behind GitHub's Copilot may be susceptible to bias (7/8/21)

We tested AI interview tools. Here's what we found. (7/7/21)

How TikTok's hate speech detection tool set off a debate about racial bias on the app (7/7/21)

THE MOST "WOKE" COMPANY COULD CONTRIBUTE MOST TO ONLINE BIAS (7/5/21)

AI legislation must address bias in algorithmic decision-making systems (7/5/21)

COULD AI KEEP PEOPLE 'ALIVE' AFTER DEATH? (7/3/21)

Pandemic Wave of Automation May Be Bad News for Workers
(7/3/21)

What happened when a 'wildly irrational' algorithm made crucial
healthcare decisions (7/2/21)

Study finds that few major AI research papers consider negative
impacts (7/1/21)

The importance of having accountability in AI ethics (7/1/21)

Facial recognition tech has been widely used across the US
government for years, a new report shows (6/30/21)

Using A.I. to Find Bias in A.I. (6/30/21)

A Hippocratic Oath for your AI doctor (6/29/21)

Fired by Bot at Amazon: 'It's You Against the Machine' (6/28/21)

DeepMind AGI paper adds urgency to ethical AI (6/26/21)

NIST proposes ways to identify and address AI bias (6/25/21)

The World Needs Deepfake Experts to Stem This Chaos (6/24/21)

IBM explores AI tools to spot, cut bias in online ad targeting (6/24/21)

Combating racial bias in AI (6/23/21)

LinkedIn's job-matching AI was biased. The company's solution? More
AI. (6/23/21)

Google & DeepMind Researchers Revamp ImageNet (6/23/21)

Navigating a surprising pandemic side effect: AI whiplash (6/21/21)

Europe's privacy regulators call for a ban on facial recognition in publicly accessible spaces (6/21/21)

'Nobody is catching it': Algorithms used in health care nationwide are rife with bias (6/21/21)

A popular algorithm to predict sepsis misses most cases and sends frequent false alarms, study finds (6/21/21)

Google searches for new measure of skin tones to curb bias in products (6/18/21)

Facial Recognition Failures Are Locking People Out of Unemployment Systems (6/18/21)

Hey computer, am I okay? (6/18/21)

The Efforts to Make Text-Based AI Less Racist and Terrible (6/17/21)

Ethical AI will not see broad adoption by 2030, study suggests (6/17/21)

Amazon Alexa head scientist on developing trustworthy AI systems (6/16/21)

Facebook's latest AI doesn't just detect deep fakes, it knows where they came from (6/16/21)

Want a job? Employers say: Talk to the computer (6/14/21)

How and Why Enterprises Must Tackle Ethical AI (6/14/21)

Fact-Checks, Info Hubs, and Shadow-Bans: A Landscape Review of Misinformation Interventions (6/14/21)

Google's Privacy Backpedal Shows Why It's So Hard Not to Be Evil (6/14/21)

The secrets behind the plastic spoon: a 'perfect' design with terrible consequences (6/14/21)

How To Make Sure That Diversity In AI Works (6/14/21)

Human Cognitive Bias And Its Role In AI (6/14/21)

How to mitigate bias in AI (6/13/21)

Rules around facial recognition and policing remain blurry (6/12/21)

These creepy fake humans herald a new age in AI (6/11/21)

Biden administration forms new AI task force (6/10/21)

[China] Across China, AI 'city brains' are changing how the government runs (6/10/21)

TikTok changed the shape of some people's faces without asking (6/10/21)

EleutherAI claims new NLP model approaches GPT-3-level performance (6/9/21)

10 steps to educate your company on AI fairness (6/9/21)

Facial Verification Won't Fight Fraud (6/9/21)

First U.S. Artificial Intelligence Czar Seeks 'Responsible Use' of AI Tools (6/8/21)

Bias and discrimination in AI: whose responsibility is it to tackle them? (6/8/21)

Instagram pushes back against 'The Algorithm' – because it tracks every action you take on the app using 'a variety of algorithms' (6/8/21)

Investors call for ethical approach to facial recognition technology (6/8/21)

Skin in the game: Video chat apps tout 'inclusive' AI features (6/7/21)

That AI scanning your X-ray for signs of COVID-19 may just be looking at your age (6/7/21)

As AI develops, so does the debate over profits and ethics (6/4/21)

Federal facial recognition ban will be reintroduced 'soon,' Sen. Jeff Merkley says (6/4/21)

China's response to COVID showed the world how to make the most of A.I. (6/4/21)

TikTok just gave itself permission to collect biometric data on US users, including 'faceprints and voiceprints' (6/3/21)

'Care bots' are on the rise and replacing human caregivers (6/3/21)

Don't End Up on This Artificial Intelligence Hall of Shame (6/3/21)

Dark Patterns that Mislead Consumers Are All Over the Internet (6/3/21)

How Amazon became an engine for anti-vaccine conspiracy theories (6/3/21)

Amazon's Cost Saving Routing Algorithm Makes Drivers Walk Into Traffic (6/2/21)

McDonald's is testing automated drive-thru ordering at 10 Chicago restaurants (6/2/21)

Machine learning is booming in medicine. It's also facing a credibility crisis (6/2/21)

Senate Democrats Urge Google To Investigate Racial Bias In Its Tools And The Company (6/2/21)

FACIAL RECOGNITION TECHNOLOGY: Federal Law Enforcement Agencies Should Better Assess Privacy and Other Risks (June 2021)

A rogue killer drone 'hunted down' a human target without being instructed to, UN report says (5/30/21)

Skin in the frame: black photographers welcome Google initiative (5/28/21)

How a largely untested AI algorithm crept into hundreds of hospitals (5/28/21)

Anthropic is the new AI research outfit from OpenAI's Dario Amodei, and it has $124M to burn (5/28/21)

[Spain] Workers vs Algorithms (5/27/21)

The ethical use of data in artificial intelligence (5/27/21)

This $5 billion insurance company likes to talk up its AI. Now it's in a mess over it (5/27/21)

Scared, human? Emotion-detection AI meets eye-tracking technology (5/27/21)

[EU] Clearview AI hit with sweeping legal complaints over controversial face scraping in Europe (5/27/21)

[EU] The Grand Normalization of Mass Surveillance: ECtHR Grand Chamber Judgments in Big Brother Watch and Centrum för rättvisa (5/26/21)

Google Strikes Deal With Hospital Chain to Develop Healthcare Algorithms (5/26/21)

AI emotion-detection software tested on Uyghurs (5/26/21)

Report finds startling disinterest in ethical, responsible use of AI among business leaders (5/25/21)

65% of execs can't explain how their AI models make decisions, survey finds (5/25/21)

How 'blind learning' could solve A.I.'s 'Little Shop of Horrors' dilemma (5/25/21)

'AI' is being used to profile people from their head vibrations – but is there enough evidence to support it? (5/24/21)

Facial Recognition Is Racist. Why Aren't More Cities Banning It? (5/24/21)

AI Can Write Disinformation Now–and Dupe Human Readers (5/24/21)

AI didn't invent privacy abuse - how the history of data privacy informs our future (5/24/21)

Citizen's dystopian new feature is mass surveillance disguised as public safety (5/22/21)

Facebook AI cuts by more than half the error rate of unsupervised speech recognition (5/21/21)

AI's Future Doesn't Have to Be Dystopian (5/20/21)

Google's New Dermatology App Wasn't Designed for People With Darker Skin (5/20/21)

Publicis Media updates ad tech evaluation process to avoid cultural stereotypes, but getting information about algorithms 'is hard' (5/20/21)

Statistical bias in context - AI didn't invent quantitative methods of bias (5/20/21)

The race to understand the exhilarating, dangerous world of language AI (5/20/21)

Spotify, don't spy: global coalition of 180+ musicians and human rights groups take a stand against speech-recognition technology (5/19/21)

Twitter finds its AI tends to crop out Black people, men from photos (5/19/21)

Innovative free course empowers citizens to advocate for ethical AI (5/19/21)

The disinformation threat from text-generating AI (5/19/21)

[France] Environmentalists do not cut video surveillance but slow down its development (5/18/21)

Why companies should carefully read government A.I. plans (5/18/21)

Is it time for an algorithm bill of rights? These analysts think so (5/18/21)

Amazon extends moratorium on police use of facial recognition software (5/18/21)

Amazon extends ban on police use of its facial recognition technology indefinitely (5/18/21)

Evolving to a more equitable AI (5/18/21)

Ethical AI Is Our Responsibility (5/17/21)

The Impact of Dark Patterns on Communities of Color (5/17/21)

AIs are getting smarter, fast. That's creating tricky questions that we can't answer (5/17/21)

The Chief AI Ethics Officer: A Champion or a PR Stunt? (5/16/21)

Police departments adopting facial recognition tech amid allegations of wrongful arrests (5/16/21)

Critics raise alarm over Big Tech's most powerful tools (5/16/21)

In charts: facial recognition technology – and how much do we trust it? (5/16/21)

[New Zealand] Jacinda Ardern calls for 'ethical algorithms' to help stop online radicalisation (5/14/21)

We need to design distrust into AI systems to make them safer (5/13/21)

[UK] AI and robots to assess patients in NHS plan to tackle record waiting lists (5/13/21)

[China] Chinese Users Do Care About Online Privacy (5/12/21)

Instagram Censored Posts About One Of Islam's Holiest Mosques, Drawing Employee Ire (5/12/21)

Trustworthy AI versus ethical AI - what's the difference, and why does it matter? (5/11/21)

Expert.ai adds emotion, style detection tools to natural language API (5/10/21)

Blind people, advocates slam company claiming to make websites ADA compliant (5/9/21)

[Israel] Sheikh Jarrah: Activists raise concerns over deleted social media content (5/7/21)

Deepfake detectors and datasets exhibit racial and gender bias, USC study shows (5/6/21)

California's "Equity" Algorithm Could Leave 2 Million Struggling Californians Without Additional Vaccine Supply (5/6/21)

Proper data hygiene critical as enterprises focus on AI governance (5/6/21)

[Denmark] Ready for face recognition at the restaurant? (5/5/21)

AI Bias Problem Needs More Academic Rigor, Less Hype (5/4/21)

Anyone can use this powerful facial-recognition tool – and that's a problem (5/4/21)

AI bias is an ongoing problem, but there's hope for a minimally biased future (5/4/21)

Fighting algorithmic bias in artificial intelligence (5/4/21)

YOUR CAR IS SPYING ON YOU, AND A CBP CONTRACT SHOWS THE RISKS (5/3/21)

Shhhh, they're listening: Inside the coming voice-profiling revolution (5/3/21)

Appen combats skewed AI data to ensure end-users have the same experience (5/3/21)

Using AI to root out unconscious bias (5/1/21)

Ethics of AI: Benefits and risks of artificial intelligence (4/30/21)

[Russia] 'Deepfake' that supposedly fooled European politicians was just a look-alike, say pranksters (4/30/21)

[Portugal] "It is worrying for democracy." TAP uses "heartless" algorithm to fire (4/30/21)

King County government must turn its back on facial recognition technology (4/30/21)

Credit Card Ads Were Targeted by Age, Violating Facebook's Anti-Discrimination Policy (4/29/21)

Black Man Accuses AI of Being Racist After Passport Photo Is Rejected in Viral Video (4/28/21)

Kind Environments for technology organisations? [part two] (4/28/21)

N.Y.P.D. Robot Dog's Run Is Cut Short After Fierce Backlash (4/28/21)

Landmark AI legislation could tackle algorithmic bias (4/28/21)

IBM and Microsoft Have Integrated AI Ethical Standards into Their Operations, So Can You (4/28/21)

Congress drags algorithms out of the shadows (4/27/21)

Antiracism in AI: How to Build Bias Checkpoints Into Your Development and Delivery Process (4/27/21)

To Fight Social Media Disinformation, Look to the Algorithms (4/27/21)

Ex-Google employee: Big Tech's biz model is 'a society that is addicted, outraged, polarized' (4/27/21)

Artificial Intelligence Is Misreading Human Emotion (4/27/21)

[UAE] Sharjah Police drones use face-recognition technology to identify wanted criminals (4/26/21)

Intersectionality, Inequity And Imminence Of AI: A Human Perspective (4/26/21)

AI at work isn't always intelligent (4/26/21)

Despite acknowledged promise: Fear, uncertainty and doubt surround AI adoption (4/26/21)

'Make Algorithmic Audits As Ubiquitous As Seatbelts'–Why Tech Needs Outside Help To Serve Humanity (4/26/21)

AI that stops you getting mad at your customers? Are you crazy? (4/24/21)

Big tech's algorithms are not color blind (4/23/21)

Stop talking about AI ethics. It's time to talk about power. (4/23/21)

Algorithmic Nudges Don't Have to Be Unethical (4/22/21)

Stanford study questions how medical AI devices are evaluated (4/22/21)

Shadow Bans, Dopamine Hits, and Viral Videos, All in the Life of TikTok Creators (4/22/21)

The global race to regulate AI (4/22/21)

To ensure inclusivity, the Biden administration must double down on AI development initiatives (4/22/21)

Microsoft and Big Tech Can't Distance Themselves From the Police Violence They Fuel (4/21/21)

Europe throws down gauntlet on AI with new rulebook (4/21/21)

In scramble to respond to Covid-19, hospitals turned to models with high risk of bias (4/21/21)

6 key battles ahead for Europe's AI law (4/21/21)

Europe's Proposed Limits on AI Would Have Global Consequences (4/21/21)

Google Ethical AI Group's Turmoil Began Long Before Public Unraveling (4/21/21)

FTC issues stern warning: Biased AI may break the law (4/20/21)

'Detoxified' language models might marginalize minorities, says study (4/20/21)

AI vs. Maya Angelou: Experimental Evidence That People Cannot Differentiate AI-Generated From Human-Written Poetry (4/19/21)

Nextdoor will alert users if it thinks they're about to post something racist (4/19/21)

She's taking Jeff Bezos to task (4/19/21)

US banks deploy AI to monitor customers, workers amid tech backlash (4/19/21)

Hackers Used to Be Humans. Soon, AIs Will Hack Humanity (4/19/21)

Google translation AI botches legal terms 'enjoin,' 'garnish' -research (4/19/21)

Here's why we should never trust AI to identify our emotions (4/18/21)

Data and computer scientists, ecologists, pathologists and legal scholars study AI's biases (4/16/21)

[Sweden] World Moving Towards a "Devastating Marriage" of Artificial Intelligence & Weapons of War (4/16/21)

[Italy] Facial recognition: Sari Real Time does not comply with the privacy policy (4/16/21)

[Netherlands] Rotterdam Court of Audit: risk of biased outcomes due to the use of algorithms (4/15/21)

Amazon and Microsoft team up to defend against facial recognition lawsuits (4/15/21)

MEPs call for European AI rules to ban biometric surveillance in public (4/15/21)

AI's social justice problem: It's amplifying human bias (4/15/21)

Coded Bias: An Insightful Look At AI, Algorithms And Their Risks To Society (4/15/21)

AI is increasingly being used to identify emotions – here's what's at stake (4/15/21)

The Slow Violence of Emerging Technologies [part one] (4/14/21)

The new lawsuit that shows facial recognition is officially a civil rights issue (4/14/21)

The Computer Got it Wrong: Why We're Taking the Detroit Police to Court Over a Faulty Face Recognition 'Match' (4/13/21)

Why AI Is Failing for Enterprises: Predetermination Bias (4/12/21)

Revealed: the Facebook loophole that lets world leaders deceive and harass their citizens (4/12/21)

Society dictated your face, according to biometrics research and a patent (4/12/21)

Hitting the Books: How biased AI can hurt users or boost a business's bottom line (4/10/21)

Black women, AI, and overcoming historical patterns of abuse (4/10/21)

[Sweden] CKP's test use of a program intended for face recognition has been reported to the Data Protection Ombudsman (4/9/21)

[South Asia] From Lahore to Lucknow, crimes against women spur more surveillance (4/9/21)

Artificial intelligence isn't helping you hire the best person for the job (4/9/21)

Shedding light on fairness in AI with a new data set (4/8/21)

A $2 Billion Government Surveillance Lab Created Tech That Guesses Your Name By Simply Looking At Your Face (4/8/21)

Students of color are getting flagged to their teachers because testing software can't see them (4/8/21)

In all police stations in France, facial recognition is used (4/7/21)

Study suggests that AI model selection might introduce bias (4/7/21)

[EU] Europeans can't talk about racist AI systems. They lack the words. (4/6/21)

Twitter nukes AI-generated twits who backed Amazon and pushed anti-union rhetoric (4/6/21)

Discover the stupidity of AI emotion recognition with this little browser game (4/6/21)

Surveillance Nation (4/6/21)

Your Local Police Department Might Have Used This Facial Recognition Tool To Surveil You. Find Out Here. (4/6/21)

Can A.I. help Facebook cure its disinformation problem? (4/6/21)

[India] Modi govt now plans a 'touchless' vaccination process, with Aadhaar-based facial recognition (4/6/21)

[France] Suresnes wants to detect suspicious behavior (4/6/21)

Your 'smart home' is watching – and possibly sharing your data with the police (4/5/21)

Government audit of AI with ties to white supremacy finds no AI (4/5/21)

Can AI read your emotions? Try it for yourself (4/5/21)

Scientists create online games to show risks of AI emotion recognition (4/4/21)

Why Silicon Valley's most astute critics are all women (4/3/21)

Hitting the Books: The bias behind AI assistants' failure to understand accents (4/3/21)

Here's how enterprises say they're deploying AI responsibly (4/2/21)

[EU] Seeing stones: pandemic reveals Palantir's troubling reach in Europe (4/2/21)

A South Korean Chatbot Shows Just How Sloppy Tech Companies Can Be With User Data (4/2/21)

These Companies Track Millions Of Cars–Immigration And Border Police Have Been Grabbing Their Data (4/1/21)

Study finds that even the best speech recognition systems exhibit bias (4/1/21)

[UK] Technology Managing People – the legal implications (March 2021)

AI experts warn Facebook's anti-bias tool is 'completely insufficient' (3/31/21)

The Foundations of AI Are Riddled With Errors (3/31/21)

[UK] We've won our lawsuit over Matt Hancock's £23m NHS data deal with Palantir (3/30/21)

Utah Gave $20 Million Contract to AI Surveillance Firm That Didn't Have AI (3/30/21)

MIT study finds labelling errors in datasets used to test AI (3/29/21)

Automated translation is hopelessly sexist, but don't blame the algorithm or the training data (3/29/21)

Americans think AI has the most potential to cause harm over next decade (3/29/21)

FBI warns of the rise of 'deepfakes' in coming months and explains how to spot them easily (3/29/21)

[Australia] 'Dangerous future' in NDIS tech experimentation (3/28/21)

[Brazil] Anti-racism, technology and artificial intelligence in Brazil: a chat at Mozfest, by Mozilla (3/28/21)

AI: Ghost workers demand to be seen and heard (3/28/21)

Why Representation Matters When Building AI (3/28/21)

[Spain] Aragón will implement an algorithm for the detection of suicide risk in his clinical history (3/26/21)

If Mark Zuckerberg won't fix Facebook's algorithms problem, who will? (3/26/21)

Scotland launches AI strategy with a focus on ethics and inclusion (3/26/21)

[Germany] Allocation of vaccination appointments: Are older people disadvantaged? (3/25/21)

AI at work: Staff 'hired and fired by algorithm' (3/25/21)

[Netherlands] Roeland (22) was kicked from TikTok because the app thinks he is under 13 (3/23/21)

Amazon Delivery Drivers Forced to Sign 'Biometric Consent' Form or Lose Job (3/23/21)

After Clearview, more bad actors in A.I. facial recognition might show up (3/23/21)

Major flaws found in machine learning for COVID-19 diagnosis (3/23/21)

How synthetic data could save AI (3/20/21)

Uber under pressure over facial recognition checks for drivers (3/19/21)

[UK] Uber drivers claim they were fired after company's identification software failed to recognise their faces (3/19/21)

What happens when your massive text-generating neural net starts spitting out people's phone numbers? If you're OpenAI, you create a filter (3/18/21)

The Pandemic Could Accelerate Job Automation–and Inequality (3/18/21)

A New York Lawmaker Wants to Ban Police Use of Armed Robots (3/18/21)

AI can be unintentionally biased: Data cleaning and awareness can help prevent the problem (3/18/21)

'Anonymized' X-ray datasets can reveal patient identities (3/17/21)

California bans 'dark patterns' that trick users into giving away their personal data (3/16/21)

Europe's artificial intelligence blindspot: Race (3/16/21)

[Netherlands] Tens of thousands of people may falsely enter police face database (3/16/21)

This is the EU's chance to stop racism in artificial intelligence (3/16/21)

Diversity And Inclusion In AI (3/16/21)

Researchers blurred faces that launched a thousand algorithms (3/15/21)

Who Is Making Sure the A.I. Machines Aren't Racist? (3/15/21)

Preventing bias in AI is hard. Bug bounties could point the way forward (3/15/21)

Tenant screening software faces national reckoning (3/14/21)

Getting to trustworthy AI (3/14/21)

Google might ask questions about AI ethics, but it doesn't want answers (3/13/21)

Dutch court rulings break new ground on gig worker data rights (3/12/21)

Auditing employment algorithms for discrimination (3/12/21)

[Spain] The 'riders' law will oblige companies to inform unions about algorithms that affect working conditions (3/11/21)

[France] In the Official Journal, "smart" cameras to measure the rate of mask wearing in transport (3/11/21)

GAO's emerging tech shop developing framework to test reliability of AI algorithms (3/11/21)

How Facebook got addicted to spreading misinformation (3/11/21)

[UK] Hundreds of sewage leaks detected thanks to AI (3/11/21)

[China] Honest passengers first! Beijing subway to pilot credit-based fast entry system (3/10/21)

Biased AI can be bad for your health – here's how to promote algorithmic fairness (3/9/21)

Artificial intelligence tool helps detect unconscious racial, cultural bias in the workplace (3/8/21)

Algorithms are increasingly treating workers like robots. Canada needs policy to protect them (3/8/21)

Underpaid Workers Are Being Forced to Train Biased AI on Mechanical Turk (3/8/21)

Inside the fall of Watson Health: How IBM's audacious plan to 'change the face of health care' with AI fell apart (3/8/21)

Building AI for the Global South (3/7/21)

I asked an AI to tell me how beautiful I am (3/5/21)

'Facebook has a blind spot': why Spanish-language misinformation is flourishing (3/3/21)

Looking For An AI Ethicist? Good Luck (3/2/21)

[Russia] In Moscow, Big Brother Is Watching and Recognizing Protesters (3/1/21)

Does your company need a Chief AI Ethics Officer, an AI Ethicist, AI Ethics Council, or all three? (Mar. 2021)

Taking on the tech giants: the lawyer fighting the power of algorithmic systems (2/28/21)

5 steps to creating a responsible AI Center of Excellence (2/28/21)

Virginia Approves Limits on Police Use of Facial Recognition (2/26/21)

An AI is training counselors to deal with teens in crisis (2/26/21)

TikTok agrees legal payout over facial recognition (2/26/21)

'This is bigger than just Timnit': How Google tried to silence a critic and ignited a movement (2/26/21)

Why Google's AI ethics blunders are a PR nightmare (2/25/21)

Can Auditing Eliminate Bias fromAlgorithms? (2/23/21)

Whistleblowers: Software Bug Keeping Hundreds Of Inmates In Arizona Prisons Beyond Release Dates (2/22/21)

HUMANS ARE TRYING TO TAKE BIAS OUT OF FACIAL RECOGNITION PROGRAMS. IT'S NOT WORKING–YET. (2/22/21)

Can We Engineer Ethical AI? (2/21/21)

The Shoddy Science Behind Emotional Recognition Tech (2/19/21)

How NSF and Amazon Are Collectively Tackling Artificial Intelligence-Based Bias (2/18/21)

DeepMind researchers say AI poses a threat to people who identify as queer (2/18/21)

Studies find bias in AI models that recommend treatments and diagnose diseases (2/18/21)

Why Tech Companies Are Limiting Police Use of Facial Recognition (2/18/21)

Biden Must Halt Face Recognition Technology to Advance Racial Equity (2/17/21)

This AI reads children's emotions as they learn (2/17/21)

Can Computer Algorithms Learn to Fight Wars Ethically? (2/17/21)

Spain: Ethnic origin, gender or clothing: Renfe's controversial system to monitor its travelers (2/17/21)

Germany: Objective or biased (2/16/21)

Study shows that federated learning can lead to reduced carbon emissions (2/16/21)

Development of an AI Technology able to 'Read Emotions' across entire cities to Stop Crime Before it Happen (2/16/21)

How computers see us: Doctoral student working to curb discrimination by artificial intelligence (2/15/21)

France: The CNIL dampens the appetite of Big Brother Bercy (2/15/21)

Washington state lawmakers seek to ban government from using discriminatory AI tech (2/13/21)

Amazon's driver monitoring app is an invasive nightmare (2/13/21)

Minneapolis poised to ban facial recognition for police use (2/12/21)

Facebook's next big AI project is training its machines on users' public videos (3/12/21)

Techno-utopianism in the workplace and the threat of excessive automation (2/12/21)

Amazon uses an app called Mentor to track and discipline delivery drivers (2/12/21)

Are AI investors shorting Black lives? (2/11/21)

Russia: "Face control": Russian police go digital against protesters (2/11/21)

Why Is Facebook Rejecting These Fashion Ads? (2/11/21)

Fighting AI bias needs to be a key part of Biden's civil rights agenda (2/11/21)

COUNCIL OF EUROPE PROPOSES BAN ON CERTAIN FACIAL RECOGNITION APPLICATIONS

Despite Scanning Millions of Faces, Feds Caught Zero Imposters at Airports Last Year (2/10/21)

EU's top privacy regulator urges ban on surveillance-based ad targeting (2/10/21)

[UK] BAILII grants Oxford University unprecedented access to case data for AI analysis in historic agreement (2/10/21)

The computers rejecting your job application (2/8/21)

How Data Can Drive Inequality (2/8/21)

Diversity in AI is awful (2/8/21)

Where Do Vaccine Doses Go, and Who Gets Them? The Algorithms Decide (2/7/21)

Rethinking Gaming: The Ethical Work of Optimization in Web Search Engines (2/7/21)

'Orwellian' AI lie detector project challenged in EU court (2/5/21)

UPDATE 1-High-tech lie detector used at Europe borders face scrutiny (2/5/21)

New AI Can Detect Emotion With Radio Waves (2/4/21)

Your iPhone's Adult Content Filter Blocks Anything 'Asian' (2/4/21)

There Are Spying Eyes Everywhere–and Now They Share a Brain (2/4/21)

AI brain drain to Google and pals threatens public sector's ability to moderate machine-learning bias (2/4/21)

The Role Of Bias In Artificial Intelligence (2/4/21)

The AI industry is built on geographic and social inequality, research shows (2/4/21)

21 States Are Now Vetting Unemployment Claims With a 'Risky' Facial Recognition System (2/3/21)

U.S. technology company Clearview AI violated Canadian privacy law: report (2/3/21)

"Even if you can do it, should you?" Researchers talk combating bias in artificial intelligence (2/3/21)

[Netherlands] Can algorithms look into a child's future? In Hjørring and Silkeborg, experiments were made on vulnerable children (2/2/21)

[Singapore] Application installed on students' devices does not track personal information: MOE (2/1/21)

Confidence, uncertainty, and trust in AI affect how humans make decisions (2/1/21)

Embracing Diversity Using AI [Infographic] (2/1/21)

Inside a Hot-Button Research Paper: Dr. Emily M. Bender Talks Large Language Models and the Future of AI Ethics (2/1/21)

Here's a Way to Learn if Facial Recognition Systems Used Your Photos (1/31/21)

Forget user experience. AI must focus on 'citizen experience' (1/31/21)

Why some companies are thinking twice about using artificial intelligence (1/31/21)

The Coup We Are Not Talking About (1/29/21)

New Spotify Patent Involves Monitoring Users' Speech to Recommend Music (1/28/21)

Spotify Secures Horrifying Patent to Monitor Users' Speech (1/28/21)

Police Say They Can Use Facial Recognition, Despite Bans (1/28/21)

[Greece] Flush with EU funds, Greek police to introduce live face recognition before the summer (1/28/21)

University will stop using controversial remote-testing software following student outcry (1/28/21)

Bumble's algorithm will now report you for body-shaming (1/28/21)

It's time to use all of Twitter's archives to teach AI about bias (1/27/21)

Independent auditors are struggling to hold AI companies accountable (1/26/21)

SCIENTISTS TRYING TO BUILD MEDICAL AI THAT'S LESS RACIST THAN DOCTORS (1/26/21)

NJ Transit will test AI-powered face-mask detection (1/25/21)

New Algorithms Could Reduce Racial Disparities in Health Care (1/25/21)

Governance: Companies mature in their use of AI know that it needs guardrails (1/25/21)

[France] Suspended feminist accounts: for once, Twitter admits a real "error" (1/25/21)

[Netherlands] The Dutch Government's Benefits Scandal Is Rooted in Stigma Against Welfare Recipients (1/23/21)

AI could make healthcare fairer–by helping us believe what patients say (1/22/21)

Center for Applied Data Ethics suggests treating AI like a bureaucracy (1/22/21)

[France] In Bercy, a cell of computer scientists to help the State regulate GAFA (1/22/21)

[India] Why Lucknow Police wanting to use AI to 'read' distressed expressions of women won't work (1/21/21)

Democrats urge tech giants to change algorithms that facilitate spread of extremist content (1/21/21)

'For Some Reason I'm Covered in Blood': GPT-3 Contains Disturbing Bias Against Muslims (1/21/21)

India's biometric ID system is eroding the rights of pregnant women (1/21/21)

This App Claims It Can Detect 'Trustworthiness.' It Can't (1/19/21)

FDA publishes Action Plan to regulate AI and ML based products (1/18/21)

The FTC Forced a Misbehaving A.I. Company to Delete Its Algorithm (1/18/21)

Social-Media Algorithms Rule How We See the World. Good Luck Trying to Stop Them. (1/17/21)

[France] Twitter and Facebook send "dykes" and "queers" to the closet (1/16/21)

Chatbot Gone Awry Starts Conversations About AI Ethics in South Korea (1/16/21)

Using AI to Make Hiring Decisions? Prepare for EEOC Scrutiny (1/15/21)

Facebook uses AI to predict if COVID-19 patients will need more care (1/15/21)

An Algorithm Is Helping a Community Detect Lead Pipes (1/14/21)

Democrats push for AI bias testing (1/14/21)

Why Ethics Matter For Social Media, Silicon Valley And Every Tech Industry Leader (1/14/21)

How explainable artificial intelligence can help humans innovate (1/13/21)

Salesforce researchers release framework to test NLP model robustness (1/13/21)

Flo gets FTC slap for sharing user data when it promised privacy (1/13/21)

[Italy] The Viminal-Guarantor of privacy clash on facial recognition in real time (1/13/21)

Huawei patent mentions use of Uighur-spotting tech (1/12/21)

The Robot Made Me Do It: Human–Robot Interaction and Risk-Taking Behavior (Research Summary) (1/12/21)

The Ethics of Emotion in AI Systems (Research Summary) (1/12/21)

Job Screening Service Halts Facial Analysis of Applicants (1/12/21)

"I Don't Want Someone to Watch Me While I'm Working": Gendered Views of Facial Recognition Technology in Workplace Surveillance (Research Summary) (1/12/21)

Facial Recognition Technology Isn't Good Just Because It's Used to Arrest Neo-Nazis (1/12/21)

Facial Recognition Technology Has A Bias Problem (1/11/21)

Listening to Black Women: The Innovation Tech Can't Figure Out (1/11/21)

[Netherlands] SyRI coalition to Senate: 'Super SyRI' blueprint for more benefits affairs (1/11/21)

Outlandish Stanford facial recognition study claims there are links between facial features and political orientation (1/11/21)

[S. Korea] CEO says controversial AI chatbot 'Luda' will socialize in time (1/11/21)

Algorithms and the coronavirus pandemic (1/9/21)

New York City Proposes Regulating Algorithms Used in Hiring (1/8/21)

ExamSoft's proctoring software has a face-detection problem (1/5/21)

[Italy] Court Rules Deliveroo Used 'Discriminatory' Algorithm (1/5/21)

Medicine's machine learning problem (1/4/21)

Singapore Police May Use Contact Tracing Data for Investigations (1/4/21)

2020 in Review: 10 AI Failures (1/1/21)

2020

Peer-reviewed

Racial Bias in Pulse Oximetry Measurement (12/17/20)

[Switzerland] Smart Criminal Justice (12/10/20)

Invited Talk: You Can't Escape Hyperparameters and Latent Variables: Machine Learning as a Software Engineering Enterprise (12/7/20)

FAIROD: Fairness-aware Outlier Detection (5/20/21)

Civil Unrest on Twitter (CUT): A Dataset of Tweets to Support Research on Civil Unrest (Nov 2020)

Protecting consumers from collusive prices due to AI (11/27/20)

The ethical questions that haunt facial-recognition research (11/18/20)

How to Measure Gender Bias in Machine Translation: Optimal Translators, Multiple Reference Points (11/12/20)

We need to talk about deception in social robotics! (11/11/20)

Redistribution and Rekognition: A Feminist Critique of Algorithmic Fairness (11/7/20)

"What We Can't Measure, We Can't Understand": Challenges to Demographic Data Procurement in the Pursuit of Fairness (10/30/20)

Recipes for Safety in Open-domain Chatbots (10/22/20)

The "black box" at work (10/19/20)

Exchanging Lessons Between Algorithmic Fairness and Domain Generalization (10/14/20)

The Ethics of Emotion in AI Systems (10/8/20)

Brave: what it means to be an AI Ethicist (10/6/20)

CrowS-Pairs: A Challenge Dataset for Measuring Social Biases in Masked Language Models (9/30/21)

REAL TOXICITY PROMPTS: Evaluating Neural Toxic Degeneration in Language Models (9/25/20)

Counterfactual Explanation and Causal Inference in Service of Robustness in Robot Control (9/18/20)

Compounding Injustice: The Cascading Effect of Algorithmic Bias in Risk Assessments (8/11/20)

Discovering and Categorising Language Biases in Reddit∗ (8/6/20)

What are you optimizing for? Aligning Recommender Systems with Human Values (2020)

A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government (Aug 2020)

Artificial Intelligence and Inequality in the Middle East: The Political Economy of Inclusion (7/20/20)

Emerging challenges in AI and the need for AI ethics education (7/15/20)

Green Algorithms: Quantifying the carbon footprint of computation (7/15/20)

Machine Learning Explainability for External Stakeholders (7/10/20)

Beyond Accuracy: Behavioral Testing of NLP Models with CheckList (July 2020)

Social Bias Frames: Reasoning about Social and Power Implications of Language (July 2020)

AI and the Global South: Designing for Other Worlds (July 2020)

Don't ask if artificial intelligence is good or fair, ask how it shifts power (7/7/20)

Speech Recognition Tech Is Yet Another Example of Bias (7/5/20)

Give Me Convenience and Give Her Death: Who Should Decide What Uses of NLP are Appropriate, and on What Basis? (7/5/20)

The Ethics of Algorithms: Key Problems and Solutions (July 2020)

Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics (6/24/20)

Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for shifting Organizational Practices (6/22/20)

Algorithmic Bias: On the Implicit Biases of Social Technology (6/20/21)

Hidden in Plain Sight – Reconsidering the Use of Race Correction in Clinical Algorithms (6/17/20)

The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation (6/17/20)

Mind the app -- Considerations on the ethical risks of COVID-19 apps (6/13/20)

In AI We Trust: Ethics, Artificial Intelligence, and Reliability (6/10/20)

Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis (6/9/20)

Understanding the Essence of Artificial Intelligence: Towards Ecological Safety of AI in Human Society (Jun 2020)

Fairness, Equality, and Power in Algorithmic Decision-Making (preprint)

Language (Technology) is Power: A Critical Survey of "Bias" in NLP (5/28/20)

IEEE 7010: A New Standard for Assessing the Well-being Implications of Artificial Intelligence (5/7/20)

Double-Hard Debias: Tailoring Word Embeddings for Gender Bias Mitigation (5/3/20)

Social Biases in NLP Models as Barriers for Persons with Disabilities (5/2/20)

Multi-Dimensional Gender Bias Classification (5/1/20)

Politeness Transfer: A Tag and Generate Approach (5/1/20)

Ethics and governance for digital disease surveillance (May 2020)

Mind Your Inflections! Improving NLP for Non-Standard English with Base-Inflection Encoding (4/30/20)

The Politics of Adversarial Machine Learning (4/26/20)

Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI (4/25/20)

Learning to Diversify from Human Judgments: Research Directions and Open Challenges (4/25/20)

A Human in the Loop is Not Enough: The Need for Human-Subject Experiments in Facial Recognition (4/25/20)

Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning (4/25/20)

StereoSet measures racism, sexism, and other forms of bias in AI language models (4/22/20)

Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims (4/15/20)

Research Collection: Research Supporting Responsible AI (4/13/20)

Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem (4/9/20)

Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal (4/7/20)

Measuring the predictability of life outcomes with a scientific mass collaboration (3/30/20)

AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media (3/29/20)

Racial disparities in automated speech recognition (3/23/20)

What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis (3/11/20)

Recommender Systems and their Ethical Challenges (2/27/20)

No computation without representation: Avoiding data and algorithm biases through diversity (2/26/20)

Informed Consent and Medical Artificial Intelligence: What to Tell the Patient? (2/26/20)

Learning the difference that makes a difference with counterfactually-augment data (2/14/20)

Diversity and Inclusion Metrics in Subset Selection (2/9/20)

Hazard Contribution Modes of Machine Learning Components (2/7/20)

The Ethics of AI Ethics: An Evaluation of Guidelines (2/1/20)

Deontological Ethics By Monotonicity Shape Constraints (1/31/20)

From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy (1/29/20)

The Explanation Game: A Formal Framework for Interpretable Machine Learning (1/17/20)

Robot Rights? Let's Talk about Human Welfare Instead (1/14/20)

Algorithmic Fairness from a Non-ideal Perspective (1/9/20)

Toward Situated Interventions for Algorithmic Equity: Lessons from the Field (Jan 2020)

Reducing sentiment polarity for demographic attributes in word embeddings using adversarial learning (Jan 2020)

"The Human Body is a Black Box": Supporting Clinical Decision-Making with Deep Learning (Jan 2020)

Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure (Jan 2020)

Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing (Jan 2020)

Fairness and utilization in allocating resources with uncertain demand (Jan 2020)

Government, NGO, and expert publications

Crucial Yet Overlooked: Why We Must Reconcile Legal and Technical Approaches to Algorithmic Bias (12/17/20)

Landmark artificial intelligence legislation advances toward becoming law (12/16/20)

How Many Jobs Will AI Destroy? As Many As We Tell It To. (12/16/20)

Creating Trustworthy AI (12/15/20)

WEF: Responsible Limits on Facial Recognition Use Case: Flow Management (12/14/20)

In AI ethics, "bad" isn't good enough (12/14/20)

The ethics of artificial intelligence (12/10/20)

Why people may not trust your AI, and how to fix it (12/8/20)

What the AI Community Can Learn From Sneezing Ferrets and a Mutant Virus Debate (12/8/20)

The tech industry needs regulation for its systemically important companies (12/8/20)

Hong Qu: Shining a Headlight on AI Blindspots (12/3/20)

Report – Algorithm-driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination? (12/3/20)

When Algorithmic Fairness Fixes Fail: The Case for Keeping Humans in the Loop (12/2/20)

How News Organizations Use Algorithms to Decide What to Show You (12/1/20)

Can AI Fairly Decide Who Gets an Organ Transplant? (12/1/20)

How AI bots and voice assistants reinforce gender bias (11/23/20)

Corsight AI launches real-time facial recognition that identifies masked faces (11/7/20)

Google AI Blog: Mitigating Unfair Bias in ML Models with the MinDiff Framework (11/16/20)

Retooling AI: Algorithm Bias and the Struggle to Do No Harm (11/10/20)

A Legal Approach to "Affirmative Algorithms" (11/9/20)

AUTOMATING SOCIETY REPORT 2020 (EU) (Nov 2020)

Stanford HAI Policy Brief: Domain Shift and Emerging Questions in Facial Recognition Technology (Nov 2020)

Stanford HAI Policy Brief: Preparing for the Age of Deepfakes and Disinformation (Nov 2020)

Root Out Bias at Every Stage of Your AI-Development Process
(10/30/20)

Spanish police plan to extend use of its lie-detector while efficacy is
unclear (10/27/20)

HBR: When Do We Trust AI's Recommendations More Than People's?
(10/16/20)

How to use AI hiring tools to reduce bias in recruiting (10/13/20)

Men drive trucks, women raise children - discriminatory gender
targeting through Facebook (10/11/20)

The Short Anthropological Guide to the Study of Ethical AI (10/10/20)

Net worthy: How to get budget and buy-in for tech ethics (and other
forms of responsible business) (Oct 2020)

Stanford HAI Policy Brief: Toward Fairness in Health Care Training Data
(Oct 2020)

HBR: Automated Hardship: How the Tech-Driven Overhaul of the UK's
Social Security System Worsens Poverty (9/29/20)

NEC and The Face Recognition Company each upgrade biometric
systems accuracy for masked faces (9/24/21)

POVERTY LAWGORITHMS: A Poverty Lawyer's Guide to Fighting
Automated Decision-Making Harms on Low-Income Communities
(9/15/20)

AI ethics groups are repeating one of society's classic mistakes (9/14/20)

AI's Promise and Peril for the U.S. Government [Stanford HAI] (Sept 2020)

Framework for Promoting Workforce Well-being in the AI-Integrated Workplace (8/27/20)

GAO Facial Recognition Technology: Privacy and Accuracy Issues Related to Commercial Uses (8/11/20)

Risks of Discrimination through the Use of Algorithms [German Federal Anti-Discrimination Agency] (8/7/20)

NIST: Four Principles of Explainable Artificial Intelligence (Aug 2020)

BMM Privacy Bot: Tool to anonymize BLM protesters

Report: Bridging AI's trust gaps: Aligning policymakers and companies (The Future Society) (7/22/20)

Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (7/17/20)

Facebook's Civil Rights Audit – Final Report (7/8/20)

AI Watch: Artificial Intelligence in public services (July 2020)

AUTOMATED SUSPICION: THE EU'S NEW TRAVEL SURVEILLANCE INITIATIVES (July 2020)

The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law: Ad Hoc Committee on Artificial Intelligence (6/24/20)

A call for a critical look at the metrics for success in the evaluation of AI (6/17/20)

ARTIFICIAL INTELLIGENCE ETHICS FRAMEWORK FOR THE INTELLIGENCE COMMUNITY (June 2020)

FACIAL RECOGNITION TECHNOLOGIES: May 29, 2020 A PRIMER (5/29/20)

When AI Writes Your Email (5/6/20)

UC Berkeley Decision Points in AI Governance THREE CASE STUDIES EXPLORE EFFORTS TO OPERATIONALIZE AI PRINCIPLES (May 2020)

Artificial Intelligence for Social Good in Latin America and the Caribbean: The Regional Landscape and 12 Country Snapshots (May 2020)

Is AI trustworthy enough to help us fight COVID-19? (5/1/20)

A Scalable Approach to Reducing Gender Bias in Google Translate (4/22/20)

How to create a trustworthy COVID-19 tracking technology (4/20/20)

Aligning AI to Human Values means Picking the Right Metrics (4/15/20)

Measuring the predictability of life outcomes with a scientific mass collaboration (4/14/20)

Statement Regarding the Ethical Implementation of Artificial
Intelligence Systems (AIS) for Addressing the COVID-19 Pandemic
(April 2020)

Interpretability [Cloudera Fast Forward Report] (Apr 2020)

FTC: Using Artificial Intelligence and Algorithms (4/8/20)

From Principles to Practice An interdisciplinary framework to
operationalise AI ethics (4/2/20)

The DoD AI Ethical Principles: Shifting From Principles to Practice
(4/1/20)

AI is transforming society. Here's what we can do to make sure it
prioritizes human needs. (3/27/20)

To Prevent Algorithmic Bias, Legal and Technical Definitions around
Algorithmic Fairness Must Align (3/23/20)

Digital Commerce, AI, and Constraining Consumer Choice (3/19/20)

Mapping coronavirus, responsibly (2/25/20)

Algorithm Groups People More Fairly to Reduce AI Bias (2/12/20)

Protecting privacy in an AI-driven world (2/10/20)

ML-fairness-gym: A Tool for Exploring Long-Term Impacts of Machine
Learning Systems (2/5/20)

The measure and mismeasure of fairness: a critical review of fair
machine learning (2/3/20)

Bringing Facial Recognition Systems To Light (Feb 2020)

Artificial Intelligence in the Asia-Pacific Region: Examining policies and strategies to maximise AI readiness and adoption (Feb 2020)

IBM: Precision Regulation for Artificial Intelligence (1/21/20)

Black-Boxed Politics:Opacity is a Choice in AI Systems (1/17/20)

What do we teach when we teach tech & AI ethics? (1/17/20)

How businesses can create an ethical culture in the age of tech (1/7/20)

Artificial Intelligence Principles For Vulnerable Populations in Humanitarian Contexts (Jan 2020)

Excavating AI: The Politics of Images in Machine Learning Training Sets (Jan 2020)

Learning from the past to create Responsible AI (evergreen)

Popular Press

Unregulated facial recognition must stop before more Black men are wrongfully arrested (12/31/20)

In 2020, COVID-19 derailed the privacy debate in the U.S (12/29/20)

Why 2020 was a pivotal, contradictory year for facial recognition (12/29/20)

Researchers find race, gender, and style biases in art-generating AI systems (12/28/20)

AI jobs in 2021: here are some key trends (12/28/20)

AI research survey finds machine learning needs a culture change (12/26/20)

He spent 10 days in jail after facial recognition software led to the arrest of the wrong man, lawsuit says (12/28/20)

Algorithms are deciding who gets the first vaccines. Should we trust them? (12/23/20)

Dozens sue Amazon's Ring after camera hack leads to threats and racial slurs (12/23/20)

Pulse Oximeter Devices Have Higher Error Rate in Black Patients (12/22/20)

This is the Stanford vaccine algorithm that left out frontline doctors (12/21/20)

How to create space for ethics in AI (12/19/20)

Stanford Apologizes After Vaccine Allocation Leaves Out Nearly All Medical Residents (12/18/20)

Pôle emploi makes it much too easy for anyone to obtain French CVs (12/17/20)

From whistleblower laws to unions: How Google's AI ethics meltdown could shape policy (12/16/20)

[Canada] Use of surveillance software to crack down on exam cheating has unintended consequences (12/16/20)

Researchers describe profound complexity of biometrics ethics puzzle (12/15/20)

Surveillance companies are using mobile ads to obtain scarily accurate location data (12/11/20)

AI needs to face up to its invisible-worker problem (12/11/20)

Sci-fi surveillance: Europe's secretive push into biometric technology (12/10/20)

How our data encodes systematic racism (12/10/20)

Study finds crime-predicting judicial tool exhibits gender bias (12/10/20)

Study finds diversity in data science teams is key in reducing algorithmic bias (12/9/20)

IBM announces new AI language, explainability, and automation services (12/9/20)

Uni revealed it killed off its PhD-applicant screening AI – just as its inventors gave a lecture about the tech (12/8/20)

Huawei tested AI software that could recognize Uighur minorities and alert police, report says (12/8/20)

You can buy a robot to keep your lonely grandparents company. Should you? (12/8/20)

Artificial Intelligence Is An Amazing Disruptor, But Has A Major Impact On Unskilled Workers (12/8/20)

Surprise, surprise: AI cameras sold to schools in New York struggle with people of color and are full of false positives (12/7/20)

How banks use AI to catch criminals and detect bias (12/7/20)

We can reduce gender bias in natural-language AI, but it will take a lot more work (12/6/20)

The coming war on the hidden algorithms that trap people in poverty (12/4/20)

Health algorithms discriminate against Black patients, also in Switzerland (12/4/20)

[France] The Interior strengthens the possibilities of political identification (12/4/20)

Study shows how AI exacerbates recruitment bias against women (12/2/20)

Microsoft apologises for feature criticised as workplace surveillance (12/2/20)

How The Department Of Defense Approaches Ethical AI (11/29/20)

Ethical AI isn't the same as trustworthy AI, and that matters (11/28/20)

[UK] The Covid data spies paid to know ALL your secrets: Town halls harvest millions of highly personal details including if you're being

unfaithful or having unsafe sex (11/27/20)

Uganda is using Huawei's facial recognition tech to crack down on dissent after anti-government protests (11/27/20)

Who will your algorithm harm next? Why businesses need to start thinking about evil AI now (11/26/20)

Can an Algorithm Prevent Suicide? (11/23/20)

[France] French tax authority pushes for automated controls despite mixed results (11/23/20)

The UK Government Isn't Being Transparent About Its Palantir Contracts (11/23/20)

China's Surveillance State Sucks Up Data. U.S. Tech Is Key to Sorting It. (11/22/20)

Watch: Facial recognition at Dubai Metro stations to identify wanted criminals (11/22/20)

Can We Make Our Robots Less Biased Than We Are? (11/22/20)

Hundreds of Facebook moderators complain: AI content moderation isn't working and we're paying for it (11/21/20)

New York City wants to restrict artificial intelligence in hiring (11/20/20)

These Algorithms Could Bring an End to the World's Deadliest Killer (11/20/20)

Facebook's latest efforts to combat hate speech aren't enough, ADL says (11/20/20)

When AI Sees a Man, It Thinks 'Official.' A Woman? 'Smile' (11/19/20)

Facebook's A.I. is getting better at finding malicious content–but it won't solve the company's problems (11/19/20)

Banks turn to AI as regulators press for Libor exit (11/19/20)

Zest AI Joins Forces With Freddie Mac to Help Make Homeownership Possible for More Americans (11/19/20)

When AI Systems Fail: Introducing the AI Incident Database (11/18/20)

The way we train AI is fundamentally flawed (11/18/20)

How regulators can get facial recognition technology right (11/17/20)

Health Care AI Systems Are Biased (11/17/20)

Online exams raise concerns of racial bias in facial recognition (11/17/20)

[Australia] Robodebt class action: Coalition agrees to pay $1.2bn to settle lawsuit (11/16/20)

Activists urge EU to ban live facial recognition in public spaces (11/12/20)

'Coded Bias' Review: When the Bots Are Racist (11/11/20)

Why AI can't move forward without diversity, equity, and inclusion (11/12/20)

Equitable tech: AI-enabled platform to reduce bias in datasets released (11/11/20)

AI research finds a 'compute divide' concentrates power and accelerates inequality in the era of deep learning (11/11/20)

"Data Trusts" Could Be the Key to Better AI (11/10/20)

The US Government Will Pay Doctors to Use These AI Algorithms (11/10/20)

Face for sale: Leaks and lawsuits blight Russia facial recognition (11/9/20)

How artificial intelligence may be making you buy things (11/9/20)

Pope urges Catholics to pray that AI does not widen inequality (11/7/20)

How IBM Is Working Toward a Fairer AI (11/5/20)

Proctorio used DMCA to take down a student's critical tweets (11/5/20)

Artificial intelligence is making the beauty industry work for everyone (11/4/20)

When AI Says, 'You're Fired' (11/4/20)

Portland, Maine has voted to ban facial recognition (11/4/20)

Police Will Pilot a Program to Live-Stream Amazon Ring Cameras
(11/3/20)

How the Police Use AI to Track and Identify You (11/3/20)

Schools Adopt Face Recognition in the Name of Fighting Covid
(11/3/20)

Artificial Intelligence in healthcare is racist (11/2/20)

When Algorithms Infer Pregnancy or Other Sensitive Information
About People (11/2/20)

Fairness Definitions Explained (Research Summary) (11/2/20)

Big tech's 'blackbox' algorithms face regulatory oversight under EU
plan (10/30/20)

The AI Company Helping the Pentagon Assess Disinfo Campaigns
(10/28/20)

Nearly half of councils in Great Britain use algorithms to help make
claims decisions (10/28/20)

Uber Is Getting Sued Over Its Allegedly Racist Ratings System
(10/27/20)

Why Getting Paid for Your Data Is a Bad Deal (10/26/20)

Uber Drivers Launch Legal Action Over 'Robo-Firing' By Algorithm
(10/26/20)

How an Algorithm Blocked Kidney Transplants to Black Patient (10/26/20)

AI bias: blame the workman, not his tools (10/26/20)

Surveillance Startup Used Own Cameras to Harass Coworkers (10/26/20)

How the Racism Baked Into Technology Hurts Teens (10/24/20)

How to make a chatbot that isn't racist or sexist (10/23/20)

AI Weekly: Constructive ways to take power back from Big Tech (10/23/20)

The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research (10/22/20)

AI researchers urge tech to go beyond scale to address systemic social issues (10/22/20)

Putting Responsible AI Into Practice (10/22/20)

Algorithms Are Making Economic Inequality Worse (10/22/20)

The true dangers of AI are closer than we think (10/21/20)

Activists Turn Facial Recognition Tools Against the Police (10/21/20)

GOOGLE AI TECH WILL BE USED FOR VIRTUAL BORDER WALL, CBP CONTRACT SHOWS (10/21/20)

Facial recognition datasets are being widely used despite being taken down due to ethical concerns. Here's how. (10/21/20)

White House Nears New Rules on Artificial Intelligence (10/21/20)

Salesforce's Simulation Cards spell out uses, risks, and bias to make AI models more transparent (10/20/20)

Photoshop's AI neural filters can tweak age and expression with a few clicks (10/20/20)

AI Fairness Isn't Just an Ethical Issue (10/20/20)

Singapore releases AI ethics, governance reference guide (10/16/20)

UK passport photo checker shows bias against dark-skinned women (10/16/20)

The real promise of synthetic data (10/16/20)

BMW develops AI ethics code for its cars (10/16/20)

A practical guide to building ethical AI (10/15/20)

AI for good: A better, more inclusive future of work (10/15/20)

Six ways machine learning threatens social justice (10/15/20)

'Machines set loose to slaughter': the dangerous rise of military AI (10/15/20)

Six ways machine learning threatens social justice (10/15/20)

AI Reads Human Emotions. Should it? (10/14/20)

Politicians have made an algorithm to fix the housing crisis. It's bad (10/14/20)

Which Company Uses the Most of Your Data? (10/14/20)

From a small town in North Carolina to big-city hospitals, how software infuses racism into U.S. health care (10/13/20)

When AI hurts people, who is held responsible? (10/12/20)

Shrinking the 'data desert': Inside efforts to make AI systems more inclusive of people with disabilities (10/12/20)

AI Can Help Diagnose Some Illnesses–If Your Country Is Rich (10/11/20)

Toddlers Are Being Scooped Up in Buenos Aires' Live Facial Recognition Dragnet (10/9/20)

UK passport photo checker shows bias against dark-skinned women (10/7/20)

The biggest barrier to humane, ethical AI: Capitalism itself (10/5/20)

Facial Recognition, Racial Recognition and the Clear and Present Issues with AI Bias (10/5/20)

Twitter vows to fix bias image cropping issue (10/2/20)

AI and ethics: One-third of executives are not aware of potential AI bias (10/2/20)

6 big ethical questions about the future of AI (Oct 2020)

A Facial Action Coding system to explain what women are thinking? You could just ask (9/30/20)

How not to build a biased algorithm (9/30/20)

Alexa, do I have COVID-19? (9/30/20)

The True Impact of Sexist and Racist Bias in Algorithms (9/29/20)

ExamSoft's remote bar exam sparks privacy and facial recognition concerns (9/29/20)

Netherlands: End dangerous mass surveillance policing experiments (9/29/20)

Our goal shouldn't be to build merely 'trustworthy' AI (9/28/20)

Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI (9/28/20)

AI Democratization in the Era of GPT-3 (9/25/20)

'I'm extremely controversial': the psychologist rethinking human emotion (9/25/20)

Students Are Pushing Back Against Proctoring Surveillance Apps (9/25/20)

How close is AI to decoding our emotions? (9/24/20)

Amazon is embracing surveillance-as-a-service (9/24/20)

GPT-3's bigotry is exactly why devs shouldn't use the internet to train AI (9/24/20)

A controversial photo editing app slammed for AI-enabled 'blackface' feature (9/24/20)

Locked Out: Can Algorithms Violate Fair Housing Laws? (9/24/20)

Regulating biometrics: Taking stock of a rapidly changing landscape (9/23/20)

The geographic bias in medical AI tools (9/23/20)

Can ethics classes actually influence students' moral behavior? (9/23/20)

To Clean Up Comments, Let AI Tell Users Their Words Are Trash (9/22/20)

To Make Fairer AI, Physicists Peer Inside Its Black Box (9/22/20)

Another reminder that bias, testing, diversity is needed in machine learning: Twitter's image-crop AI may favor white men, women's chests (9/21/20)

Twitter and Zoom's algorithmic bias issues (9/21/20)

Twitter apologises for 'racist' image-cropping algorithm (9/21/20)

Diversity in AI: The Invisible Men and Women (9/21/20)

Twitter is looking into why its photo preview appears to favor white faces over Black faces (9/20/20)

Algorithms may never really figure us out – thank goodness (9/20/20)

Is an Algorithm Less Racist Than a Loan Officer? (9/18/20)

Facebook Testing Implications of Privacy-Invading Tech By Invading People's Privacy (9/18/20)

Cutting-edge language models can produce convincing misinformation if we don't stop them (9/18/20)

Fake Data Could Help Solve Machine Learning's Bias Problem–if We Let It (9/17/20)

Unethical AI unfairly impacts protected classes - and everybody else as well (9/16/20)

Ethical Tech Starts With Addressing Ethical Debt (9/16/20)

Where is the accountability for AI ethics gatekeepers? (9/15/20)

AI researchers devise failure detection method for safety-critical machine learning (9/15/20)

AI ethics groups are repeating one of society's classic mistakes (9/14/20)

How Algorithms Can Fight Bias Instead of Entrench It (9/14/20)

Why the AI industry urgently needs more uncomfortable conversations about BAME representation (9/14/20)

New standards for AI clinical trials will help spot snake oil and hype (9/11/20)

Could you live here? Decades-old redlining still shapes San Diego's neighborhoods today, advocates say (9/11/20)

Portland passes broadest facial recognition ban in the US (9/9/20)

Portland Passes Groundbreaking Ban on Facial Recognition in Stores, Banks, Restaurants and More (9/9/20)

Mutant Algorithms Are Coming for Your Education (9/8/20)

When Algorithms Give Real Students Imaginary Grades (9/8/20)

Here are the biased algorithms the UK government uses to make high-level decisions (9/7/20)

From viral conspiracies to exam fiascos, algorithms come with serious side effects (9/6/20)

Facebook AI open-sources Opacus, a new high-speed library for training PyTorch models with differential privacy (DP) (9/6/20)

Pasco's sheriff created a futuristic program to stop crime before it happens: It monitors and harasses families across the county. (9/3/20)

What Does Building a Fair AI Really Entail? (9/3/20)

Racial biases infect artificial intelligence (9/2/20)

Biometrics Have Crept Into Humanitarian Aid, But the Systems May Disadvantage Women Who Need Help Most (9/2/20)

Microsoft launches a deepfake detector ahead of US election (9/2/20)
Bodies into Bits (8/31/20)

Google Offers to Help Others With the Tricky Ethics of AI (8/28/20)

Rooting out racism in AI systems -- there's no time to lose (8/28/20)

Facebook's discriminatory ad targeting illustrates the dangers of biased algorithms (8/28/20)

PopID's face-based payments pose privacy and security risks (8/27/20)

The utopian promise and dystopian potential of real-time detection of police, fire, and medical emergencies (8/27/20)

Participation-washing could be the next dangerous fad in machine learning (8/25/20)

Does Facebook Still Sell DiscriminatoryAds? (8/25/20)

LinkedIn open-sources toolkit to measure AI model fairness (8/25/20)

Councils scrapping use of algorithms in benefit and welfare decisions (8/24/20)

The term 'ethical AI' is finally starting to mean something (8/23/20)

Algorithms can drive inequality. Just look at Britain's school exam chaos (8/23/20)

Best Practices for Indigenous Keywording for Stock Images (8/20/20)

Why 'Ditch the algorithm' is the future of political protest (8/19/20)

INSIGHT: Facial Recognition Is a Threat to People of Color (8/18/20)

Too many AI researchers think real-world problems are not relevant (8/18/20)

Does technology increase the problem of racism and discrimination? (8/17/20)

An A.I. Training Tool Has Been Passing Its Bias to Algorithms for Almost Two Decades (8/17/20)

Governments have collected large amounts of data to fight the coronavirus. That's raising privacy concerns (8/17/20)

Inbuilt biases and the problem of algorithms (8/17/20)

Facebook algorithm found to 'actively promote' Holocaust denial (8/16/20)

Human Rights Commission warns government over 'dangerous' use of AI (8/15/20)

U.K. Regulator Issues Data Guidance for Companies Working With AI (8/14/20)

Horrific AI surveillance experiment uses convicted felons as human guinea pigs (8/14/20)

Problematic study on Indiana parolees seeks to predict recidivism with AI (8/14/20)

NYPD Used Facial Recognition Technology In Siege Of Black Lives Matter Activist's Apartment (8/14/20)

Facebook's AI for detecting hate speech is facing its biggest challenge yet (8/14/20)

The Quiet Growth of Race-Detection Software Sparks Concerns Over Bias (8/14/20)

Governments should close the AI trust gap with businesses (8/13/20)

The pandemic is speeding up automation, putting jobs in question (8/11/20)

Ethics Is More Important than Technology (8/10/20)

WOULD YOU TRUST A LAWYER BOT WITH YOUR LEGAL NEEDS? (8/10/20)

Research summary: Mass Incarceration and the Future of AI (8/9/20)

Researchers quantify bias in Reddit content sometimes used to train AI (8/7/20)

Hypotenuse AI wants to take the strain out of copywriting for e-commerce (8/7/20)

Can language models learn morality? (8/7/20)

AI Project Failure Rates Near 50%, But It Doesn't Have to Be That Way, Say Experts (8/7/20)

Here are a few ways GPT3 can go wrong (8/7/20)

AI models need to be 'interpretable' rather than just 'explainable' (8/6/20)

Researchers discover evidence of gender bias in major computer vision APIs (8/6/20)

Researchers say 'The Whiteness of AI' in pop culture erases people of color (8/6/20)

Police built an AI to predict violent crime. It was seriously flawed (8/6/20)

Explainable AI: A guide for making black box machine learning models explainable (8/6/20)

How a Popular Medical Device Encodes Racial Bias (8/5/20)

Evil AI: These are the 20 most dangerous crimes that artificial intelligence will create (8/5/20)

Can Artificial Intelligence Help Increase Diversity in IT? (8/5/20)

Meet the computer scientist and activist who got Big Tech to stand down (8/4/20)

New Zealand Has a Radical Idea for Fighting Algorithmic Bias: Transparency (8/4/20)

Cloak your photos with this AI privacy tool to fool facial recognition (8/4/20)

Home Office to scrap 'racist algorithm' for UK visa applicants (8/4/20)

The problem of underrepresented languages snowballs from data sets to NLP models (8/4/20)

On Diversity, Silicon Valley Failed to Think Different (8/3/20)

Why artificial intelligence models are often biased, according to the Google exec who heads Alphabet's internal tech incubator Jigsaw (8/3/20)

AI is struggling to adjust to 2020 (8/2/20)

The 6 unholy AI systems thou shalt not develop (7/31/20)

The problems AI has today go back centuries (7/31/20)

Researchers examine the ethical implications of AI in surgical settings (7/31/20)

AI-powered tool aims to help reduce bias and racially charged language on websites (7/30/20)

OPENAI'S LATEST BREAKTHROUGH IS ASTONISHINGLY POWERFUL, BUT STILL FIGHTING ITS FLAWS (7/30/20)

Over Half of Americans Do Not Trust Companies to Ethically Collect, Use or Sell Personal Data (7/29/20)

Google releases Model Card Toolkit to promote AI model transparency (7/29/20)

Service that uses AI to identify gender based on names looks incredibly biased (7/29/20)

Power, Harms, and Data (7/28/20)

Rite Aid deployed facial recognition systems in hundreds of U.S. stores (7/28/20)

Even Google CEO Sundar Pichai agrees that it is imperative to embed ethics into AI (7/27/20)

Researchers find evidence of bias in facial expression data sets (7/25/20)

Here's why AI didn't save us from COVID-19 (7/24/20)

Four Steps for Drafting an Ethical Data Practice Blueprint (7/24/20)

An AI hiring firm says it can predict job hopping based on your interviews (7/24/20)

An online image database will remove 600,000 pictures after an art project revealed the system's racist bias. (9/24/20)

AI says men are lazy (7/24/20)

Google Ad Portal Equated "Black Girls" with Porn (7/23/20)

Facebook ignored racial bias research, employees say (7/23/20)

AI system detects posts by foreign 'trolls' on Facebook and Twitter (7/22/20)

Blind to black people (7/21/20)

Dermatology faces a reckoning: Lack of darker skin in textbooks and journals harms care for patients of color (7/21/20)

Facebook Creates Teams to Study Racial Bias, After Previously Limiting Such Efforts (7/21/20)

Clients loved this designer's work. Turns out, he was an AI (7/20/20)

Tackling the misinformation epidemic with "In Event of Moon Disaster" (7/20/20)

Why Hundreds of Mathematicians Are Boycotting Predictive Policing (7/20/20)

WHY RACIAL BIAS STILL HAUNTS SPEECH-RECOGNITION AI (7/17/20)

Predictive policing algorithms are racist. They need to be dismantled. (7/17/20)

This grading algorithm is failing students (7/16/20)

MIT researchers find 'systematic' shortcomings in ImageNet data set (7/15/20)

Deepfake used to attack activist couple shows new disinformation frontier (7/15/20)

DARPA Pays $1 Million For An AI App That Can Predict An Enemy's Emotions (7/15/20)

An invisible hand: Patients aren't being told about the AI systems advising their care (7/15/20)

THE MICROSOFT POLICE STATE: MASS SURVEILLANCE, FACIAL RECOGNITION, AND THE AZURE CLOUD (7/14/20)

Facial recognition technology is one of the most racist weapons in the police arsenal (7/14/20)

Massachusetts Edges Toward First State Facial Recognition Ban (7/14/20)

Op-Ed: Bias against African American English speakers is a pillar of systemic racism (7/14/20)

Why are Artificial Intelligence systems biased? (7/12/20)

DeepMind researchers propose rebuilding the AI industry on a base of anticolonialism (7/11/20)

Controversial Detroit facial recognition got him arrested for a crime he didn't commit (7/10/20)

Meet the Secret Algorithm That's Keeping Students Out of College (7/10/20)

AI for self-driving cars doesn't account for crime (7/10/20)

Automatic for the Bosses (7/9/20)

Deepfakes are becoming the hot new corporate training tool (7/7/20)

RACIAL EQUITY IN DATA INTEGRATION: HOW TO EXCLUDE RACIAL BIAS IN DATA (7/6/20)

How AI can empower communities and strengthen democracy (7/4/20)

Study: Only 18% of data science students are learning about AI ethics (7/3/20)

9 emerging job roles for the future of AI (7/2/20)

MIT Takes Down Popular AI Dataset Due to Racist, Misogynistic Content (7/2/20)

Salesforce researchers claim new method mitigates AI models' gender bias (7/1/20)

MIT removes huge dataset that teaches AI systems to use racist, misogynistic slurs (7/1/20)

MIT apologizes, permanently pulls offline huge dataset that taught AI systems to use racist, misogynistic slurs (7/1/20)

Vivienne Ming on Understanding the Impact of Courage & Fear Using AI (6/30/20)

Vestager warns against predictive policing in Artificial Intelligence (6/30/20)

New mathematical idea reins in AI bias towards making unethical and costly commercial choices (6/30/20)

Machine learning systems, fair or biased, reflect our moral standards (6/30/20)

Montreal AI Ethics Institute suggests ways to counter bias in AI models (6/30/20)

Objective Algorithms Are a Myth (6/29/20)

Here Is How The United States Should Regulate Artificial Intelligence
(6/28/20)

AI Weekly: A deep learning pioneer's teachable moment on AI bias
(6/26/20)

AI gatekeepers are taking baby steps toward raising ethical standards
(6/26/20)

Why Statistics Don't Capture The Full Extent Of The Systemic Bias In
Policing (6/25/20)

If Done Right, AI Could Make Policing Fairer (6/25/20)

Lessons from the PULSE Model and Discussion (6/24/20)

Wrongfully Accused by an Algorithm (6/24/20)

If AI is going to help us in a crisis, we need a new kind of ethics
(6/24/20)

California city bans predictive policing in U.S. first (6/24/20)

What a machine learning tool that turns Obama white can (and can't)
tell us about AI bias (6/23/20)

Over 1,000 AI Experts Condemn Racist Algorithms That Claim to
Predict Crime (6/23/20)

Police Forces in Canada Are Quietly Adopting Facial Recognition Tech
(6/23/20)

The flaws that make today's AI architecture unsafe and a new approach that could fix it (6/22/20)

Big Tech Is Using the Pandemic to Push Dangerous New Forms of Surveillance (6/22/20)

US government doesn't know how it uses facial recognition in public housing (6/22/20)

The flaws that make today's AI architecture unsafe and a new approach that could fix it (6/22/20)

WE DON'T HAVE TO SETTLE FOR BIASED AI (6/18/20)

Here's an Algorithm for Defunding the Police (6/18/20)

Racial bias skews algorithms widely used to guide care from heart surgery to birth, study finds (6/17/20)

How is Face Recognition Surveillance Technology Racist? (6/16/20)

Riding Out Quarantine With a Chatbot Friend: 'I Feel Very Connected' (6/16/20)

Fresh concerns about AI bias in the age of COVID-19 (6/15/20)

The Liabilities of Artificial Intelligence Are Increasing (6/15/20)

Who Is Responsible When Autonomous Systems Fail? (6/15/20)

Mind the App–Considerations on the Ethical Risks of COVID-19 Apps (6/13/20)

Researchers find racial discrimination in 'dynamic pricing' algorithms used by Uber, Lyft, and others (6/12/20)

Researchers propose framework to measure AI's social and environmental impact (6/12/20)

US government doesn't know how it uses facial recognition in public housing (6/11/20)

The Protests Prove the Need to Regulate Surveillance Tech (6/9/20)

How to detect unwanted bias in machine learning models (6/5/20)

This startup is using AI to give workers a "productivity score" (6/4/20)

Amazon "Stands in Solidarity" Against Police Racism While Selling Racist Tech to Police (6/3/20)

Of course technology perpetuates racism. It was designed that way. (6/3/20)

Microsoft researchers say NLP bias studies must consider role of social hierarchies like racism (6/1/20)

Gender Bias In Predictive Algorithms: How Applied AI Research Can Help Us Build A More Equitable Future (5/30/20)

Access Denied: Faulty Automated Background Checks Freeze Out Renters (5/28/20)

ACLU sues facial recognition firm Clearview AI, calling it a 'nightmare scenario' for privacy (5/28/20)

How people with Down syndrome are teaching Google (5/20/20)

This robot can guess how you're feeling by the way you walk (5/18/20)

Algorithms associating appearance and criminality have a dark past (5/15/20)

AI and me: friendship chatbots are on the rise, but is there a gendered design flaw? (5/7/20)

Facebook's AI detects gender bias in text (5/6/20)

Do I sound sick to you? Researchers are building AI that would diagnose COVID-19 by listening to people talk. (4/30/20)

Aerial surveillance planes to begin flying over Baltimore Friday (4/30/20)

Algorithmic Risk Assessment and COVID-19: Why PATTERN Should Not Be Used (4/30/20)

Google Fixes Gender Bias in Google Translate (Again) (4/29/20)

Silicon Valley needs a new approach to studying ethics now more than ever (4/24/20)

The role of demographic data in addressing algorithmic bias (4/24/20)

StereoSet measures racism, sexism, and other forms of bias in AI language models (4/22/20)

AI researchers propose 'bias bounties' to put ethics principles into practice (4/17/20)

Cambridge Researchers Tackle Neural Machine Translation's Gender Bias (4/16/20)

Google's AutoML Zero lets the machines create algorithms to avoid human bias (4/14/20)

Researchers find AI is bad at predicting GPA, grit, eviction, job training, layoffs, and material hardship 3/30/20

AI BIAS: A THREAT TO WOMEN'S LIVES? (3/26/20)

Trust at the center: Building an ethical AI framework (3/26/20)

Pardon the Intrusion #13: Policing using AI (3/23/20)

There Is a Racial Divide in Speech-Recognition Systems, Researchers Say (3/23/20)

Should I Worry About... the philosophy behind AI? (3/15/20)

Lie detectors have always been suspect. AI has made the problem worse. (3/13/20)

Could we forgive a machine? Study explores forgiveness in the context of robotics and AI (3/12/20)

ATIH: Natalie Garrett on Ethical Debt and Ethics Education in Tech (3/12/20)

Police Used Facial Recognition to Arrest Over 1,100 People in India Last Month (3/11/20)

A Crisis of Ethics in Technology Innovation (3/10/20)

International Women's Day: how can algorithms be sexist? (3/10/20)

Algorithms that run our lives are racist and sexist. Meet the women trying to fix them (3/10/20)

Algorithms Learn Our Workplace Biases. Can They Help Us Unlearn Them? (3/10/20)

Maryland's face recognition system is one of the most invasive in the nation (3/9/20)

Why the AI we rely on can't get privacy right (yet) (3/7/20)

The case for an AI that puts nature and ethics first, not humans (3/7/20)

Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich (3/5/20)

Clearview AI: We Are 'Working to Acquire All U.S. Mugshots' From Past 15 Years (3/5/20)

This Small Company Is Turning Utah Into a Surveillance Panopticon (3/4/20)

Humans Are The Cause Of Bias In AI, But We're Also The Solution (3/3/20)

This Is The Year Of AI Regulations (3/1/20)

TOP 8 FUNNIEST AND SHOCKING AI FAILURES OF ALL TIME (3/1/20)

How To Combat The Dark Side Of AI (2/28/20)

HOW FACIAL RECOGNITION TECHNOLOGY COULD CHANGE COLLEGE CAMPUSES COMPLETELY (2/28/20)

Why is TikTok creating filter bubbles based on your race? (2/28/20)

Pope Francis joins IBM and Microsoft in call for AI regulation (2/28/20)

FIRST EVER DECISION OF A FRENCH COURT APPLYING GDPR TO FACIAL RECOGNITION (2/27/20)

Clearview's Facial Recognition App Has Been Used By The Justice Department, ICE, Macy's, Walmart, And The NBA (2/27/20)

Scientists propose new regulatory framework to make AI safer (2/26/20)

Clearview AI, The Company Whose Database Has Amassed 3 Billion Photos, Hacked (2/26/20)

This Filter Makes Your Photos Invisible to Facial Recognition (2/26/20)

Suckers List: How Allstate's Secret Auto Insurance Algorithm Squeezes Big Spenders (2/25/20)

AI Deception: When Your Artificial Intelligence Learns to Lie (2/24/20)

Pentagon Adopts New Ethical Principles for Using AI in War (2/24/20)

Study finds quarter of climate change tweets from bots (2/22/20)

Privacy commissioners launch investigation into facial recognition technology tested by Toronto police, other GTA forces (2/21/20)

Google launches TensorFlow library for optimizing fairness constraints
(2/21/20)

Google AI will no longer use gender labels like 'woman' or 'man' on
images of people to avoid bias (2/20/20)

Artificial Intelligence (AI) And The Law: Helping Lawyers While
Avoiding Biased Algorithms (2/18/20)

Facial Expressions Don't Tell The Whole Story Of Emotion (2/17/20)

Mastercard working with transport partners on payments via gait or
face biometrics (2/17/20)

Mastercard is pioneering new payment technology that identifies
commuters by the way they walk (2/17/20)

EU's new AI rules will focus on ethics and transparency (2/17/20)

AI systems claiming to 'read' emotions pose discrimination risks
(2/16/20)

The Future of Artificial Emotional Awareness (2/13/20)

Fake news? A.I. algorithm reveals political bias in the stories you read
(2/13/20)

Cost Cutting Algorithms Are Making Your Job Search a Living Hell
(2/12/20)

Attorneys in unemployment fraud cases join forces, call for state
review of AI in government (2/12/20)

The end of privacy as we know it (2/10/20)

Public bodies are secretly using AI for decisions on people's lives, warns standards watchdog (2/10/20)

An AI regulation strategy that could really work (2/8/20)

An Algorithm That Grants Freedom, or Takes It Away (2/6/20)

How Algorithmic Bias Hurts People With Disabilities (2/6/20)

Who owns your DNA? You should, according to this biodata bill of rights (2/6/20)

Google's ML-fairness-gym lets researchers study the long-term effects of AI's decisions (2/5/20)

Algorithms on social media need regulation, says UK's AI adviser (2/4/20)

EPIC Asks Federal Trade Commission To Regulate Use Of Artificial Intelligence In Pre-Employment Screenings (2/3/20)

Inside the future of online dating: AI swiping and concierge bots (1/31/20)

Why asking an AI to explain itself can make things worse (1/29/20)

Why Amazon's Ring and facial recognition technology are a clear and present danger to society (1/31/20)

An AI Epidemiologist Sent the First Warnings of the Wuhan Virus (1/25/20)

New Jersey cops told to halt all use of controversial facial-recognition technology (1/25/20)

AI Can Do Great Things–if It Doesn't Burn the Planet (1/21/20)

The Secretive Company That Might End Privacy as We Know It (1/18/20)

Your online activity is now effectively a social 'credit score' (1/17/20)

Siri, Alexa and unconscious bias: the case for designing fairer AI assistants (1/17/20)

There's a new obstacle to landing a job after college: Getting approved by AI (1/15/20)

What Does Fairness in AI Mean? (1/15/20)

Are Your Students Bored? This AI Could Tell You (1/13/20)

Cambridge Votes To Ban Face Surveillance Technology (1/13/20)

'Smile with your eyes': How to beat South Korea's AI hiring bots and land a job (1/13/20)

AI can now read emotions – should it? (1/13/20)

Troll Watch: AI Ethics (1/11/20)

Samsung's Neon AI has an ethics problem, and it's as old as sci-fi canon (1/10/20)

Artificial Intelligence Makes Bad Medicine Even Worse (1/10/20)

Germany's plans for automatic facial recognition meet fierce criticism (1/10/20)

Could an AI become your friend? (1/10/20)

Technology Can't Fix Algorithmic Injustice (1/9/20)

San Diego's massive, 7-year experiment with facial recognition technology appears to be a flop (1/9/20)

THE 'ROBOT TAX' DEBATE HEATS UP (1/8/20)

AI Governance in 2019: A Year in Review in Japan (1/8/20)

Twitter bots and trolls promote conspiracy theories about Australian bushfires (1/7/20)

The US just released 10 principles that it hopes will make AI safer (1/7/20)

Dating apps need women. Advertisers need diversity. AI companies offer a solution: Fake people (1/7/20)

Google's AI beats doctors at detecting breast cancer. (Except when it doesn't.) (1/6/20)

AIRBNB CLAIMS ITS AI CAN PREDICT WHETHER GUESTS ARE PSYCHOPATHS (1/4/20)

Rise of #MeTooBots: scientists develop AI to detect harassment in emails (1/3/2020)

Illinois says you should know if AI is grading your online job interviews
(1/1/2020)

Technology Can't Fix Algorithmic Injustice (Jan 2020)

2019

Peer-reviewed

Detect Toxic Content to Improve Online Conversations (2019)

When Hate Speech Leads to Hateful Actions: A Corpus and Discourse
Analytic Approach to Linguistic Threat Assessment of Hate Speech
(2019)

Technology on the margins: AI and global migration management
from a human rights perspective (Dec 2019)

Lessons from Archives: Strategies for Collecting Sociocultural Data in
Machine Learning (12/22/19)

What's Next for AI Ethics, Policy, and Governance? A Global Overview
(12/18/19)

It's easy to fool yourself: Case studies on identifying bias and
confounding in bio-medical datasets (12/12/19)

Explainable Artificial Intelligence (XAI): Concepts, taxonomies,
opportunities and challenges toward responsible AI (12/12/19)

An Unethical Optimization Principle (11/12/19)

Measurement and Fairness (12/11/19)

BERT has a Moral Compass: Improvements of ethical and moral values of machines (12/11/19)

Mapping the Ethicality of Algorithmic Pricing: A Review of Dynamic and Personalized Pricing (12/10/19)

Automatically Neutralizing Subjective Bias in Text (12/5/19)

On the Legal Compatibility of Fairness Definitions (11/25/19)

Mapping the Ethicality of Algorithmic Pricing: A Review of Dynamic and Personalized Pricing (11/25/19)

The Second Wave of Algorithmic Accountability (11/25/19)

The Risks of Using AI to Interpret Human Emotions (11/18/19)

The wrong kind of AI? Artificial intelligence and the future of labour demand (11/18/19)

The Debate on the Ethics of AI in Health Care: a Reconstruction and Critical Review (11/13/19)

In bot we trust: A new methodology of chatbot performance measures (11/1/19)

Principles alone cannot guarantee ethical AI (Nov 2019)

Release Strategies and the Social Impacts of Language Models (Nov 2019)

AI SYSTEMS AS STATE ACTORS (Nov 2019)

Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri (Nov 2019)

Evaluating the Factual Consistency of Abstractive Text Summarization (10/28/19)

Fairness and Privacy in AI/ML Systems (10/28/19)

Dissecting racial bias in an algorithm used to manage the health of populations (10/25/19)

Toward a better trade-off between performance and fairness with kernel-based distribution matching (10/25/19)

In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions (10/23/19)

The Woman Worked as a Babysitter: On Biases in Language Generation (10/23/19)

Quantifying the Carbon Emissions of Machine Learning (10/21/19)

An Information-Theoretic Perspective on the Relationship Between Fairness and Accuracy (10/17/19)

Recommender systems and their ethical challenges (10/13/19)

Bias Detect Neutral Network (10/8/19)

A RIGHT TO REASONABLE INFERENCES: RE-THINKING DATA PROTECTION LAW IN THE AGE OF BIG DATA AND AI (10/5/19)

A fairer way forward for AI in health care (9/25/19)

The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings (9/25/19)

Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead (9/22/19)

Pretrained AI Models: Performativity, Mobility, and Change (9/7/19)

The global landscape of AI ethics guidelines (9/2/19)

'Computer says no': was your mortgage application rejected unfairly (8/8/19)

Selection Bias Explorations and Debias Methods for Natural Language Sentence Matching Datasets (8/2/19)

NHS AI Lab: why we need to be ethically mindful about AI for healthcare (August 2019)

Racial Bias in AI Isn't Getting Better and Neither Are Researchers' Excuses (7/29/19)

Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence (7/18/19)

A Causal Bayesian Networks Viewpoint on Fairness (7/15/19)

Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements (7/13/19)

Artificial Intelligence Governance and Ethics: Global Perspectives (7/4/19)

Artificial Intelligence: the global landscape of ethics guidelines (6/24/19)

A Unified Framework of Five Principles for AI in Society (6/22/19)

"With great power comes great responsibility": keeping public sector algorithms accountable (6/11/2019)

Notes on bias in the socio-material realization of AI technologies (June, 2019)

Transparency as design publicity: explaining and justifying inscrutable algorithms (June, 2019)

Detecting Bias with Generative Counterfactual Face Attribute Augmentation (6/18/19)

Counterfactual Data Augmentation for Mitigating Gender Stereotypes in Languages with Rich Morphology (6/11/19)

SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter (6/9/20)

Evaluating Gender Bias in Machine Translation (6/3/19)

Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies and the Potential of Rational Discourse (Jun 2019)

Racial Bias in Hate Speech and Abusive Language Detection Datasets
(May 2019)

Ethical Dimensions of Visualization Research (May 2019)

Semantics derived automatically from language corpora necessarily
contain human biases (5/25/19)

Translating Principles into Practices of Digital Ethics: Five Risks of Being
Unethical (5/23/19)

Tutorial: Socially Responsible Natural Language Processing (5/13/19)

Stop explaining black box machine learning models for high stakes
decisions and use interpretable models instead (5/13/19)

Limits of Deepfake Detection: A Robust Estimation Viewpoint (5/9/19)

Error terrain analysis for machine learning: Tools and visualizations
(5/6/19)

Data Is the New What? Popular Metaphors & Professional Ethics in
Emerging Data Culture (5/2/19)

Evaluating the Underlying Gender Bias in Contextualized Word
Embeddings (4/18/19)

Discrimination through optimization: How Facebook's ad delivery can
lead to skewed outcomes (4/3/19)

Biased Language in Patient Medical Records: Detection through
Natural Language Processing (NLP) and Impact on Quality (4/2/19)

That's not fair! (April 2019)

The EU-Emotion Voice Database (April 2019)

DATA PROTECTION, ARTIFICIAL INTELLIGENCE AND COGNITIVE
SERVICES IS THE GENERAL DATA PROTECTION REGULATION (GDPR)
"ARTIFICIAL INTELLIGENCE-PROOF" ? (April 2019)

Tay is you. The attribution of responsibility in the algorithmic culture
(3/31/19)

On Measuring Social Biases in Sentence Encoders (3/25/19)

Adversarial attacks on medical machine learning (3/22/19)

Optimized Scoring Systems: Toward Trust in Machine Learning for
Healthcare and Criminal Justice (3/19/20)

DIRTY DATA, BAD PREDICTIONS: HOW CIVIL RIGHTS VIOLATIONS
IMPACT POLICE DATA, PREDICTIVE POLICING SYSTEMS, AND JUSTICE
(3/5/19)

Big data and discrimination (March 2019)

What We Really Mean When We Say "Ethics" (video) (February 2019)

Predictive Inequity in Object Detection (2/21/19)

Artificial Intelligence Crime: An Interdisciplinary Analysis of
Foreseeable Threats and Solutions (2/14/19)

Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation
in 60 Societies (2/8/19)

Discrimination in the Age of Algorithms (2/7/19)

Private Accountability in the Age of Artificial Intelligence (2019)

No training required: Exploring random encoders for sentence classification (Jan 2019)

Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure (Jan 2019)

Fairness in representation: quantifying stereotyping as a representational harm (1/28/19)

Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment (1/23/19)

Attenuating Bias in Word Vectors (1/23/19)

Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions (1/23/19)

Identifying and Correcting Label Bias in Machine Learning (1/1/5/19)

Giving Algorithms a Sense of Uncertainty Could Make Them More Ethical (1/18/19)

Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? (1/7/19)

Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products (Jan 2019)

Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning (Jan 2019)

Government, NGO, and expert publications

AI Safety, Security, and Stability Among Great Powers: Options, Challenges, and Lessons Learned for Pragmatic Engagement (Dec 2019)

Recommender systems and their ethical challenges (2019)

Beyond Bias: Contextualizing "Ethical AI" Within the History of Exploitation and Innovation in Medical Research (12/21/19)

NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software (12/19/19)

Fairness Indicators: Scalable Infrastructure for Fair ML Systems (12/11/19)

2019 AI Index Report (December 2019)

Disability, Bias, and AI (Nov 2019)

AI 360: Hold, Fold, or Double Down (Nov 2019)

The value of a shared understanding of AI models (Nov 2019)

Supporting rights-respecting AI (11/26/19)

The Second Wave of Algorithmic Accountability (11/25/19)

Opinion: AI For Good Is Often Bad (11/18/19)

The Ethical Dilemma at the Heart of Big Tech Companies (11/14/19)

Delivering the benefits of Custom Neural Voice (11/12/19)

AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense (Nov 2019)

How Machine Learning Pushes Us to Define Fairness (11/6/19)

Expanding Frameworks: An economic justice approach to digital privacy (11/6/19)

Using Tensorflow Fairness Indicators (11/1/19)

Fair ML for Health Resources (Oct 2019)

Artificial Intelligence Research Needs Responsible Publication Norms (10/24/19)

The AI Ethics Charter by Korea Artificial Intelligence Ethics Association (10/23/19)

Introducing AI Blindspot: A Call for Tech to Think Holistically and Spot Risks (10/22/19)

AI-enabled human rights monitoring (10/17/19)

Fairness in Clustering with Multiple Sensitive Attributes (10/11/19)

Austria's employment agency rolls out discriminatory algorithm, sees no problem (10/6/19)

New York City Police Department Surveillance Technology (10/4/19)

This machine read 3.5 million books then told us what it thought about men and women (9/30/19)

Building ML models for everyone: understanding fairness in machine learning (9/25/19)

Unpacking "Ethical AI": A curated reading list (9/24/19)

LITIGATING ALGORITHMS 2019 US REPORT: New Challenges to Government Use of Algorithmic Decision Systems (9/24/19)

STREET–LEVEL ALGORITHMS AND SEEING THE FOREST FOR THE TREES IN AI (9/23/19)

Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy (9/17/19)

AI ethics and the limits of code(s) (9/16/19)

Face recognition, bad people and bad data (9/9/19)

A Draft Syllabus/Curriculum for an Ethics of AI Course (9/1/19)

How New A.I. Is Making the Law's Definition of Hacking Obsolete (8/21/19)

U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools (8/9/19)

AI Justice: When AI Principles Are Not Enough (8/5/19)

WEF: Responsible Use of Technology (August 2019)

Research Summary: Social Robots and Empathy: The Harmful Effects of Always Getting What We Want (7/27/19)

AI & Global Governance: Human Rights and AI Ethics – Why Ethics Cannot be Replaced by the UDHR (7/19/19)

Gov.UK: Interim report: Review into bias in algorithmic decision-making (7/19/19)

Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns (7/17/19)

Gender diversity in AI research (7/17/19)

Open View: AI/ML Fairness (7/13/19)

IDC Survey Finds Artificial Intelligence to be a Priority for Organizations But Few Have Implemented an Enterprise-Wide Strategy (7/8/19)

Differential Privacy in the 2020 Decennial Census and the Implications for Available Data Products (7/8/19)

The Foundation of Responsible Artificial Intelligence (7/7/19)

AI Differential Privacy and Federated Learning (7/1/19)

Fluxus Landscape: An Expansive View of AI Ethics and Governance (6/26/19)

Five practical steps to make Artificial Intelligence (AI) interpretable (6/21/19)

Researchers develop 'vaccine' against attacks on machine learning (6/20/19)

Ethics of Emotion AI - Part 1 (6/19/19)

The future of women at work: Transitions in the age of automation (6/12/19)

Ethical AI: Tensions and trade-offs (6/11/19)

Ethical Codex for Data- Based Value Creation (6/10/19)

AI is not as advanced as you might think (6/10/19)

Introducing the Principled Artificial Intelligence Project (6/7/19)

Does object recognition work for everyone? A new method to assess bias in CV systems (6/7/19)

Deepfakes and Synthetic Media: Updated Survey of Solutions against Malicious Usages (June 2019)

Emotional Artificial Intelligence guidelines for ethical use (June 2019)

Systemic Algorithmic Harms (5/31/19)

Responsible AI: A Global Policy Framework (5/29/19)

Accountable AI: Can the Law Reduce Bias and Increase Transparency? (5/20/19)

Tech workers call time on moving fast and breaking things (May 2019)

Designing AI for Social Good: Seven Essential Factors (May 2019)

From What to How: An Overview of AI Ethics Tools, Methods and
Research to Translate Principles into Practices (May 2019)

GARBAGE IN, GARBAGE OUT: FACE RECOGNITION ON FLAWED DATA
(5/16/19)

What the Fifth Industrial Revolution is and why it matters (5/15/19)

When Humans Attack: Re-thinking safety, security, and AI (5/14/19)

People + AI Guidebook (May 2019)

World Economic Forum's AI head on how to protect human rights
without stifling innovation (4/29/19)

Recommender Systems and their Ethical Challenges (April 2019)

DATA PROTECTION, ARTIFICIAL INTELLIGENCE AND COGNITIVE
SERVICES IS THE GENERAL DATA PROTECTION REGULATION (GDPR)
"ARTIFICIAL INTELLIGENCE-PROOF"? (April 2019)

How AI can enable a sustainable future (April 2019)

Artificial Intelligence and Human Rights (April 2019)

DISCRIMINATING SYSTEMS: Gender, Race, and Power in AI (April
2019)

Gender, Race, and Power in AI (4/17/19)

From Principles to Action: How do we Implement Tech Ethics? (4/17/19)

Untold History of AI: Algorithmic Bias Was Born in the 1980s (4/15/19)

The Need to Regulate AI Implementation in Public Assistance Programs (4/12/19)

Laying down the law on AI: ethics done, now the EU must focus on human rights (4/8/19)

The Automated Administrative State (4/8/19)

UK to introduce world first online safety laws (4/8/19)

Artificial Intelligence: Australia's Ethics Framework: A discussion (4/5/19)

AI Ethics: Seven Traps (3/25/19)

Google: Perspectives on Issues in AI Governance (March 2019)

Agile Ethics: Managing Ethical Complexity in Technology (3/14/19)

BENEFITS & RISKS OF ARTIFICIAL INTELLIGENCE (March 2019)

Pentagon Seeks 'Ethical Principles' for AI Use (3/4/19)

Intel's Recommendations for the U.S. National Strategy on Artificial Intelligence (White Paper) (2/6/19)

Ethics of AI in Radiology: European and North American Multisociety Statement (February 2019)

In Favor of Developing Ethical Best Practices in AI Research (2/21/19)

The Environment for Ethical Action (2/18/19)

Council of Europe: Algorithmic Used To Manipulate Social And Political Behaviours (2/18/19)

How the Economics of Data Science is Creating New Sources of Value (2/16/19)

Public deliberation could help address AI's legitimacy problem in 2019 (2/8/19)

Start with values: How innovation can be bold, fast and responsible (1/31/19)

AI in Context: The Labor of Integrating New Technologies (1/30/19)

The ethical and political questions raised by AI (1/29/19)

Illinois Supreme Court Holds That Biometric Privacy Law Does Not Require Actual Harm for Private Suits (1/29/19)

Fairness in representation: quantifying stereotyping as a representational harm (1/28/19)

Ignorance and Distrust Prevail about What Companies and Governments Do with Personal Data (1/25/19)

Automation and artificial intelligence: How machines are affecting people and places (1/24/19)

Building Ethically Aligned AI (1/23/19)

The Era of "Move Fast and Break Things" Is Over [HBR] (1/22/19)

Ethics and technology in the Fourth Industrial Revolution (1/21/19)

Why AI is failing the next generation of women (1/18/19)

Creating AI Systems That Take Culture into Account (1/17/19)

Overcoming barriers to AI adoption (1/16/19)

Facebook Algorithms and Personal Data (Pew Survey) (1/16/19)

Case study: Algorithmic decision-making and accountability (Jan 2019)

Singapore Proposed Model Artificial Intelligence Governance Framework (Jan 2019)

Artificial Intelligence: American Attitudes and Trends (Jan 2019)

Future of work research papers (Jan 2019)

News/popular press

Fight against facial recognition hits wall across the West (12/30/19)

Designing a Moral Compass for AI (12/30/19)

Artificial Intelligence Is Rushing Into Patient Care–And Could Raise Risks (12/24/19)

AI year in review: Opportunities grow, but ethics loom large (12/23/19)

'The Algorithm Made Me Do It': Artificial Intelligence Ethics Is Still On Shaky Ground (12/22/19)

AI is biased. Here's how scientists are trying to fix it (12/21/19)

Beyond Bias: Contextualizing "Ethical AI" Within the History of Exploitation and Innovation in Medical Research (12/20/19)

Federal study of top facial recognition algorithms finds 'empirical evidence' of bias (12/20/19)

THE INVENTION OF "ETHICAL AI": How Big Tech Manipulates Academia to Avoid Regulation (12/20/19)

AI Is Biased. Here's How Scientists Are Trying to Fix It (12/19/19)

Ethics In AI: Why Values For Data Matter (12/18/19)

Human Rights Commission wants privacy laws adjusted for an AI future (12/17/19)

Researchers were about to solve AI's black box problem, then the lawyers got involved (12/17/19)

Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement (12/13/19)

Emotion recognition technology should be banned, says an AI research institute (12/13/19)

The AI community needs to take responsibility for its technology and its actions (12/13/19)

AI expert calls for end to UK use of 'racially biased' algorithms (12/12/19)

Emotion-detecting tech should be restricted by law - AI Now (12/12/19)

Stanford, Kyoto & Georgia Tech Model 'Neutralizes' Biased Language (12/11/19)

The Utility of Interpretable AI (12/10/19)

Microsoft tech expert warns of bias and sexism in artificial intelligence (12/10/19)

As AI moves into content creation, researchers aim to battle its biases (12/9/19)

How practitioners and academics think (and then forget) about fairness when building AI systems (12/6/19)

Biased Algorithms Are Easier to Fix Than Biased People (12/6/19)

Life Under the Algorithm: How a relentless speedup is reshaping the working class (12/4/19)

Human insight remains essential to beat the bias of algorithms (12/4/19)

UK GPDR Watchdog: Explain Your AI (12/3/19)

Portland plans to propose the strictest facial recognition ban in the country (12/2/19)

Chinese tech groups shaping UN facial recognition standards (12/1/19)

Dealing With Bias in Artificial Intelligence (Nov 2019)

An epidemic of AI misinformation (11/30/19)

Explain yourself, mister: Fresh efforts at Google to understand why an AI system says yes or no (11/29/19)

The first effort to regulate AI was a spectacular failure (11/26/19)

Tainted Data Can Teach Algorithms the Wrong Lessons (11/25/19)

Human biases are baked into algorithms. Now what? (11/25/19)

Scientists developed a new AI framework to prevent machines from misbehaving (11/25/19)

Can We Force AIs to Be Fair Towards People? Scientists Just Invented a Way (11/25/19)

Tuning Out Toxic Comments, With the Help of AI (11/21/19)

Stanford, UMass Amherst develop algorithms that train AI to avoid specific misbehaviors (11/21/19)

Researchers Want Guardrails to Help Prevent Bias in AI (11/21/19)

Dealing with bias in AI (11/19/19)

The Apple Card Didn't 'See' Gender–and That's the Problem
(11/19/19)

There's an easy way to make lending fairer for women. Trouble is, it's
illegal. (11/15/19)

The Real AI Threat to Democracy (11/14/19)

When Algorithms Decide Whose Voices Will Be Heard (11/12/19)

Google's secret cache of medical data includes names and full details
of millions – whistleblower (11/12/19)

These Black Women Are Fighting For Justice In A World Of Biased
Algorithms (11/12/19)

Apple co-founder Steve Wozniak says Apple Card discriminated against
his wife (11/11/19)

'You sound worried': would you let an AI rephrase the tone of your
emails? (11/6/19)

AI Can Make Bank Loans More Fair (11/6/20)

Rights group files federal complaint against AI-hiring firm HireVue,
citing 'unfair and deceptive' practices (11/6/19)

Artificial Intelligence Can Be Biased. Here's What You Should Know.
(11/5/19)

Microsoft's AI Research Draws Controversy Over Possible
Disinformation Use (11/4/19)

How to train artificial intelligence that won't destroy the environment (11/4/19)

AI is making literary leaps – now we need the rules to catch up (11/2/19)

Pentagon's draft AI ethics guidelines fight bias and rogue machines (11/2/19)

AI-Related Lawsuits Are Coming (11/1/20)

NYPD Built Bias Safeguards Into Pattern-Spotting AI System (10/31/19)

Using AI to Eliminate Bias from Hiring (10/29/19)

Big Data and Racial Bias: Can That Ghost Be Removed from the Machine? (10/28/19)

Why did Microsoft fund an Israeli firm that surveils West Bank Palestinians? (10/28/19)

The hidden costs of AI (10/26/19)

A health care algorithm affecting millions is biased against black patients (10/24/19)

Artificial Intelligence Research Needs Responsible Publication Norms (10/24/19)

How much "fake news" is on Twitter?

When Binary Code Won't Accommodate Nonbinary People (10/23/19)

AN ALGORITHM FOR DE-BIASING AI SYSTEMS (10/23/19)

After Backlash, Portland-Area Outback Steakhouse Cancels New Artificial Intelligence Surveillance Test (10/23/19)

A face-scanning algorithm increasingly decides whether you deserve the job (10/22/19)

These Startups Are Building Tools to Keep an Eye on AI (10/21/19)

Military artificial intelligence can be easily and dangerously fooled (10/21/19)

Facial recognition: This new AI tool can spot when you are nervous or confused (10/21/19)

New framework makes AI systems more transparent without sacrificing performance (10/21/19)

Fighting hate with AI-powered retorts (1019/19)

Can you make AI fairer than a judge? Play our courtroom algorithm game (10/17/19)

Facial recognition AI can't identify trans and non-binary people (10/16/19)

Zombie debts are hounding struggling Americans. Will you be next? (10/15/19)

Bias and Algorithmic Fairness: The modern business leader's new responsibility in a brave new world ruled by data. (10/13/19)

Ethical AI: Why governments, companies are a worried lot (10/10/19)

Rasa's conversational AI can selectively ignore bits of dialogue to improve its responses (10/9/19)

Beware of automated hiring (10/8/19)

How to operationalize AI ethics (10/7/19)

Austria's employment agency rolls out discriminatory algorithm, sees no problem (10/6/19)

Google causing more facial recognition problems, machine learning goes quantum and losing a job if an AI doesn't like your face (10/6/19)

Artificial stupidity: 'Move slow and fix things' could be the mantra AI needs (10/5/19)

An Open Source License That Requires Users to Do No Harm (10/4/19)

Congress Is Investigating the Military's Use of Facial Recognition (10/3/19)

The Unsettling Ways Tech Is Changing Your Personal Reality (10/3/19)

Bias and Algorithmic Fairness (10/3/19)

Blind Spots in AI Just Might Help Protect Your Privacy (10/2/19)

British workers are deliberately sabotaging robots amid fears they will take their jobs, study finds (9/29/19)

"The New Jim Code" – Ruha Benjamin on racial discrimination by algorithm (9/26/19)

Reddit and Gab's most toxic communities inadvertently train AI to combat hate speech (9/25/19)

Africa Is Building an A.I. Industry That Doesn't Look Like Silicon Valley (9/25/19)

The Viral App That Labels You Isn't Quite What You Think (9/19/20)

See how an AI system classifies you based on your selfie (9/17/19)

AI ethics and the limits of code(s) (9/16/19)

120 million workers will need to be retrained due to AI, says IBM study (9/6/19)

An AI-Run World Needs to Better Reflect People of Color (9/6/19)

'Sense of urgency', as top tech players seek AI ethical rules (9/2/19)

How to Build an AI Ethics Committee (8/30/19)

Silicon Valley's Secret Philosophers Should Share Their Work (8/27/19)

'Dangerous' AI offers to write fake news (8/27/19)

Amazon, Microsoft, 'putting world at risk of killer AI': study (8/21/19)

Flawed Algorithms Are Grading Millions of Students' Essays (8/20/19)

An AI privacy conundrum? The neural net knows more than it says (8/19/19)

AI can read your emotions. Should it? (8/17/19)

'Stereotyping' emotions is getting in the way of artificial intelligence. Scientists say they've discovered a better way. (8/16/19)

Do Tech Companies Really Need to Snoop Into Private Conversations to Improve Their A.I.? (8/14/19)

Training the next generation of ethical techies (8/14/19)

Training bias in AI "hate speech detector" means that tweets by Black people are far more likely to be censored (8/14/19)

What your voice reveals about you (8/13/19)

Google's Artificial Intelligence Hate Speech Detector Has a 'Black Tweet' Problem (8/13/19)

Google's algorithm for detecting hate speech is racially biased (8/13/19)

GOOGLE'S HATE SPEECH-DETECTING AI IS BIASED AGAINST BLACK PEOPLE (8/12/19)

Facial recognition software mistook 1 in 5 California lawmakers for criminals, says ACLU (8/12/2019)

IBM Research launches explainable AI toolkit (8/8/19)

AI Needs Your Data–and You Should Get Paid for It (8/8/19)

'Computer says no': was your mortgage application rejected unfairly (8/8/19)

AI Ethics Guidelines Every CIO Should Read (8/7/19)

China has started a grand experiment in AI education. It could reshape how the world learns (8/2/19)

Artificial intelligence could improve healthcare for all–unless it doesn't (8/1/19)

WE TESTED EUROPE'S NEW LIE DETECTOR FOR TRAVELERS – AND IMMEDIATELY TRIGGERED A FALSE POSITIVE (7/26/19)

AI 'emotion recognition' can't be trusted (7/25/19)

Fitbits and other wearables may not accurately track heart rates in people of color (7/24/19)

Microsoft and the learnings from its failed Tay artificial intelligence bot (7/24/19)

HIGH-STAKES AI DECISIONS NEED TO BE AUTOMATICALLY AUDITED (7/18/19)

How companies can avoid ethics washing (7/17/19)

Facial Recognition Tech Is Growing Stronger, Thanks to Your Face (7/13/19)

Ethical Artificial Intelligence Becomes A Supreme Competitive Advantage (7/7/19)

Botched Adidas Twitter Campaign Spits out Racist and Anti-Semitic Arsenal Jerseys (7/2/19)

What Does an AI Ethicist Do? (6/24/19)

AN AI "VACCINE" CAN BLOCK ADVERSARIAL ATTACKS (6/20/19)

AI ain't for everyone – who trusts bots, and why (6/14/29)

Credit Scores Could Soon Get Even Creepier and More Biased (6/13/19)

San Francisco says it will use AI to reduce bias when charging people with crimes (6/12/19)

THE NEXT BIG PRIVACY HURDLE? TEACHING AI TO FORGET (6/12/19)

AI is worse at identifying household items from lower-income countries (6/11/19)

Illinois Bill Aims to Limit Use of AI in Video Job Interviews (6/11/19)

How Big Tech funds the debate on AI ethics (6/6/19)

How companies like Google are dealing with the ethics of AI (6/4/19)

Stop saying "robots are coming for your job"; start saying "Your boss wants to replace you with a robot" (6/3/19)

Ethics as a competitive advantage in the booming artificial intelligence industry (6/3/19)

This Creepy AI Predicts What You Look Like Based on Your Voice (6/1/19)

Artificial intelligence, the future of work, and inequality (5/31/19)

Addressing the Biases Plaguing Algorithms (5/31/19)

10 things we should all demand from Big Tech right now (5/29/19)

AI is only human (5/28/19)

To Fight Deepfakes, Researchers Built a Smarter Camera (5/28/19)

Amazon venturing into emotionally-savvy Artificial Intelligence at the cost of your Privacy (5/28/19)

When algorithms mess up, the nearest human gets the blame (5/28/19)

Top 5 Tools Data Scientists Can Use To Mitigate Biases In Algorithms (5/28/19)

America and its economic allies have announced five "democratic" principles for AI (5/22/19)

The Best Reason for Your City to Ban Facial Recognition (5/17/19)

Voice Recognition Still Has Significant Race and Gender Biases (5/10/19)

The Legislation That Targets the Racist Impacts of Tech (5/7/19)

All the ways hiring algorithms can introduce bias (5/6/19)

Who to sue when a robot loses your fortune (5/5/19)

The danger of AI is weirder than you think (April 2019)

Why your board needs an AI council (4/28/19)

WILL ARTIFICIAL INTELLIGENCE ENHANCE OR HACK HUMANITY? (4/28/19)

FOR THE SUCCESSFUL ADOPTION OF AI, WE NEED MORE WOMEN LEADERS (4/24/19)

How AI can help close the gender pay gap and eliminate bias (4/24/19)

Microsoft is winning the techlash (4/24/19)

Shazeda Ahmed on the Messy Truth About Social Credit (4/23/19)

Ethics committee raises alarm over 'predictive policing' tool (4/20/19)

How killer robots overran the UN (4/19/19)

Some AI just shouldn't exist (4/19/19)

Emotionally intelligent AI will respond to how you feel (4/17/19)

A.I. Bias Isn't the Problem. Our Society Is (4/16/19)

Google reportedly disbands review panel monitoring DeepMind Health AI (4/15/19)

Discrimination's Digital Frontier (4/15/19)

What is fair when it comes to AI bias? (4/12/19)

Do no evil: why we need a public conversation about AI ethics (4/12/19)

A new bill would force companies to check their algorithms for bias (4/10/19)

Microsoft's Brad Smith on How to Responsibly Deploy AI (4/10/19)

AI's leading developers share their fears about the tech developing too fast (4/9/19)

Why Artificial Intelligence Needs Democratic Governance (4/9/19)

How will AI change your life? AI Now Institute founders Kate Crawford and Meredith Whittaker explain. (4/8/19)

It's disturbingly easy to trick AI into doing something deadly (4/8/19)

Tech titans declare AI ethics concerns (4/8/19)

What we lost when we lost Google ATEAC (4/7/19)

UK businesses using artificial intelligence to monitor staff activity (4/7/19)

Real or artificial? Tech titans declare AI ethics concerns (4/7/19)

Hey Google, sorry you lost your ethics council, so we made one for you (4/6/19)

Facebook's ad-serving algorithm discriminates by gender and race (4/5/19)

"Color-blindness" is a bad approach to solving bias in algorithms (4/3/19)

Algorithms have gotten out of control. It's time to regulate them. (4/3/19)

The problem with AI ethics (4/3/19)

The Pitfalls of Data's Gender Gap (4/1/19)

'Bias deep inside the code': the problem with AI 'ethics' in Silicon Valley (3/29/19)

Inmates in Finland are training AI as part of prison labor (3/28/19)

RISK ASSESSMENT: EXPLAINED (3/27/19)

Rethinking Privacy For The AI Era (3/27/19)

Ethical question takes center stage at Silicon Valley summit on artificial intelligence (3/27/19)

How malevolent machine learning could derail AI (3/25/19)

CAN AI BE A FAIR JUDGE IN COURT? ESTONIA THINKS SO (3/25/19)

How Pope Francis could shape the future of robotics (3/24/19)

Human Contact Is Now a Luxury Good (3/23/19)

Our Software Is Biased Like We Are. Can New Laws Change That?
(3/23/19)

The Dog That Did Not Bark In the Night: The Danger of Survivor Bias in
AI (3/22/19)

Deciding how to decide: Six key questions for reducing AI's democratic
deficit (3/22/19)

Warnings of a Dark Side to A.I. in Health Care (3/12/19)

When AI speaks on behalf of humans: Proposing ethical guidelines
based on Google Duplex assistant (3/12/19)

AI, ethics and data bias - getting it right (3/20/19)

The world's first genderless AI voice is here. Listen now (3/19/19)

To Be Ethical, AI Must Become Explainable. How Do We Get There?
(3/19/19)

As AI Spreads, Tech Needs 'Chief Bias Officers' (3/18/19)

Tackling Bias in Machine Learning (3/18/19)

Courts and police departments are turning to AI to reduce bias, but
some argue it'll make the problem worse (3/17/19)

Is Microsoft AI Helping to Deliver China's 'Shameful' Xinjiang
Surveillance State? (3/15/19)

Without Humans, A.I. Can Wreak Havoc (3/12/19)

The NYPD Now Has an Automated Tool to Help Them Recognize
Patterns in Crimes (3/10/19)

The Achilles' Heel Of AI (3/7/19)

Does AI Ethics Have A Bad Name? (3/7/19)

Don't look now: why you should be worried about machines reading
your emotions (3/6/19)

A new study finds a potential risk with self-driving cars: failure to detect
dark-skinned pedestrians (3/6/19)

The infamous AI gaydar study was repeated – and, no, code can't tell if
you're straight or not just from your face (3/5/19)

Should we be treating algorithms the same way we treat hazardous
chemicals (3/5/19)

AI will reproduce and enshrine age-old biases–if we let it (3/5/19)

How AI Will Rewire Us (3/5/19)

Some self-driving car systems have trouble detecting darker skin, study
says (3/5/19)

China's tech billionaires back ethical rules to guide development of AI
and other technologies (3/4/19)

Seeking Ground Rules for A.I. (3/4/19)

Apple is hiring an analyst to explain Siri complaints to executives (3/4/19)

16 Uncomfortable Questions Everyone Needs to Ask About Artificial Intelligence (3/3/19)

Is Ethical A.I. Even Possible? (3/1/19)

Why AI is a threat to democracy—and what we can do to stop it (2/26/19)

Chinese police test gait-recognition technology from AI start-up Watrix that identifies people based on how they walk (2/26/19)

Troubling Trends Towards Artificial Intelligence Governance (2/25/19)

AI researchers debate the ethics of sharing potentially harmful programs (2/21/19)

China Uses DNA to Track Its People, With the Help of American Expertise (2/21/19)

Pope Francis and Microsoft team up to promote prize for artificial intelligence (2/19/19)

OpenAI: Social science, not just computer science, is critical for AI (2/19/19)

Over 2000 people died after receiving Centrelink robo-debt notice, figures reveal (2/18/19)

AI in the UK: The only way is ethics (2/18/19)

Bentham, Hobbes, and The Ethics of Artificial Intelligence (2/17/19)

Are the algorithms that power dating apps racially biased? (2/17/19)

New AI fake text generator may be too dangerous to release, say creators (2/14/19)

THE REAL REASON TECH STRUGGLES WITH ALGORITHMIC BIAS (2/12/19)

GOOGLE AND MICROSOFT WARN THAT AI MAY DO DUMB THINGS (2/11/19)

YouTube announces it will no longer recommend conspiracy videos (2/10/19)

How Vodafone's Chatbot Technology Is Helping It Cut Jobs (2/8/19)

Women Stand Against Social Injustice In AI (2/7/19)

I Cut the 'Big Five' Tech Giants From My Life. It Was Hell (2/7/19)

Microsoft warns investors that its artificial-intelligence tech could go awry and hurt its reputation (2/6/19)

Goodbye, trolley problem. This is Silicon Valley's new ethics test. (2/5/19)

SUPPOSEDLY 'FAIR' ALGORITHMS CAN PERPETUATE DISCRIMINATION (2/5/19)

Microsoft warned investors that biased or flawed AI could hurt the company's image (2/5/19)

Five Takeaways from an AI Fairness Conference (2/2/19)

JUSTICE BY THE NUMBERS: MEET THE STATISTICIAN TRYING TO FIX BIAS IN CRIMINAL JUSTICE ALGORITHMS (2/1/19)

Guidelines for human-AI interaction design (2/1/19)

Automated background checks are deciding who's fit for a home (2/1/19)

Getting efficient with "What-happens-if …" (2/1/19)

A.I. Could Worsen Health Disparities (1/31/19)

Google's head of translation on fighting bias in language and why AI loves religious texts (1/30/19)

This Is Your Brain Off Facebook (1/30/19)

Prisons Are Building Databases of Inmates' Voice Prints (1/30/19)

PRISONS ACROSS THE U.S. ARE QUIETLY BUILDING DATABASES OF INCARCERATED PEOPLE'S VOICE PRINTS (1/30/19)

New York Insurers Can Evaluate Your Social Media Use–If They Can Prove Why It's Needed (1/30/19)

San Francisco proposal would ban government facial recognition use in the city (1/29/19)

ML Integrity: Four Production Pillars For Trustworthy AI (1/29/19)

IBM releases Diversity in Faces, a dataset of over 1 million annotations to help reduce facial recognition bias (1/29/19)

The Unnatural Ethics of AI Could Be Its Undoing (1/29/19)

How to lift the veil off hidden algorithms (1/28/19)

SECURITY ISN'T ENOUGH. SILICON VALLEY NEEDS 'ABUSABILITY' TESTING (1/28/19)

Being Ethical: How to Beat Bias in AI and Be Transparent (1/28/19)

MIT hopes to automatically 'de-bias' face detection AI (1/27/19)

Speaking Black Dialect in Courtrooms Can Have Striking Consequences (1/25/19)

Bias in, bias out: the Stanford scientist out to make AI less white and male (1/25/19)

The Hidden Automation Agenda of the Davos Elite (1/25/19)

MIT researchers: Amazon's Rekognition shows gender and ethnic bias (updated) (1/24/19)

Amazon Is Pushing Facial Technology That a Study Says Could Be Biased (1/24/19)

Quarter of U.S. jobs could be jeopardized by AI, research shows (1/24/19)

A.I. and Automation Will Hit Low-Skill Jobs and Trump Swing States Hardest (1/24/19)

Workers in heartland states most at risk of losing jobs to AI, new study finds (1/24/19)

Automation is a bigger threat to inland California workers, study finds (1/24/19)

Can we make artificial intelligence ethical? (1/23/19)

Artificial Intelligence and Ethics (Jan/Feb 2019)

The AI Arms Race Means We Need AI Ethics (1/22/19)

He Said, She Said: Addressing Gender in Neural Machine Translation (1/22/19)

What makes AI ethicists "the top hire companies need to succeed"? (1/21/19)

A.I. Is Sending People to Jail—and Getting It Wrong (1/21/19)

The goal is to automate us': welcome to the age of surveillance capitalism (1/20/19)

What makes AI ethicists "the top hire companies need to succeed"? (1/21/19)

Why Is AI And Machine Learning So Biased? The Answer Is Simple Economics (1/20/19)

Giving algorithms a sense of uncertainty could make them more ethical (1/18/19)

How This Chicago Company Is Using Predictive Analytics To Decrease Bias (1/17/19)

US military trusted more than Google, Facebook to develop AI: survey (1/14/19)

The weaponization of artificial intelligence (1/14/19)

AI can wreak havoc if left unchecked by humans (1/14/19)

Facial and emotional recognition; how one man is advancing artificial intelligence (1/13/19)

One day your voice will control all your gadgets, and they will control you (1/11/19)

How a Google Researcher Is Making AI Easier to Understand (1/10/19)

Americans want to regulate AI but don't trust anyone to do it (1/10/19)

Never mind killer robots–here are six real AI dangers to watch out for in 2019 (1/7/19)

The Future Of Recruiting In The Age of Automation And Artificial Intelligence (1/4/19)

2019 Is the Year to Stop Talking About Ethics and Start Taking Action (1/4/19)

How a Feel-Good AI Story Went Wrong in Flint (1/3/19)

New App Helps People Remember Faces (1/1/19)

2018

Peer-reviewed

IMPLEMENTING ETHICS INTO ARTIFICIAL INTELLIGENCE: A CONTRIBUTION, FROM A LEGAL PERSPECTIVE, TO THE DEVELOPMENT OF AN AI GOVERNANCE REGIME (2018)

Ethics as an Escape from Regulation: From ethics-washing to ethics shopping (2018)

The Bias Bias in Behavioral Economics (12/31/18)

Fairness Behind a Veil of Ignorance: A Welfare Analysis for Automated Decision Making (Dec 2018)

Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions (12/16/18)

Improving fairness in machine learning systems: What do industry practitioners need? (12/13/18)

Racial Influence on Automated Perceptions of Emotions (12/6/18)

Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data (12/1/18)

Artificial Intelligence in the Context of Crime and Criminal Justice (Dec 2018)

Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science (Dec 2018)

Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved (11/27/18)

50 Years of Test (Un)fairness: Lessons for Machine Learning (11/25/18)

An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations

Even If AI Can Cure Loneliness – Should It? (11/9/18)

Some Requests for Machine Learning Research from the East African Tech Scene (11/8/18)

Fairness and Abstraction in Sociotechnical Systems (11/7/18)

Consumer-Lending Discrimination in the Era of FinTech (Oct 2018)

The dangers of automating social programs (Oct 2018)

AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations (10/28/18)

Learning Adversarially Fair and Transferable Representations (10/22/18)

Gender-Aware Natural Language Translation (10/8/18)

Model Cards for Model Reporting (10/5/18)

Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations (10/1/18)

Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice (September 2018)

Language Modeling Teaches You More Syntax than Translation Does: Lessons Learned Through Auxiliary Task Analysis (9/26/18)

TRANSPARENCY AND EXPLANATION IN DEEP REINFORCEMENT LEARNING NEURAL NETWORKS (9/17/18)

World Economic Forum - The Future of Work 2018 Report (Sept 2018)

A Moral Framework for Understanding of Fair ML through Economic Models of Equality of Opportunity (9/10/18)

Assessing Gender Bias in Machine Translation -- A Case Study with Google Translate (9/6/18)

How AI can be a force for good (8/24/18)

Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual & Group Unfairness via Inequality Indices (8/19/18)

The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning (8/14/18)

Building Safer AGI by introducing Artificial Stupidity (8/11/18)

A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics (7/2/18)

GENDER SHADES: INTERSECTIONAL ACCURACY DISPARITIES IN COMMERCIAL GENDER CLASSIFICATION (July 2018)

A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions (July 2018)

Achieving Fairness through Adversarial Learning: an Application to
Recidivism Prediction (6/30/18)

A review of possible effects of cognitive biases on interpretation of
rule-based machine learning models (6/27/18)

Towards Accountable AI: Hybrid Human-Machine Analyses for
Characterizing System Failure (6/15/18)

AI and the economy (6/13/18)

Fairness definitions explained (May 2018)

Why is my classifier discriminatory? (5/30/18)

ENHANCING THE ACCURACY AND FAIRNESS OF HUMAN DECISION
MAKING (5/25/18)

To Build Truly Intelligent Machines, Teach Them Cause and Effect
(5/15/18)

The Dataset Nutrition Label: A Framework To Drive Higher Data Quality
Standards (5/9/18)

Addressing Age-Related Bias in Sentiment Analysis (April 2018)

Let's talk about race: Identity, chatbots, and AI (April 2018)

Fairness and Accountability Design Needs for Algorithmic Support in
High-Stakes Public Sector Decision-Making (4/21/18)

Modeling and Simultaneously Removing Bias via Adversarial Neural
Networks (4/18/18)

Word embeddings quantify 100 years of gender and ethnic stereotypes (4/17/18)

Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions (4/10/18)

It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process (3/29/18)

DATASHEETS FOR DATASETS (3/23/18)

Data Statements for NLP: Toward Mitigating System Bias and Enabling Better Science (2/23/18)

DELAYED IMPACT OF FAIR MACHINE LEARNING (3/12/18)

A Reductions Approach to Fair Classification (3/6/18)

Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making (2/318)

Measuring and Mitigating Unintended Bias in Text Classification (2/1/18)

AN AI RACE FOR STRATEGIC ADVANTAGE: RHETORIC AND RISKS (February 2018)

Transparent Model Distillation (1/26/18)

Mitigating Unwanted Biases with Adversarial Learning (1/22/18)

The accuracy, fairness, and limits of predicting recidivism (1/17/18)

AI: from rational agents to socially responsible agents (2018)

Government, NGO, and expert publications

Discrimination, artificial intelligence, and algorithmic decision-making (2018)

Detecting Bias in Amazon reviews (2018)

AI Index 2018 Annual Report (December 2018)

AI Narratives Report (12/11/18)

The AI Maturity Playbook (12/11/18)

Artificial Intelligence and the Future of Humans (Pew Survey) (12/10/18)

AI Now 2018 Report (12/6/18)

AI Now 2018 Recommendations (12/6/18)

AI Fairness for People with Disabilities: Point of View (11/26/18)

Public Attitudes Toward Computer Algorithms (Pew Survey) (11/16/18)

Gender, Race and Power: Outlining a New AI Research Agenda (11/15/18)

AI4People–An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations (11/2/18)

Portrayals and perceptions of AI and why they matter (Nov 2018)

Applying AI for Social Good: McKinsey Report (Nov 2018)

WHY DOES ARTIFICIAL INTELLIGENCE DISCRIMINATE? (10/24/18)

Declaration on Ethics and Data Protection in Artificial Intelligence by 40th International Conference of Data Protection and Privacy Commissioners (10/23/18)

Big data and AI: Ethical and societal implications (10/16/18)

Governing Artificial Intelligence: Upholding Human Rights & Dignity (10/10/18)

In Advanced and Emerging Economies Alike, Worries About Job Automation (Pew: Survey) (9/13/18)

Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System (Sept 2018)

Cognitive diversity: AI and the future of work (Sept 2018)

KPMG: AI Risk and Controls Matrix (Sept 2018)

Playing with fairness (Sept. 2018)

An ethics checklist for data scientists

Brookings survey finds divided views on artificial intelligence for warfare, but support rises if adversaries are developing it (8/29/18)

Factsheets for AI Services (8/22/18)

Ethical OS: A guide to anticipating the future impact of today's technology (2018)

Humans Plus Robots: Why the Two Are Better Than Either One Alone (7/12/18)

Critical Perspectives on Artificial Intelligence and Human Rights (6/19/18)

Data ethics framework (6/13/18)

Analyzing & Preventing Unconscious Bias in Machine Learning (6/12/18)

How good are Google's new AI ethics principles? (6/7/18)

Google's AI Principles (6/6/18)

Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems (6/5/18)

The State of Machine Learning (ML) Adoption in the Enterprise - O'Reilly Survey (June 2018)

The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems (5/16/18)

Designing ethically pt. 2 (5/4/18)

Designing ethically pt. 1 (5/2/18)

Skill shift: Automation and the future of the workforce (May 2018)

Addressing Age-Related Bias in Sentiment Analysis (4/21/18)

Algorithmic Accountability: A Primer (4/18/18)

Algorithmic impact assessments: A practical framework for public agency accountability (4/9/18)

How to Prevent Discriminatory Outcomes in Machine Learning (3/12/18)

PERSISTENT SURVEILLANCE'S CYNICAL ATTEMPT TO PROFIT OFF BALTIMORE'S TRAUMA (3/1/18)

21 definitions of fairness and their politics [pdf] [video] (2/20/18)

The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation (2/20/18)

Face Off: Law Enforcement Use of Face Recognition Technology (2/12/18)

The problem with building a "fair" system (1/30/18)

Eight Futures of Work: Scenarios and their Implications (January 2018)

News/popular press

The Verge 2018 tech report card: AI (12/30/18)

Are AI toys ethical? (12/27/18)

The Welfare State Is Committing Suicide by Artificial Intelligence (12/25/18)

The case for taking AI seriously as a threat to humanity (12/23/18)

HELLO 2019: 3 Predictions On How AI Will Learn and Impact Society (12/21/18)

The prison-reforming First Step Act has a critical software bug (12/21/18)

'Kill your foster parents': Amazon's Alexa talks murder, sex in AI experiment (12/21/18)

The rise of A.I. could hurt women's careers in a major way (12/21/18)

Amazon error allowed Alexa user to eavesdrop on another home (12/20/18)

AI makers get political (12/19/18)

Google AI Principles updates, six months in (12/18/18)

Super Human or Less Human? (12/16/18)

It's time for a Bill of Data Rights (12/14/18)

A.I. 'bias' could create disastrous results, experts are working out how to fight it (12/14/18)

Nine charts that really bring home just how fast AI is growing (12/12/18)

The invisible workers of the AI era (12/12/18)

Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret (12/10/18)

2018 in Review: 10 AI Failures (12/10/18)

Fixing Bias In Algorithms Is Possible, And This Scientist Is Doing It (12/9/18)

120 predictions for AI for 2019 (12/9/18)

The Seductive Diversion of 'Solving' Bias in Artificial Intelligence (12/8/18)

Bank of America Tech Chief Defines Responsible AI Projects (12/5/18)

Big tech has your kid's data – and you probably gave it to them (12/5/18)

Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias (Dec 2018)

Technologist Vivienne Ming: 'AI is a human right' (12/7/18)

New Research Reveals 75 Percent of Customers Still Favour Live Agent Support for Customer Service vs 25 Percent Self-Service and Chatbots (12/6/18)

Are Silicon Valley's Autoreplies Taking Over Our Minds? (12/6/18)

Facial recognition has to be regulated to protect the public, says AI report (12/6/18)

Facial recognition: It's time for action (12/6/18)

AI has a culturally-biased worldview that Google has a plan to change (12/2/18)

10 Fundamental Insights about the Tech-Driven Future for Humanity and why women, POC, and other underrepresented people in tech should lead it (11/29/18)

Why AI Needs To Reflect Society (11/29/18)

INSTAGRAM WILL USE AI TO DESCRIBE IMAGES FOR VISUALLY IMPAIRED USERS (11/28/18)

10 Most Important AI Research Papers for 2018 (11/27/18)

UK police wants AI to stop violent crime before it happens (11/26/18)

Tech giants offer empty apologies because users can't quit (11/26/18)

THE DANGEROUS JUNK SCIENCE OF VOCAL RISK ASSESSMENT (11/25/18)

How Cheap Labor Drives China's A.I. Ambitions (11/25/18)

Wanted: The 'perfect babysitter.' Must pass AI scan for respect and attitude (11/23/18)

"Siri, is artificial intelligence biased?" (11/20/18)

How to use science fiction to teach tech ethics (11/19/18)

One of the Fathers of AI Is Worried About Its Future (11/19/18)

A new DNA sequencing service wants to reward you for sharing your data (11/15/18)

Google 'betrays patient trust' with DeepMind Health move (11/14/18)

Harvard researchers want to school Congress about AI (11/14/18)

If You Drive in Los Angeles, the Cops Can Track Your Every Move (11/13/18)

Don't Believe Every AI You See (11/13/18)

Algorithmic Bias or Fairness: The Importance of the Economic Context (11/13/18)

In the Age of A.I., Is Seeing Still Believing? (11/12/18)

The newest Jim Crow (11/8/18)

Breaking down AI's trustability challenges (11/8/18)

Google Workers' Walkout Signals Crisis of Faith in Company Culture (11/2/18)

How to Stop Technology From Becoming a Digital Frankenstein (10/31/18)

Should a Self-Driving Car Kill the Baby or the Grandma? Depends on Where You're From (10/29/18)

4 human-caused biases we need to fix for machine learning (10/27/18)

The Digital Gap Between Rich and Poor Kids Is Not What We Expected (10/26/18)

A Dark Consensus About Screens and Kids Begins to Emerge in Silicon Valley (10/26/18)

IBM explores the intersection of AI, ethics–and Pac-Man (10/25/18)

The AI cold war that threatens us all (10/23/18)

No Innovation Without Representation (10/23/18)

Companies are on the hook if their hiring algorithms are biased (10/22/18)

No, A.I. Won't Solve the Fake News Problem (10/20/18)

When AI Misjudgment Is Not an Accident (10/19/18)

An Ethical Awakening in AI (10/18/18)

Weapons of Micro Destruction: How Our 'Likes' Hijacked Democracy (10/17/18)

When Alexa Can't Understand You (10/16/18)

Black Mirror, Light Mirror: Teaching Technology Ethics Through Speculation (10/15/18)

MICROSOFT'S NADELLA SAYS AI CAN MAKE THE WORLD MORE INCLUSIVE (10/15/18)

WIRED25: Ethical AI: Intel's Genevieve Bell On Living with Artificial Intelligence (10/15/18)

A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI (10/11/18)

Building trust in AI applications (10/11/18)

Amazon scraps secret AI recruiting tool that showed bias against women (10/9/18)

WHEN TECH KNOWS YOU BETTER THAN YOU KNOW YOURSELF (10/4/18)

Think computers are less biased than people? Think again. (10/3/18)

Why Technology Favors Tyranny (Oct. 2018)

Build a minimum ethical product (9/30/18)

5 takeaways on the state of AI from Disrupt SF (9/28/18)

Senators introduce the 'Artificial Intelligence in Government Act' (9/26/18)

Taking algorithms to court (9/24/18)

Senators are asking whether artificial intelligence could violate US civil rights laws (9/21/18)

Artificial intelligence hates the poor and disenfranchised (9/21/18)

Artificial Intelligence: The Robots Are Now Hiring (9/20/18)

IBM launches cloud tool to detect AI bias and explain automated decisions (9/19/18)

The exploitation, injustice, and waste powering our AI (9/18/18)

How AI is already affecting jobs (9/17/18)

The Human Promise of the AI Revolution (9/14/18)

High Time to Regulate Face Recognition A.I. (9/13/18)

What Algorithms Know About You Based on Your Grocery Cart (9/13/18)

Google Knows Where You've Been, but Does It Know Who You Are? (9/12/18)

Many voice assistant users have 'trust issues' with their device, study claims (9/12/18)

Ethics + Data Science (9/10/18)

Underneath all the AI hype is the likelihood it threatens the poor, says this former Microsoft and Google exec (9/8/18)

Left Unchecked, Artificial Intelligence Can Become Prejudiced All On Its Own (9/7/18)

Artificial Intelligence is greater concern than climate change or terrorism, says new head of British Science Association (9/6/18)

From headlines to headway: Defining data ethics and its impact on data workers (9/5/18)

IBM collaborated with the NYPD on an AI system that can search for people by race (9/6/18)

Everyday Ethics for Artificial Intelligence (9/5/18)

Don't believe the algorithm (9/5/18)

Silicon Valley Thinks Everyone Feels the Same Six Emotions (9/5/18)

California just replaced cash bail with algorithms (9/4/18)

The New AI Tech Turning Heads in Video Manipulation (9/3/18)

Japan developing 'pre-crime' artificial intelligence to predict money laundering and terror attacks (9/1/18)

How Not to Teach Ethics (Sept 2018)

Franken-algorithms: The Deadly Consequences of Unpredictable Code (8/30/18)

Now is the time to start thinking about AI's impact on xenophobia (8/28/18)

Watson is helping heal America's broken criminal-sentencing system (8/25/18)

To Build Trust In Artificial Intelligence, IBM Wants Developers To Prove Their Algorithms Are Fair (8/22/18)

Technologists are trying to fix the "filter bubble" problem that tech helped create (8/22/18)

Who needs democracy when you have data? (8/20/18)

Bias In Maternal AI Could Hurt Expectant Black Mothers (8/17/18)

AI could make dodgy lip sync dubbing a thing of the past (8/17/18)

AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind (8/16/18)

The Role of Trust in an Era of AI Bots (8/15/18)

Children are susceptible to peer pressure from robots (8/15/18)

Yuval Noah Harari on what the year 2050 has in store for humankind (8/12/18)

Confronting demons of the computer age (8/12/18)

Tech companies use "persuasive design" to get us hooked. Psychologists say it's unethical (8/8/18)

What HBR Gets Wrong About Algorithms and Bias (8/7/18)

Case studies in data ethics (8/7/18)

Why artificial intelligence will have very human frailties (8/3/18)

Using artificial intelligence to fix Wikipedia's gender problem (8/3/18)

Facial Recognition Is the Perfect Tool for Oppression (8/2/18)

Five of the scariest predictions about artificial intelligence (8/1/18)

The U.K. Wants to Become the World Leader in Ethical A.I. (8/1/18)

Fostering a Human-Centered Approach to Artificial Intelligence (7/31/18)

Data's day of reckoning (7/31/18)

The Future of Artificial Intelligence Depends on Trust (7/31/18)

Microsoft's politically correct chatbot is even worse than its racist one (7/31/18)

This VR Founder Wants to Gamify Empathy to Reduce Racial Bias (7/20/18)

The accent gap (7/19/18)

Health Insurers Are Vacuuming Up Details About You – And It Could Raise Your Rates (7/17/18)

Alexa alternatives have a secret weapon: Privacy (7/14/18)

UX for AI: Trust as a Design Challenge (7/12/18)

The Economics Of Artificial Intelligence - How Cheaper Predictions Will Change The World (7/10/18)

'I was shocked it was so easy': meet the professor who says facial recognition can tell if you're gay (1/7/18)

London police chief 'completely comfortable' using facial recognition with 98 percent error rate (7/5/18)

4 ways AI can change democracy for the better (7/4/18)

Technology Alone Can't Preserve Endangered Languages (6/30/18)

BIAS EVERYWHERE: Tech companies just woke up to a big problem with their AI (6/30/18)

AI, Ain't I A Woman? (6/28/18)

When impact measures fail (And what to do about it) (6/26/18)

Techstars: AI startups must be wary of 'move fast and break things' mantra (6/26/18)

Facial recognition software is not ready for use by law enforcement (6/25/18)

Unmasking A.I.'s Bias Problem (6/25/18)

Bias detectives: the researchers striving to make algorithms fair (6/20/18)

Bing researchers develop a novel way of collecting high-quality AI training data (6/18/18)

UK report warns DeepMind Health could gain 'excessive monopoly power' (6/15/18)

A Tough Week for Tech Workers, and It Won't Be the Last (6/15/18)

Plum uses AI to hire people 'that never would have been discovered through a traditional hiring process' (6/13/18)

When it Comes to AI and Weapons, the Tech World Needs Philosophers (6/12/18)

This tool lets you see–and correct–the bias in an algorithm (6/12/18)

Tackling the Ethical Challenges of Slippery Technology (6/11/18)

Accenture wants to beat unfair AI with a professional toolkit (6/9/18)

Silicon Valley insiders say Facebook, Snapchat, and Twitter are using 'behavioural cocaine' to turn people into addicts (6/5/18)

The digital poorhouse (6/4/18)

Are you scared yet? Meet Norman, the psychopathic AI (6/2/18)

Artificial Intelligence: Have No Fear (June 2018)

China has turned Xinjiang into a police state like no other (5/31/18)

Does China's digital police state have echoes in the West? (5/31/18)

Pymetrics open-sources Audit AI, an algorithm bias detection tool (5/31/18)

Machine Un-Learning: Why Forgetting Might Be the Key to AI (5/31/18)

Who is writing the future? Designing infrastructure for ethical AI (5/31/18)

The (holy) ghost in the machine: Catholic thinkers tackle the ethics of artificial intelligence (5/26/18)

Navigating the risks of artificial intelligence and machine learning in low income countries (5/24/18)

Enhanced AI Tools, Enhanced AI Trust (5/24/18)

How Advances in AI and Automation Will Upend Our Traditional Models of Economic Development (5/20/18)

Finkle floats an ethical AI plan (5/18/18)

Artificial Intelligence: What's Human Rights Got To Do With It? (5/14/18)

Algorithms are making the same mistakes assessing credit scores that humans did a century ago (5/14/18)

Society needs a reboot for the Fourth Industrial Revolution (5/14/18)

Aided by Palantir, the LAPD used predictive policing to monitor specific people and neighborhoods (5/11/18)

This company audits algorithms to see how biased they are (5/9/18)

Facebook says it has a tool to detect bias in its artificial intelligence (5/3/18)

Data Violence and How Bad Engineering Choices Can Damage Society (4/30/18)

Revealed: how bookies use AI to keep gamblers hooked (4/30/18)

PwC: Lack of trust in AI assistants like Alexa could hinder adoption (4/29/18)

It's time to address the reproducibility crisis in AI (4/24/18)

What you need to know about Facebook and ethics (4/19/18)

Artificial intelligence must be 'for common good' (4/16/18)

When algorithms surprise us (4/13/18)

Microsoft has given up 'significant sales' over concerns that the customer will use AI for evil, says a top scientist (4/13/18)

Democracy vs. the Algorithm (4/12/18)

Human bias is a huge problem for AI. Here's how we're going to fix it (4/10/18)

Drawing the Ethical Line on Weaponized Deep Learning Research (4/5/18)

Leading AI researchers threaten Korean university with boycott over its work on 'killer robots' (4/4/18)

7 Short-Term AI ethics questions (4/4/18)

A.I. Engineers Must Open Their Designs To Democratic Control (4/2/18)

What happens when an algorithm cuts your health care (3/21/18)

OpenAI Wants to Make Safe AI, but That May Be an Impossible Task (3/15/18)

What Developers Really Think About AI And Bias (3/15/18)

Fun fact of the day: Voice recognition tech is naturally sexist (3/14/18)

The case for fairer algorithms (3/14/18)

Google's DeepMind Has An Idea For Stopping Biased AI (3/13/18)

How to Make A.I. That's Good for People (3/7/18)

The ART of AI–Accountability, Responsibility, Transparency (3/4/18)

Is Artificial Intelligence the Ultimate Test for Privacy? (3/2/18)

New Report on Emerging AI Risks Paints a Grim Future (2/21/18)

Interpreting machine learning models (2/20/18)

"We're in a diversity crisis": cofounder of Black in AI on what's poisoning algorithms in our lives (2/14/18)

Google's DeepMind Has An Idea For Stopping Biased AI (2/13/18)

He Predicted The 2016 Fake News Crisis. Now He's Worried About An Information Apocalypse. (2/11/18)

Facial Recognition Is Accurate, if You're a White Guy (2/9/18)

SHOULD DATA SCIENTISTS ADHERE TO A HIPPOCRATIC OATH? (2/8/18)

Dealing with Imbalanced Classes in Machine Learning (2/2/18)

WHY DO WE KEEP GENDERING OUR AI ASSISTANTS? (1/26/18)

Algorithms are making American inequality worse (1/26/18)

Sorry, Alexa Is Not a Feminist (1/24/18)

A Popular Algorithm Is No Better at Predicting Crimes Than Random People (1/17/18)

A CHILD ABUSE PREDICTION MODEL FAILS POOR FAMILIES (1/15/18)

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech (1/12/18)

Do algorithms reveal sexual orientation or just expose our stereotypes? (1/11/18)

Why bots go bad: Curbing transgressive tendencies in AI (1/3/18)

Can an Algorithm Tell When Kids Are in Danger? (1/2/18)

The Impact of Machine Learning on Economics (Jan 2018)

Artificial Intelligence Trends to Watch in 2018 - CB Insights

2017

Peer-reviewed

Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences (12/2/17)

Are Algorithms Building the New Infrastructure of Racism? How we use big data can reinforce our worst biases–or help fix them (12/2/17)

Avoiding discrimination through causal reasoning (12/2017)

Ethical issues in research using datasets of illicit origin (Nov 2017)

The real risks of artificial intelligence (2017)

Distill-and-Compare: Auditing Black-Box Models Using Transparent
Model Distillation (10/17/17)

Is that social bot behaving unethically? (Sept 2017)

Toward algorithmic transparency and accountability (Sept 2017)

Algorithmic decision making and the cost of fairness (8/13/17)

Human decisions and machine predictions (2017)

Men Also Like Shopping: Reducing Gender Bias Amplification using
Corpus-level Constraints (7/29/17)

Decoupled classifiers for fair and efficient machine learning (2017)

Algorithmic decision making and the cost of fairness (6/10/17)

Responses to Critiques on Machine Learning of Criminality Perceptions
(5/26/17)

Semantics derived automatically from language corpora necessarily
contain human biases (5/25/17)

Evaluating Quality of Chatbots and Intelligent Conversational Agents
(4/15/17)

Ethical by Design: Ethics Best Practices for Natural Language
Processing (4/4/17)

Goal-Oriented Design for Ethical Machine Learning and NLP (4/4/17)

Social Bias in Elicited Natural Language Inferences (April 2017)

Racial Discrimination in the Sharing Economy: Evidence from a Field
Experiment. (April 2017)

Decision Making With Quantized Priors Leads to Discrimination (Feb
2017)

Identifying Unknown Unknowns in the Open World: Representations
and Policies for Guided Exploration (2017)

Refractive Surveillance: Monitoring Customers to Manage Workers
(submitted 2/8/17)

Government, NGO, and expert publications

IEEE Standards Association Ethically Aligned Design guidelines v2
(12/2017)

Montreal Responsible AI Declaration (11/17)

WHAT'S NEXT FOR MACHINE LEARNING: ETHICS AND
EXPLAINABILITY IN AI FOR FRAUD (11/2017)

AI Now 2017 Report (10/2017)

Bots Won't Just Help Us Buy Stuff. They'll Help Us Become Better
Versions of Ourselves (6/1/17)

Physiognomy's New Clothes (5/6/17)

When Good Intentions Backfire ...And why we need a hacker mindset (2/15/17)

News/popular press

ATIH: Natalie Garrett on Ethical Debt and Ethics Education in Tech (3/12/20)

Can we teach morality to machines? Three perspectives on ethics for artificial intelligence (12/19/17)

Inside China's vast new experiment in social ranking (12/14/17)

New York City's Bold, Flawed Attempt to Make Algorithms Accountable (12/12/17)

The trouble with bias (12/10/17)

ARTIFICIAL INTELLIGENCE SEEKS AN ETHICAL CONSCIENCE (12/7/17)

Researchers Combat Gender and Racial Bias in Artificial Intelligence (12/4/17)

Global AI Dialogue Series: Observations from the China-US Workshop in Beijing (12/2/2017)

Ethics at scale (11/29/17)

Google Translate's gender bias pairs "he" with "hardworking" and "she" with lazy, and other examples (11/29/17)

Can A.I. Be Taught to Explain Itself? (11/21/17)

This Chatbot Is Trying Hard To Look And Feel Like Us (11/15/17)

Trustworthy AI? Yes, by design (11/8/17)

Who's the fairest of them all? Not AI (11/8/17)

When Data Science Destabilizes Democracy and Facilitates Genocide (11/2/17)

The 10 Top Recommendations for the AI Field in 2017 (10/18/17)

What Silicon Valley gets wrong about universal basic income (10/16/17)

Can we teach robots ethics? (10/15/17)

Asking the right questions about AI (10/11/17)

Forget killer robots - Bias is the real AI danger (10/3/17)

Artificial Empathy Systems and Design Research at Scale (10/2/17)

Artificial intelligence is about the people, not the machines (9/30/17)

How to recognize exclusion in AI (9/26/17)

Instagram uses 'I will rape you' post as Facebook ad in latest algorithm mishap (9/21/17)

Artificial intelligence can make our societies more equal. Here's how (9/21/17)

Training soft skills into AI technology (9/19/17)

AI research is in desperate need of an ethical watchdog (9/18/17)

New AI can guess whether you're gay or straight from a photograph
(9/7/17)

How to regulate artificial intelligence (9/1/17)

Turns out Algorithms are racist (8/31/17)

The IRS Is Mining Taxpayer Data On Social Media In Violation Of
Federal Privacy Law (8/28/17)

New app scans your face and tells companies whether you're worth
hiring (8/25/17)

Machines taught by photos learn a biased view of women (8/21/17)

AI's biggest challenge is human, not technological. (8/19/17)

AI programs are learning to exclude African-American voices
(8/16/17)

Teaching AI systems to behave themselves (8/13/17)

Why we desperately need women to design AI (8/4/17)

Biased AI is a threat to civil liberties. The ACLU has a plan to fix it.
(7/25/17)

Technology Is Biased Too. How Do We Fix It? (7/20/17)

Why Google's PAIR initiative to take bias out of AI will never be complete (7/18/17)

AI and the future of ethics (7/1/17)

Volvo admits its self-driving cars are confused by kangaroos (6/30/17)

When a Computer Program Keeps You in Jail (6/13/17)

To live in harmony with AI we must create a modern Magna Carta (7/6/17)

Corporate Surveillance in Everyday Life (June 2017)

A discussion about AI's conflicts and challenges (6/17/17)

Elon Musk and linguists say that AI is forcing us to confront the limits of human language (6/14/19)

Facebook taught bots to negotiate (and lie) like humans (6/14/17)

This backflipping noodle has a lot to teach us about AI safety (6/14/17)

Don't Grade Teachers With a Bad Algorithm (5/15/17)

The dark secrets at the heart of AI (4/11/17)

Will Using Artificial Intelligence To Make Loans Trade One Kind Of Bias For Another? (3/31/17)

We need to talk about Accessibility on Chatbots (1/30/17)

Algorithms and bias: What lenders need to know (1/20/17)

2016

Peer Reviewed

Policing Predictive Policing (2016)

Man is to computer programmer as woman is to homemaker? Debiasing word embeddings (2016)

Equality of Opportunity in Supervised Learning (10/7/2016)

The Other Question: Socialbots and the Question of Ethics (12/1/2016)

Designing AI systems that obey our laws and values (2016)

Generating Visual Explanations (Oct 2016)

On the (im)possibility of fairness (9/23/16)

The Ethics of Algorithms: Mapping the Debate (July 2016)

"Why Should I Trust You?": Explaining the Predictions of Any Classifier (2/16/18)

A Confidence-Based Approach for Balancing Fairness and Accuracy (1/22/16)

Government, NGO, and expert publications

A Guide to Solving Social Problems with Machine Learning (12/8/16)

Racism and inequity are products of design. They can be redesigned.
(11/15/16)

Big Data's Disparate Impact (2016)

News/popular press

Facebook Privacy: Social Network Buys Data From Third-Party Brokers
To Fill In User Profiles (12/28/16)

Top 10 AI failures of 2016 (12/2/2016)

Facebook Lets Advertisers Exclude Users by Race (10/28/16)

The ethics of good design: A principle for the connected age
(11/20/16)

The perpetual line-up (10/18/16)

There is a blind spot in AI (10/13/16)

A beauty contest was judged by AI and the robots didn't like dark skin
(9/8/2016)

How algorithms rule our working lives (9/1/16)

Chicago's predictive policing tool just failed a major test (8/19/16)

Artificial Intelligence Has a 'Sea of Dudes' Problem (6/23/16)

Machine Bias (5/23/2016)

Is Amazon same-day delivery service racist? (4/23/16)

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day (3/24/2016)

Uber seems to offer better service in areas with more white people (3/10/16)

2015

Race effects on eBay (Oct 2015)

Women less likely to be shown ads for high-paid jobs on Google, study shows (7/28/15)

Questioning the Fairness of Targeting Ads Online (7/7/15)

Automated Experiments on Ad Privacy Settings (4/18/15)

What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. (2015)

2014

Social attributions from faces bias human choices (Nov 2014)

How big data is unfair (9/26/14)

How to Build a Strong Ethical Culture at Your [Government] Agency (7/14/14)

2013

Data brokers selling lists of rape victims, AIDS patients (12/19/13)

Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language (9/11/13)

Linguistic Models for Analyzing and Detecting Biased Language (August 2013)

Discrimination in Online Ad Delivery (1/28/13)

2012

Fairness through awareness (Jan 2012)

2011

A Framework for Ethical Reasoning (12/2/11)

Testing a Bayesian Measure of Representativeness Using a Large Image Database (2011)

2010

Perceptual cues in non-verbal vocal expressions of emotion (4/29/10)

2009

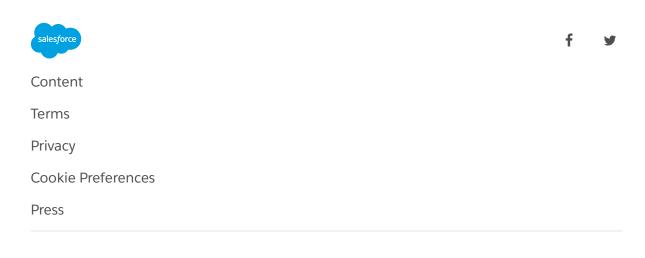HP Investigates Claims of 'Racist' Computers (12/22/09)

Camera Misses the Mark on Racial Sensitivity (5/19/09)

2008

Understanding Information Ethics and Policy: Integrating Ethical Reflection and Critical Thinking into Policy Development (2008)

2004

[Ethical Fading: The Role of Self-Deception in Unethical Behavior](#)
(7/2004)

Content

Terms

Privacy

Cookie Preferences

Press