

INTELIGÊNCIA ARTIFICIAL RESPONSÁVEL

AUDIÊNCIA PÚBLICA | SENADO - CTIA

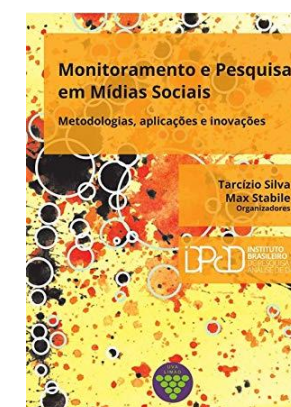
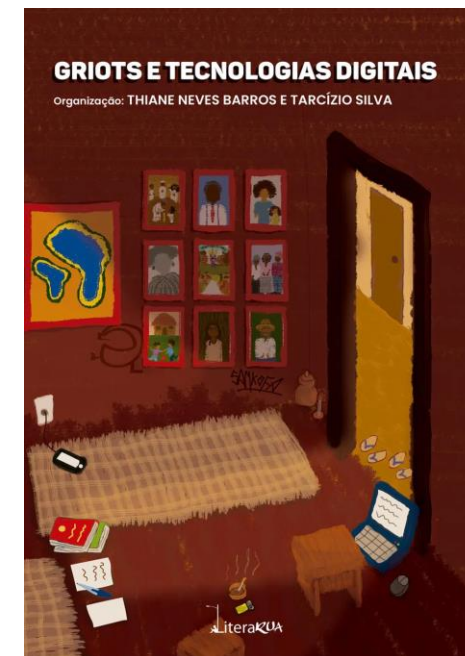
Tarcízio Silva

TARCIZIO SILVA

moz://a

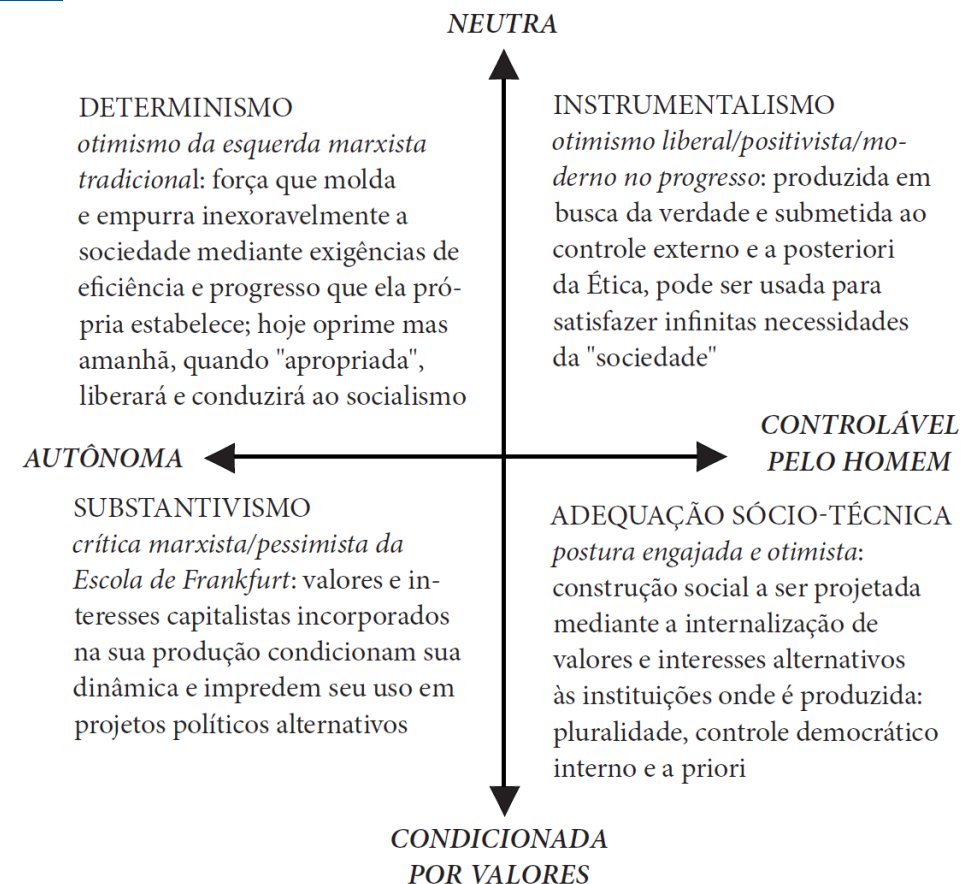


- Tech Policy Fellow | Mozilla Foundation
- Doutorando em Ciências Humanas e Sociais | UFABC
- Colaborador Tech + Society | Tecla
- Pesquisador Desvelar
- Pesquisador Sumaúma
- Membro do Grupo de Pesquisa em Comunicação Antirracista e Afrodiaspórica | Intercom
- Membro da Coalizão Direitos na Rede



NORMATIVIDADE, FUNDAMENTOS E DIRETRIZES CONTRA DISCRIMINAÇÃO NEGATIVA

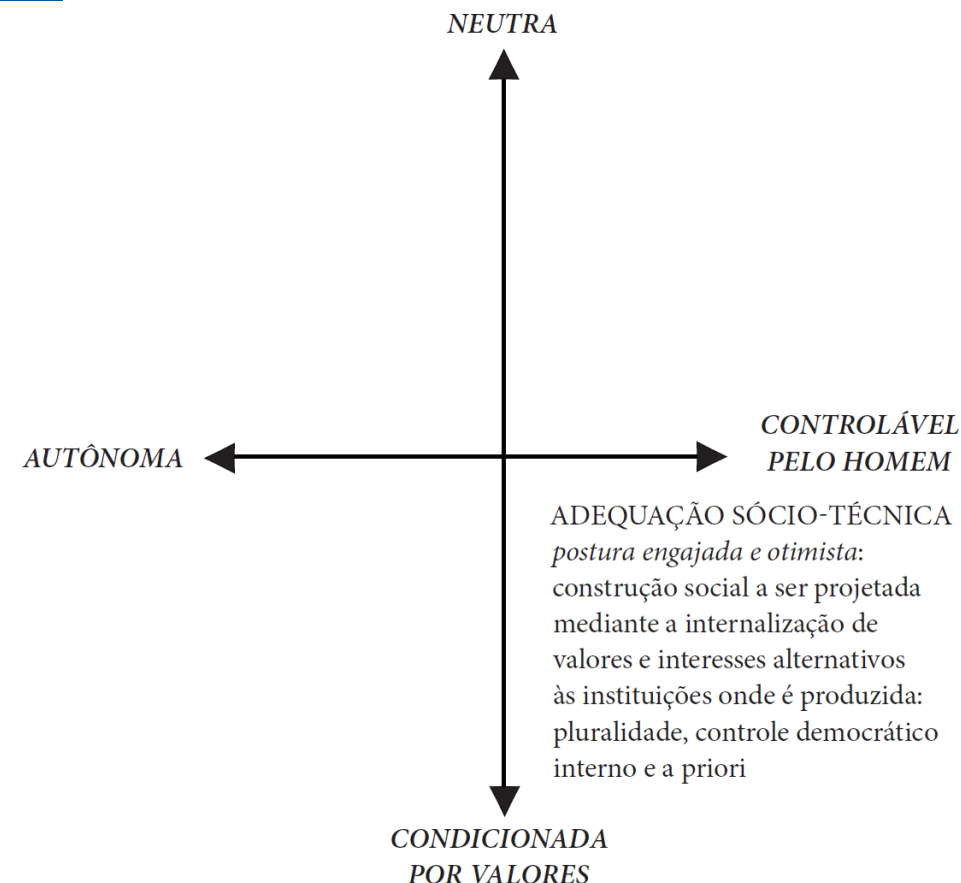
- Não existe **neutralidade** na tecnologia ou qualquer implementação, fundamento importante para a busca de **adequação sociotécnica**



[Esquema produzido por Renato Dagnino a partir das ideias de Andrew Feenberg]

NORMATIVIDADE, FUNDAMENTOS E DIRETRIZES CONTRA DISCRIMINAÇÃO NEGATIVA

- Não existe **neutralidade** na tecnologia ou qualquer implementação, fundamento importante para a busca de **adequação sociotécnica**
- Principialismo x Pragmatismo: realidade material das desigualdades interseccionais e registro do papel da IA em suas intensificações político-econômicas antes que éticas
- Inclusão explícita de mecanismos discriminações negativas interseccionais



[Esquema produzido por Renato Dagnino a partir das ideias de Andrew Feenberg]

O QUE É RACISMO ALGORÍTMICO

“modo pelo qual a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados. Tal ordenação pode ser vista como uma camada adicional do racismo estrutural, que vai além e molda o futuro e horizontes de relações de poder adicionando mais opacidade sobre exploração e opressão global racializada desde o projeto colonial do século XVI”

(SILVA, 2022, p.120)

MATURIDADE SOBRE PROBLEMAS REGISTRADOS

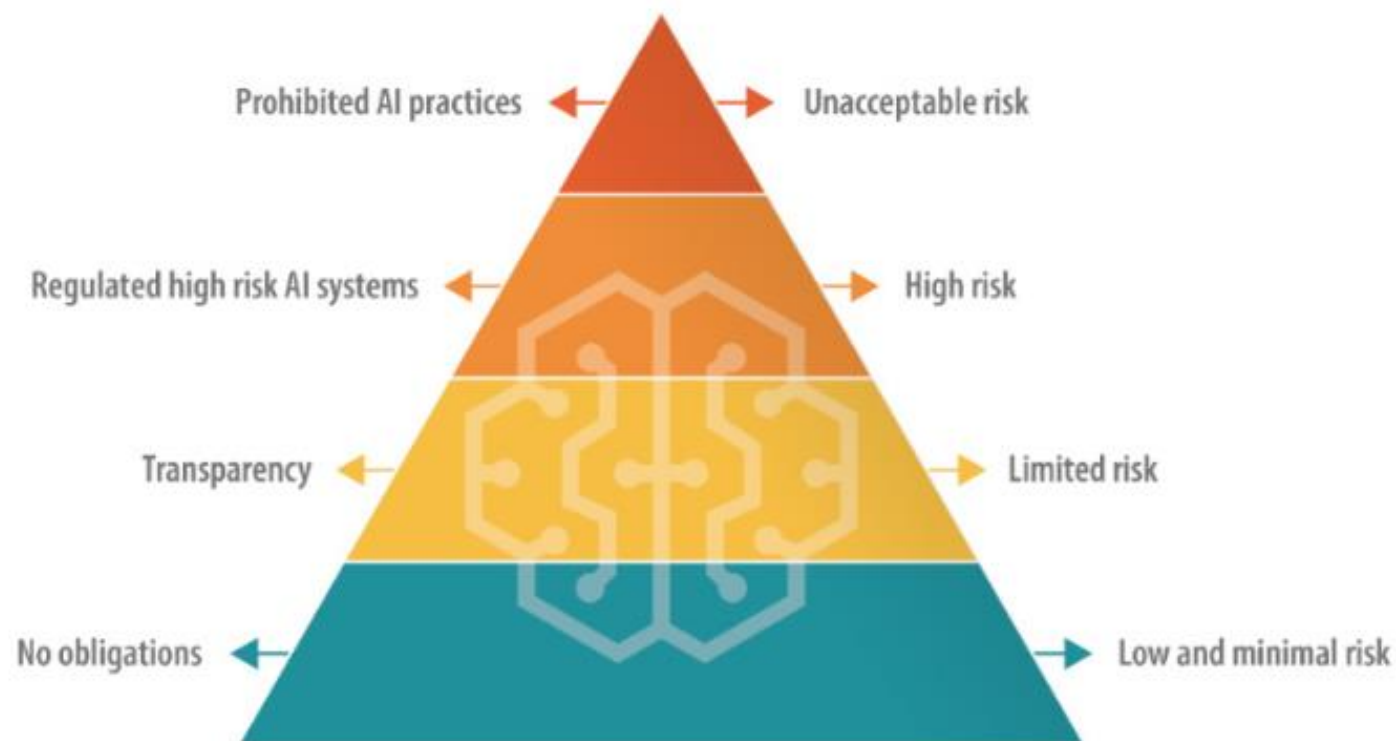


- Danos individuais e coletivos de tecnologias digitais emergentes são registrados de forma acelerada nos mais diversos campos da sociedade
- Já temos acúmulo de conhecimento sobre quais tipos de implementações oferecem danos e/ou riscos inaceitáveis

The screenshot shows the AI Incident Database (AIID) website. The header includes the AIID logo, the text 'AI INCIDENT DATABASE', a language selector set to 'English', social media icons, and a 'Subscribe' button. The main content area is titled 'Welcome to the AI Incident Database' and features a search bar with the placeholder text 'Search over 2000 reports of AI harms'. Below the search bar are 'Search' and 'Discover' buttons. A sidebar on the left contains navigation options: 'Discover', 'Submit', 'Discover Incidents', 'Spatial View', 'Table View', 'Entities', 'Taxonomies', 'Word Counts', 'Submit Incident Reports', and 'Submission'. The main content area displays a news article titled 'A man was encouraged by a chatbot to kill Queen Elizabeth II in 2021. He was sentenced to 9 years' with a 'Latest Incident Report' link. The article text mentions a Star Wars fanatic who was encouraged by a chatbot 'girlfriend' to assassinate Queen Elizabeth II and was sentenced to nine years in prison for taking his plot to Windsor Castle.

BRASIL NÃO É EUROPA: PRECISAMOS DE MAIS DEFESA DE DIREITOS

- A regulação no Brasil deve ser mais rigorosa do que a referência da U.E.
- Pobreza, desigualdades, violência policial e fragilidades do sistema criminal e judiciário devem ser motivo de mais cautela e proteção de direitos
- Compreensão sistêmica dos impactos econômicos e sociais da posição do Brasil frente a empresas estrangeiras focadas em I.A.



COMPROMISSOS DO ESTADO BRASILEIRO E IMPACTOS NA I.A. RESPONSÁVEL

dever do Estado e da sociedade **garantir a igualdade de oportunidades**, reconhecendo a todo cidadão brasileiro, independentemente da etnia ou da cor da pele, **o direito à participação na comunidade**, especialmente nas atividades políticas, econômicas, empresariais, educacionais, culturais e esportivas, defendendo sua dignidade e seus valores religiosos e culturais

Lei 12.288/2010

realizar **pesquisas sobre a natureza, as causas e as manifestações** do racismo, da discriminação racial e formas correlatas de intolerância; coletar, compilar e divulgar dados sobre a situação de grupos ou indivíduos que sejam vítimas do racismo, da discriminação racial e formas correlatas de intolerância

Convenção Interamericana contra o Racismo, a Discriminação Racial e Formas Correlatas de Intolerância

COMPROMISSOS DO ESTADO BRASILEIRO E IMPACTOS NA I.A. RESPONSÁVEL

dever do Estado e da sociedade **garantir a igualdade de oportunidades**, reconhecendo a todo cidadão brasileiro, independentemente da etnia ou da cor da pele, **o direito à participação na comunidade**, especialmente nas atividades políticas, econômicas, empresariais, educacionais, culturais e esportivas, defendendo sua dignidade e seus valores religiosos e culturais

Lei 12.288/2010

realizar **pesquisas sobre a natureza, as causas e as manifestações** do racismo, da discriminação racial e formas correlatas de intolerância; coletar, compilar e divulgar dados sobre a situação de grupos ou indivíduos que sejam vítimas do racismo, da discriminação racial e formas correlatas de intolerância

Convenção Interamericana contra o Racismo, a Discriminação Racial e Formas Correlatas de Intolerância



tornar **avaliações de impactos em direitos humanos, igualdade racial e não-discriminação um pré-requisito** para a adoção de sistemas baseados em tais tecnologias por autoridades públicas. Estas avaliações de impacto devem incorporar oportunidades significativas para **co-design e co-implementação com representantes de grupos étnico-raciais marginalizados**

estados devem garantir **transparência e prestação de contas** sobre o uso de tecnologias digitais emergentes pelo setor público e permitir análise e supervisão independente, incluindo através do **uso apenas de sistemas que sejam auditáveis**

A/HRC/44/57 – United Nations

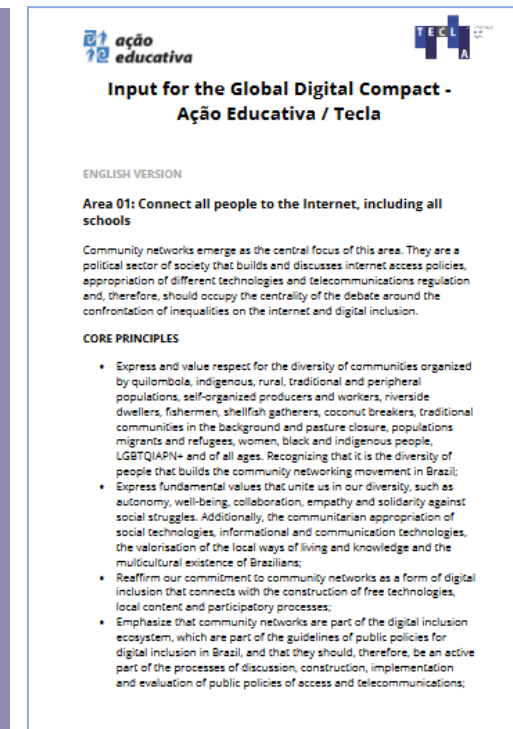
RECOMENDAÇÕES

PL 2338/2023: manter os avanços e melhorar pontos necessários

- Pontos positivos como abordagem baseada em balanceamento de framework de riscos, inovação e direitos;
- Aperfeiçoar mecanismos de direitos de informação aos sujeitos afetados;
- Avançar o debate sobre mecanismos de avaliação de gastos energéticos e outros impactos ambientais;
- Avançar itens de transparência sobre emprego de trabalhadores nas funções de treinamento, modelagem, produção de conteúdo e mais.



Nota técnica sobre o PL 2338/2023, desenvolvida pela Coalizão Direitos na Rede

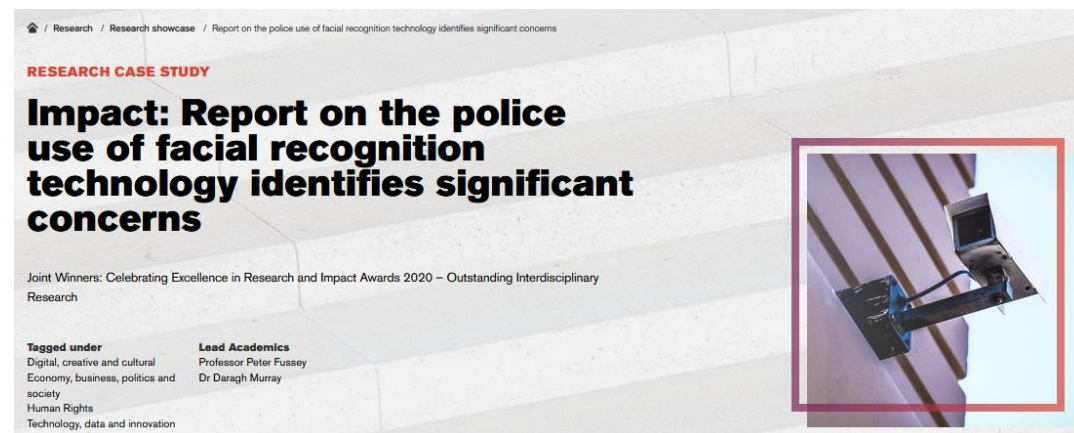


Submissão da Tecla / Ação Educativa e colaboradores à consulta da Nações Unidas

RECOMENDAÇÕES

Inclusão em implementações de risco inaceitável:

- Reconhecimento facial e identificação biométrica no espaço público e/ou em escala
- Reconhecimento de emoções, fisionomia e frenologia
- Policiamento preditivo e práticas de “*hot spotting*”
- Manipulação psicométrica e modelos comportamentais para marketing
- Precificação e outros tipos de discriminação em serviços baseadas em estados emocionais
- Sistemas de rastreamento e modelagem preditiva de comportamento de trabalhadores



RECOMENDAÇÕES

Transparência e relatórios de avaliação de impacto algorítmico

- Perspectiva de justificação prévia – sistemas de alto risco e alta escala devem justificar não-infração de direitos e controle de danos **antes** de implementados
- **Compreensão sistêmica** das tecnologias algorítmicas, incluindo não só inputs, outputs e modelos, mas também contextos, objetivos, implementações, práticas dos setores e grupos afetados

Composição da(s) entidade(s) supervisora(s):

- **Participação ampla da sociedade civil** a partir da inclusão em espaços de tomada de decisão de grupos representativos de populações vulnerabilizadas
- **Experts de domínio** e setores: academia, sociedade civil organizada e sindicatos em entidade com autonomia funcional, decisória, administrativa e financeira
- Papel do Estado na **promoção de boas práticas** ao exigir níveis óptimos de transparência, escuta pública e mecanismos de revisão

REFERÊNCIAS
BIBLIOGRÁFICAS
SELECIONADAS

BIBLIOTECA SELECIONADA SOBRE INTELIGÊNCIA ARTIFICIAL, SISTEMAS ALGORÍTMICOS E JUSTIÇA

- [*Nota Técnica sobre Projeto de Lei nº 2338/2023*](#), por Coalizão Direitos na Rede
- [*Racismo Algorítmico: inteligência artificial e discriminação nas redes digitais*](#), por Tarcizio Silva
- [*Reconhecimento facial e segurança pública: os perigos do uso da tecnologia no sistema penal seletivo brasileiro*](#), por Rosane Silva e Fernanda Rodrigues
- [*Inteligência artificial e discriminação: Desafios e perspectivas para a proteção de grupos vulneráveis frente às tecnologias de reconhecimento facial*](#), por Ramon Costa e Bianca Kremer
- [*Tecnologia de Reconhecimento Facial e Segurança Pública nas Capitais Brasileiras*](#), por Paulo Victor Melo e Paulo Serra
- [*Estudos sobre a vigilância: do panóptico ao big-other*](#), por Ana Viana e Thiago Marcilio
- [*O Slave Ship como paradigma de análise dos aparatos de vigilância no genocídio antinegro pelo Estado Brasileiro*](#), por Elizandra Salomão e Pedro Monteiro
- [*Prever crimes a que custo?*](#), por Pablo Nunes
- [*Reconhecimento facial no setor público e identidades trans*](#), por Mariah Rafaela Silva e Joana Varon
- [*Todos têm alguma luz sob o sol: luminosidade negra e a cultura visual de vigilância*](#), por Simone Browne
- [*O lobo reluzente e high-tech em pele de cordeiro*](#), por Ruha Benjamin
- [*Lei Geral de Proteção de Dados e os possíveis impactos da não obrigatoriedade de revisão humana de decisões automatizadas*](#), por Natane Santos
- [*Colaboração escrita do grupo Juristas Negras à CJSUBIA*](#), por Juristas Negras
- [*Colaboração escrita do grupo Mulheres na Privacidade à CJSUBIA*](#), por Mulheres na Privacidade
- [*Colaboração escrita do Instituto de Referência em Internet e Sociedade à CJSUBIA*](#), por IRIS

BIBLIOTECA SELECIONADA SOBRE INTELIGÊNCIA ARTIFICIAL, SISTEMAS ALGORÍTMICOS E JUSTIÇA

- [*Racial discrimination and emerging digital technologies : a human rights analysis : report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*](#), por Dra. Tendayi Achiume
- [*Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*](#), por University of Essex
- [*The grey-zones of public-private surveillance: Policy tendencies of facial recognition for public security in Brazilian cities*](#), por André Ramiro e Luã Cruz
- [*Creating Trustworthy AI*](#), por Mozilla Foundation
- [*Data Capitalism and Algorithmic Racism*](#), por Data4BlackLives & Demos
- [*On the dangers of stochastic parrots: Can language models be too big?*](#), por Emily Bender, Timnit Gebru, Angelina McMillan-Major e Margaret Mitchell
- [*How We Analyzed the COMPAS Recidivism Algorithm*](#), por Jeff Larson e colaboradores
- [*Advancing Racial Equity Through Technology Policy*](#), por AI Now Institute
- [*Assessing Risk, Automating Racism*](#), por Ruha Benjamin
- [*Robot Rights? Let's Talk about Human Welfare Instead*](#), por Abeba Birhane e Jelle van Dijk
- [*Persistent anti-muslim bias in large language models*](#), por Abubakar Abid, Maheen Farooqi e James Zou
- [*Against Scale: Provocations and Resistances to Scale Thinking*](#), por Alex Hanna, Tina M. Park
- [*Gender shades: Intersectional accuracy disparities in commercial gender classification*](#), por Joy Buolamwini e Timnit Gebru
- [*Guilty until proven guilty: policing caste through preventive policing registers in India*](#), por Shivangi Narayan
- [*Cities Should Act NOW to Ban Predictive Policing...and Stop Using ShotSpotter, Too*](#), por Electronic Frontier Foundation
- [*Ban Biometric Surveillance*](#), por Accessnow e centenas de organizações