UNIVERSITÉ DE FRIBOURG FACULTÉ DE DROIT
UNIVERSITÄT FREIBURG RECHTSWISSENSCHAFTLICHE FAKULTÄT

Institut für Europarecht
Institut de droit européen

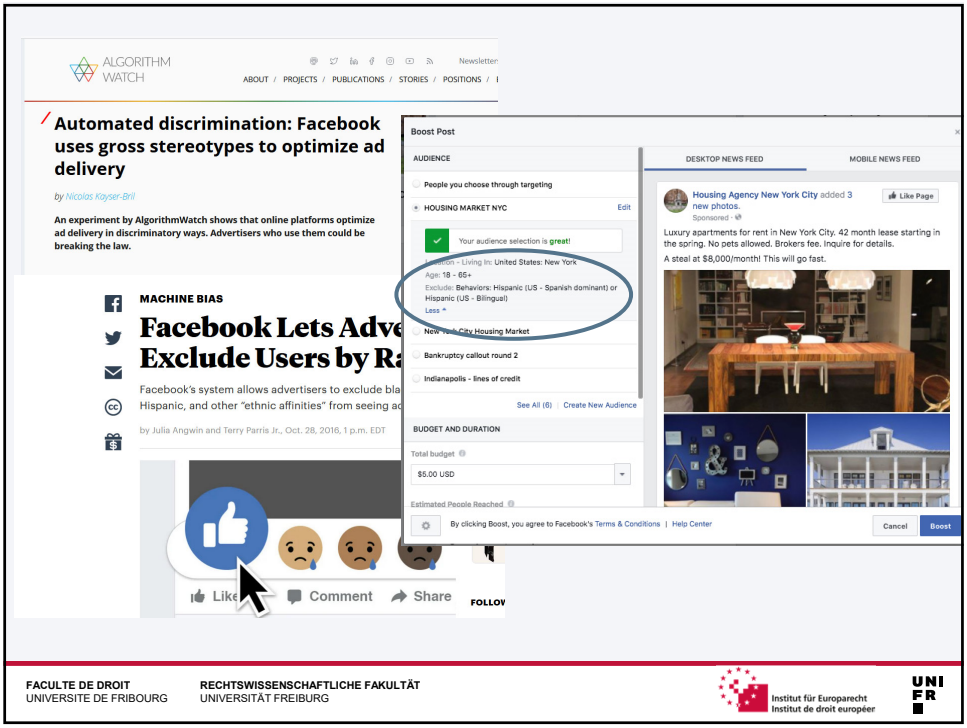
15. Dezember 2022

DISKRIMINIERUNG IN UND DURCH *SOCIAL MEDIA*

NEUE HERAUSFORDERUNGEN FÜR DAS GLEICHHEITSRECHT?

ICJ-CH Vortragsreihe «Soziale Medien und Menschenrechte», Universität Bern

Dr. Nula Frei
Institut für Europarecht, Universität Freiburg i.Ue.



FACULTE DE DROIT
UNIVERSITE DE FRIBOURG

RECHTSWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT FREIBURG

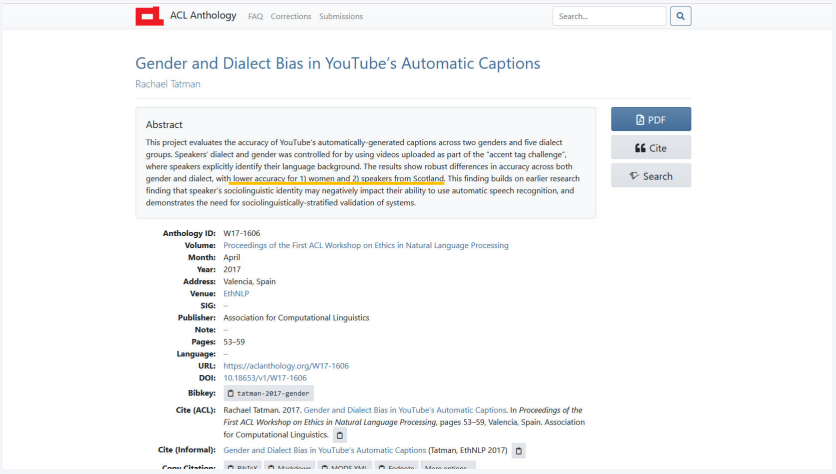
Institut für Europarecht
Institut de droit européen

Hackathon Points to More Biases in Twitter Algorithm

At Def Con 29, held online and in-person in Las Vegas in August, hackathon participants uncovered how a Twitter algorithm was coded with implicit bias against older people, those with disabilities, and Muslims, in addition to Black people.

Author: Curt Wagner

The second place winner of an algorithmic bias bounty contest sponsored by Twitter showed how Twitter's image-cropping algorithm cropped out people with white or grey hair. (Courtesy of Twitter)



DISKRIMINIERUNG *DURCH (?)* SOZIALE MEDIEN: RECHTLICHE FRAGESTELLUNGEN

1. DISKRIMINIERUNG IM JURISTISCHEN SINN?
   - Benachteiligung?
   - Urheber:in?
   - Nachweis?
   - Abhilfemassnahmen?
2. STAATLICHER HANDLUNGSBEDARF?
   - Schutzpflichten?
   - Prävention, Repression?

https://en.wikipedia.org/wiki/Online_gender-based_violence#/media/File:Online_harassment_of_women_journalists.png

FACULTE DE DROIT
UNIVERSITE DE FRIBOURG
RECHTSWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT FREIBURG
Institut für Europarecht
Institut de droit européen
UNI FR

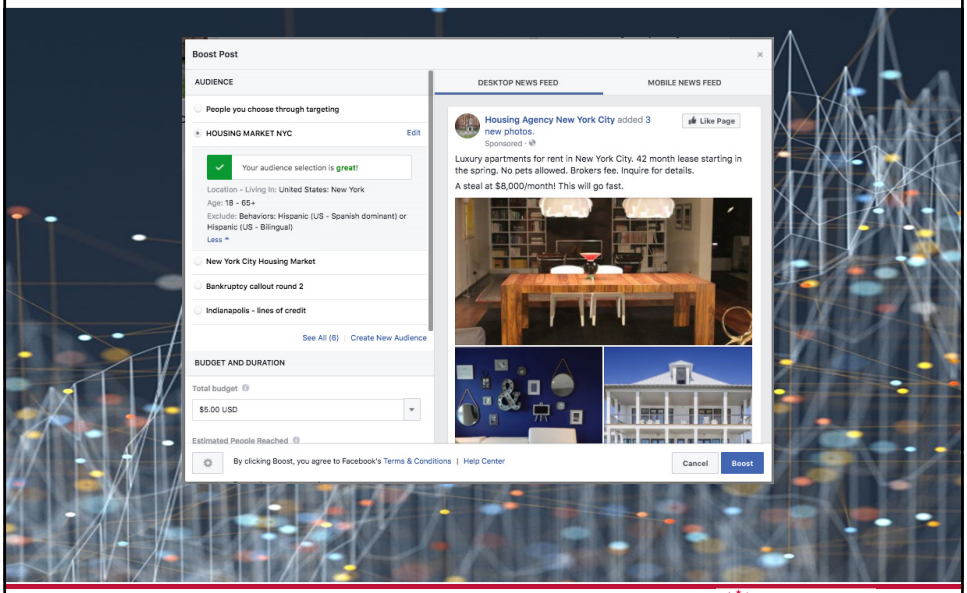
DISKRIMINIERUNG *IN (?)* SOZIALEN MEDIEN: RECHTLICHE FRAGESTELLUNGEN

1. DIGITALE GEWALT ALS FORM DER DISKRIMINIERUNG?
2. MENSCHENRECHTLICHE SCHUTZPFLICHTEN?
3. ABHILFEMASSAHMEN?

THEMENKREISE

1. **TARGETED ADVERTISING** MIT DISKRIMINIERENDER ABSICHT UND/ODER WIRKUNG
2. **FALSCHERKENNUNG** DISKRIMINIERUNGSRECHTLICH GESCHÜTZTER GRUPPEN ODER MERKMALE
3. **CYBERGEWALT** ALS FORM DER DISKRIMINERUNG

1. THEMENKREIS «TARGETED ADVERTISING»



variation	Facebook: % female impressions	Google: % female impressions	image
Image of **cosmetics**	88%	47%	
Image of a **road**	53%	46%	
Text of the ad in **gendered** form	22%	49%	
Text of the ad in **feminine** form	19%	51%	
Baseline	15%	50%	

Based on 11,563 ad impressions in France on 3 Sep.
Table: AlgorithmWatch • Get the data • Created with Datawrapper

DISKRIMINIERUNG IM JURISTISCHEN SINN?

- Unmöglichkeit, Stelleninserate zu sehen, als *Benachteiligung;* zusätzlich auch gesellschaftlich unerwünschte, strukturelle und Langzeiteffekte; Stereotypisierung
- *Urheber:in* der Diskriminierung: Soziales Medium? Werbekund:in? Algorithmus?
- Anknüpfung an ein *Diskriminierungsmerkmal* bei Ausschluss geschützter Gruppen / bei Verwendung von Proxy-Variablen
- *Rechtfertigung* der Ungleichbehandlung: zielgruppengenaue Angebotsunterbreitung als legitimes Interesse?

RECHTLICHE GELTENDMACHUNG

1. DISKRIMINIERUNGSVERBOT IM PRIVATBEREICH: ENGER ANWENDUNGSBEREICH
   - Art. 28 ZGB (Schutz der Persönlichkeit)
   - Art. 3 GIG (Gleichstellung im Erwerbsleben): Stellenausschreibung?
2. NACHWEIS EINER DISKRIMINIERUNG
   - M.E. kein Zugang zum Algorithmus notwendig
   - Ggf. *situational testing*
   - Aber: Beweislastumkehr nur im GIG
3. STAATLICHE SCHUTZPFLICHT

Art. 3*a*[19] Diskriminierung im Fernhandel

[1] Unlauter handelt insbesondere, wer im Fernhandel ohne sachliche Rechtfertigung einen Kunden in der Schweiz aufgrund seiner Nationalität, seines Wohnsitzes, des Ortes seiner Niederlassung, des Sitzes seines Zahlungsdienstleisters oder des Ausgabeorts seines Zahlungsmittels:

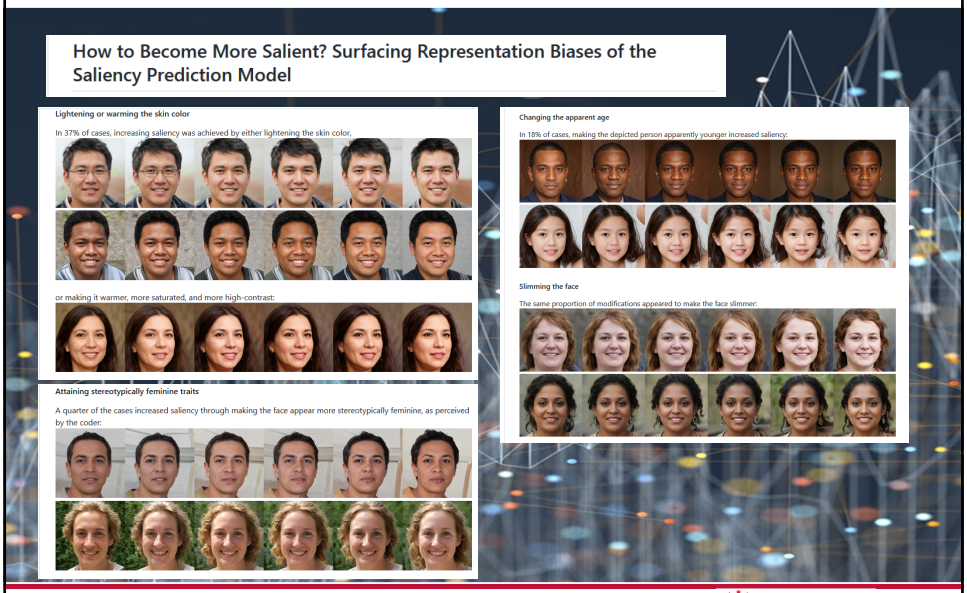
a. beim Preis oder bei den Zahlungsbedingungen diskriminiert;

b. ihm den Zugang zu einem Online-Portal blockiert beziehungsweise beschränkt; oder

c. ihn ohne sein Einverständnis zu einer anderen als der ursprünglich aufgesuchten Version des Online-Portals weiterleitet.

[2] Diese Bestimmung findet keine Anwendung auf nichtwirtschaftliche Dienstleistungen von allgemeinem Interesse; Dienstleistungen im Finanzbereich; Dienstleistungen der elektronischen Kommunikation; Dienstleistungen des öffentlichen Verkehrs; Dienstleistungen von Leiharbeitsagenturen; Gesundheitsdienstleistungen; Glücksspiele, die einen geldwerten Einsatz verlangen, einschliesslich Lotterien, Glücksspiele in Spielbanken und Wetten; private Sicherheitsdienste; soziale Dienstleistungen aller Art; Dienstleistungen, die mit der Ausübung hoheitlicher Gewalt verbunden sind; Tätigkeiten von Notaren sowie von Gerichtsvollziehern, die durch staatliche Stellen bestellt werden; audiovisuelle Dienste.

2. THEMENKREIS «FALSCHERKENNUNG»



DISKRIMINIERUNG IM JURISTISCHEN SINN?

1. BENACHTEILIGUNG?
   INDIVIDUELLE BETROFFENHEIT?

DISKRIMINIERUNG IM JURISTISCHEN SINN?

1. BENACHTEILIGUNG?
   INDIVIDUELLE BETROFFENHEIT?
2. GGF. STAATLICHE SCHUTZPFLICHTEN
   - z.B. Art. 5 CEDAW

Art. 5

Die Vertragsstaaten treffen alle geeigneten Massnahmen,

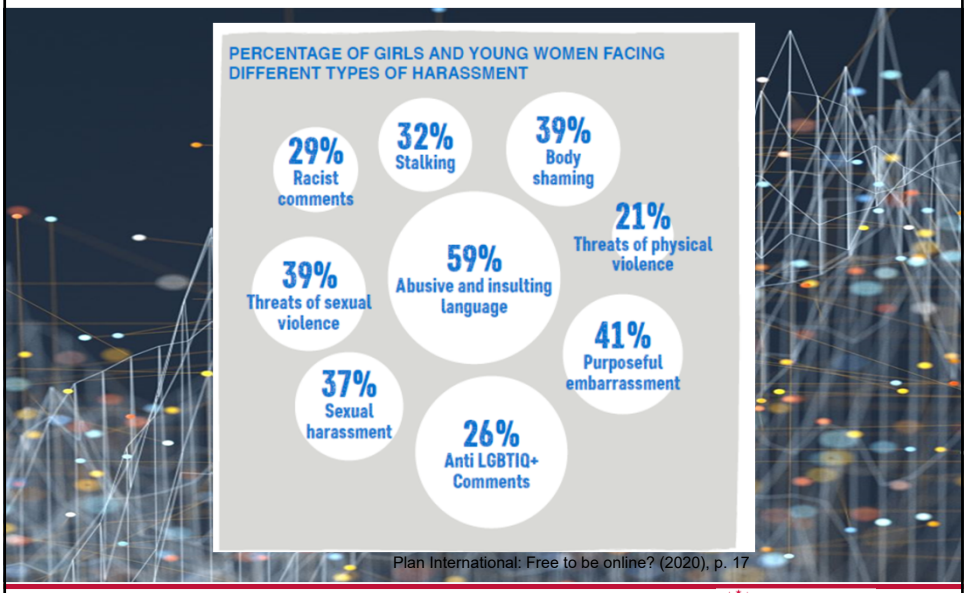
a) um einen Wandel in den sozialen und kulturellen Verhaltensmustern von Mann und Frau zu bewirken, um so zur Beseitigung von Vorurteilen sowie von herkömmlichen und allen sonstigen auf der Vorstellung von der Unterlegenheit oder Überlegenheit des einen oder anderen Geschlechts oder der stereotypen Rollenverteilung von Mann und Frau beruhenden Praktiken zu gelangen;

DISKRIMINIERUNG IM JURISTISCHEN SINN?

1. BENACHTEILIGUNG?
   INDIVIDUELLE BETROFFENHEIT?
2. GGF. STAATLICHE SCHUTZPFLICHTEN
   - z.B. Art. 8 Abs. 2 lit. c BRK («Bewusstseinsbildung»

c) die Aufforderung an alle Medienorgane, Menschen mit Behinderungen in einer dem Zweck dieses Übereinkommens entsprechenden Weise darzustellen;

3. THEMENKREIS «CYBERGEWALT»



Plan International: Free to be online? (2020), p. 17

GEWALT ALS FORM DER DISKRIMINIERUNG

1. **ALLGEMEIN**
   - CEDAW General Recommendation No. 35 (2017) updating No. 19 (1992)
   - EGMR *Opuz* v. Turkey (2009); *Dordevic* v. Croatia (2012); *Identoba* et al. v. Georgia (2015) u.a.
   - Istanbul-Konvention (2011)
2. **IM NETZ**
   - GREVIO General Recommendation No. 1 (2021) on on the digital dimension of violence against women
   - EGMR Urteile *Buturugă* v. Romania (2020); *Volodina* v. Russia no. 2 (2021)

SCHUTZPFLICHTEN: «DUE DILIGENCE»

CEDAW GENERAL RECOMMENDATION NO. 35 (2017)

incentives, including economic incentives.[38] Under the obligation of due diligence, States parties must adopt and implement diverse measures to tackle gender-based violence against women committed by non-State actors, including having laws, institutions and a system in place to address such violence and ensuring that they function effectively in practice and are supported by all State agents and bodies who diligently enforce the laws.[39] The failure of a State party to take all appropriate measures to prevent acts of gender-based violence against women in cases in which its authorities are aware or should be aware of the risk of such violence, or the failure to investigate, to prosecute and punish perpetrators and to provide reparations to victims/survivors of such acts, provides tacit permission or encouragement to perpetrate acts of gender-based violence against women.[40] Such failures or omissions constitute human rights violations.

GREVIO GENERAL RECOMMENDATION NO. 1 (2021) ON ON THE DIGITAL DIMENSION OF VIOLENCE AGAINST WOMEN

established concept of due diligence adopted by international and regional human rights instruments, policy documents and jurisprudence is framed as an obligation of means, not of results, and requires States Parties to set up the necessary legal and policy framework to allow for the prevention of all forms of violence against women and their effective investigation in order to hold perpetrators accountable for their action and to compensate victims. It is a provision of the Istanbul Convention that is central to ending impunity for gender-based violence against women and to ensuring access to justice for women and girl victims of such violence. GREVIO considers this obligation to cover all expressions of violence against women, including digital expressions and violence perpetrated with the help of or through technology. Current experiences of women and girls of such violence show that too little is done to hold perpetrators to account. As noted by the Council of Europe Commissioner for Human Rights, the lack of awareness about this issue causes cyberattacks and violence against women to not be taken as seriously as offline violence by national authorities.[17] Moreover, law-enforcement agencies and judicial authorities often lack the necessary technical training to be able to investigate and prosecute such incidents of violence effectively.

GREVIO EMPFEHLUNGEN (AUSWAHL)

1. PREVENTION
   - Eradicate gender **stereotypes** and sexist attitudes that play online and offline
   - Support **empowerment** and representation of women online by enhancing digital literacy and participation
   - Incorporate digital manifestations of violence against women in any existing intervention programmes for **perpetrators** of violence
   - **Encourage the ICT sector and internet intermediaries, including social media platforms**, to make an active effort to **avoid gender bias in the design** of smart products, mobile phone applications and video games, as well as the development of **artificial intelligence** and - respectively - to create internal monitoring mechanisms towards **ensuring the inclusion of victim-centric perspectives** as well as to advocate stronger awareness of the perspective and experiences of **female users**, in particular those exposed to or at risk of intersecting forms of discrimination. Internet intermediaries as well as technology companies should be **incentivised** to co-operate with NGOs working on violence against women in their awareness-raising and other efforts

GREVIO EMPFEHLUNGEN (AUSWAHL)

2. PROTECTION
   - ensure that the **legal framework** relating to violence against women refers and applies to all forms of violence committed in the digital sphere
   - develop and **disseminate accessible information** on the legal avenues and support services available to victims of violence against women perpetrated in the digital sphere and create online and offline complaints mechanisms that are easily and immediately accessible to victims
   - make **support services**, including legal and psychological counselling accessible to all victims of violence against women perpetrated in the digital sphere
   - **incentivise** internet intermediaries including ISPs, search engines and **social media platforms** to ensure robust **moderation** of content that falls within the scope of the Istanbul Convention through **removal** of account or content, in multiple languages on the basis of **transparent** principles that protect the human rights of all, including women's right to live free from violence and to provide easily accessible user guidance to **flag** abusive content and request its removal

GREVIO EMPFEHLUNGEN (AUSWAHL)

3. PROSECUTION
   - equip **law enforcement** and other criminal justice actors with the **necessary human, financial and technical resources** to effectively investigate and prosecute the digital dimension of violence against women in line with their due diligence obligation
   - increase **capacity-building efforts** for criminal justice and law-enforcement professionals to equip them with the necessary expertise and resources on how to use existing legal frameworks to address the digital dimension of violence against women, as well as to develop their **forensic capabilities** on the gathering and securing of electronic evidence without causing secondary victimization and re-traumatisation of the victim. International co-operation and mutual legal assistance
   - take measures to put an end to impunity for digital acts of violence against women by **encouraging the responsibility of all relevant actors, including ICT companies and internet intermediaries**, in particular through robust content moderation and removal; and by encouraging media companies to work collaboratively with law-enforcement agencies

FAZIT

- «Discrimination by Design» / «algorithmic harms are not only «bugs»
- Strukturelle Konsequenzen sind häufig folgenschwerer als die Beeinträchtigung im Einzelfall
- Rückkoppelungseffekte (weiterhin zunehmende Datenverknüpfungen, chilling effect…)
- Schwierigkeiten bei der gerichtlichen Geltendmachung, u.a. aufgrund engem Anwendungsbereich, fehlender Beweislastumkehr
- Menschenrechtlich begründete Schutzpflichten, deren Umsetzung aber noch weitgehend unklar ist


VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!
FACULTE DE DROIT
UNIVERSITE DE FRIBOURG
RECHTSWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT FREIBURG
Institut für Europarecht
Institut de droit européen
UNI
FR