



Compte amb la bretxa: dades de gènere i biaix d'IA

La bretxa de dades de gènere = dades que falten/desiguals sobre les dones. Com a resultat, les decisions de la IA esbiaixen els homes a través de biaixos representatius, algorítmics, culturals i interseccionals.

PER QUÈ PASSA?

- La majoria dels equips d'IA són homes; les opcions de disseny reflecteixen els seus punts febles.
- Les dades d'entrenament infrarepresenten les dones.
- Optimitzar només per a la precisió o el benefici general pot marginar les minories.
- L'accés al telèfon/Internet de les dones és més baix → menys rastres de dades.
- Les etiquetes binàries i les mitjanes amaguen els subgrups.

CASOS PER TIPUS DE BIAIX

Biaix de representació

- Reconeixement facial: 35% d'errors per a dones de pell més fosca (0,8% homes de pell més clara).
- Imatges de treball DALL-E 2: dones 38% (homes 62%); dones somrients 2,2 vegades més.
- Imatges STEM: 75–100% homes.

Biaix algorítmic

- Contractació: formació amb biaix masculí → currículums de dones amb una classificació inferior.
- Avaluació creditícia: dones menys/préstecs més petits malgrat un millor reemborsament.
- IA sanitària: passa per alt els símptomes de les dones → diagnòstic retardat.
- Recomanadors musicals: fins a +6. 3 pp més precisos per als homes.

Biaix cultural

- Assistents femenines per defecte → estereotip de "dones com a ajudants".
- Històries de Llama-2: dones en rols domèstics ~4 vegades més; els homes aconsegueixen treballs de més alt estatus.

Biaix interseccional

- Subtítols interseccionals de YouTube: més errors per a les dones.
- Anuncis de treball STEM: homes ~+20% d'impressions.
- Recomanadors de llibres: menys exposició per a autores dones.

TANCANT LA BRETXA: QUÈ CAL FER

Biaix representatiu

- Conjunts de dades inclusius; documentar i publicar cobertura; punts de referència interseccionals.

Biaix algorítmic

- Eliminar els marcadors de gènere de les entrades; ajustar els llindars de decisió per equilibrar els errors; biaix en la formació i el postprocessament; monitorització contínua.

Biaix cultural

- Veus per defecte neutres/no binàries; comprovació de disseny d'estereotips; equips diversos en el disseny i les proves.

Biaix interseccional

- Publicar mètriques de subgrups (gènere, to de pell, edat, dialecte); auditories prèvies/posteriors al desplegament; classificar recursos humans/finances com a d'alt risc; declaracions d'impacte de gènere; estàndards de transparència.

Públic objectiu: 📖 Personal de biblioteques, 🎓 Professorat, 🎓 Estudiantat

REFERÈNCIES

Jobin, Anna, Marcello Lenca, and Effy Vayena. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1, no. 2 (September 2019). <https://doi.org/10.1038/s42256-019-0088-2>

UNESCO and IRCAI. "Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models", 2024. <https://unesdoc.unesco.org/ark:/48223/pf0000388971>

Context del projecte GEDIS

Aquest recurs forma part del projecte europeu GEDIS (Gender Diversity in Information Science), que promou eines educatives obertes per abordar les desigualtats de gènere en l'educació superior, amb un èmfasi especial en les disciplines relacionades amb la informació i la documentació. Aquest material educatiu s'ha desenvolupat en el marc de la Summer School Barcelona, dins del projecte GEDIS.

GEDIS - Gender Diversity in Information Science: Challenges in Higher Education. <https://ub.edu/gedis>

Citació: Bosshammer, Svetlana and Daniela Varbanova. 2025. *Mind the Gap: Gender Data and AI Bias*. DOI: 10.5281/zenodo.17164102. *Compte amb la bretxa: dades de gènere i biaix d'IA*. Traduït per Claudia San José i Juan-José Boté-Vericad. DOI: 10.5281/zenodo.17316245

Cofinançat per la Unió Europea. Les opinions i punts de vista expressats són únicament responsabilitat del/de la/dels/les autor(s/es) i no reflecteixen necessàriament els de la Unió Europea ni els del Servei Espanyol per a la Internacionalització de l'Educació (SEPIE). Ni la Unió Europea ni l'autoritat concedent no se'n poden considerar responsables.

