

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
12 December 2024 (12.12.2024)



(10) International Publication Number
WO 2024/253705 A1

(51) International Patent Classification:
G06N 20/20 (2019.01) *G06N 3/0895* (2023.01)
G06N 3/091 (2023.01)

(21) International Application Number:
PCT/US2023/086020

(22) International Filing Date:
27 December 2023 (27.12.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/507,381 09 June 2023 (09.06.2023) US

(71) Applicant: **VISA INTERNATIONAL SERVICE ASSOCIATION** [US/US]; PO Box 8999, San Francisco, California 94128 (US).

(72) Inventors: **ISLAM, Rashidul**; PO Box 8999, San Francisco, California 94128 (US). **CAI, Yiwei**; PO Box 8999, San Francisco, California 94128 (US). **YANG, Hao**; PO Box 8999, San Francisco, California 94128 (US). **UDDIN, Md**

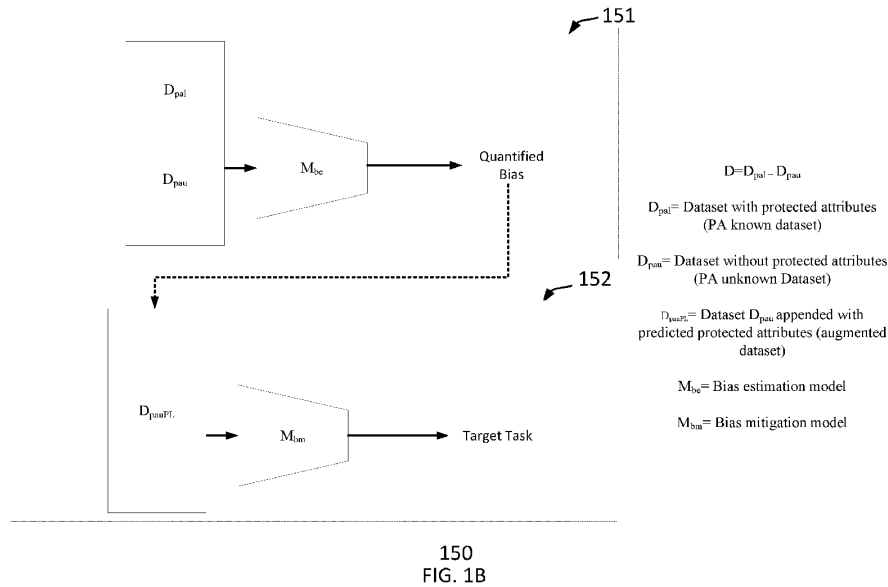
Taufeeq; PO Box 8999, San Francisco, California 94128 (US).

(74) Agent: **SHARMA, Vaibhav** et al.; Kilpatrick Townsend & Stockton LLP, Mailstop: IP Docketing – 22, 1100 Peachtree Street, Suite 2800, Atlanta, Georgia 30309 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,

(54) Title: MACHINE LEARNING FAIRNESS WITH LIMITED PROTECTED ATTRIBUTES



150
FIG. 1B

(57) Abstract: A two stage model allows for fairness of machine learning models to be increased despite limited access to categories of information (e.g., protected attributes) in the underlying data on which the models are trained. A first stage is used to train a protected attributes estimation model through active learning. A second stage is used to supplement data not containing protected attributes with estimates from the trained protected attributes estimation model and allow for a bias mitigation model to be trained with increased less bias.

[Continued on next page]

SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*