



(51) International Patent Classification:
G10L 17/02 (2013.01)

(21) International Application Number:
PCT/US2021/055267

(22) International Filing Date:
15 October 2021 (15.10.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/092,956 16 October 2020 (16.10.2020) US

(71) Applicant: **PINDROP SECURITY, INC.** [US/US]; 817
West Peachtree Street, NW, Suite 770, Atlanta, Georgia
30308 (US).

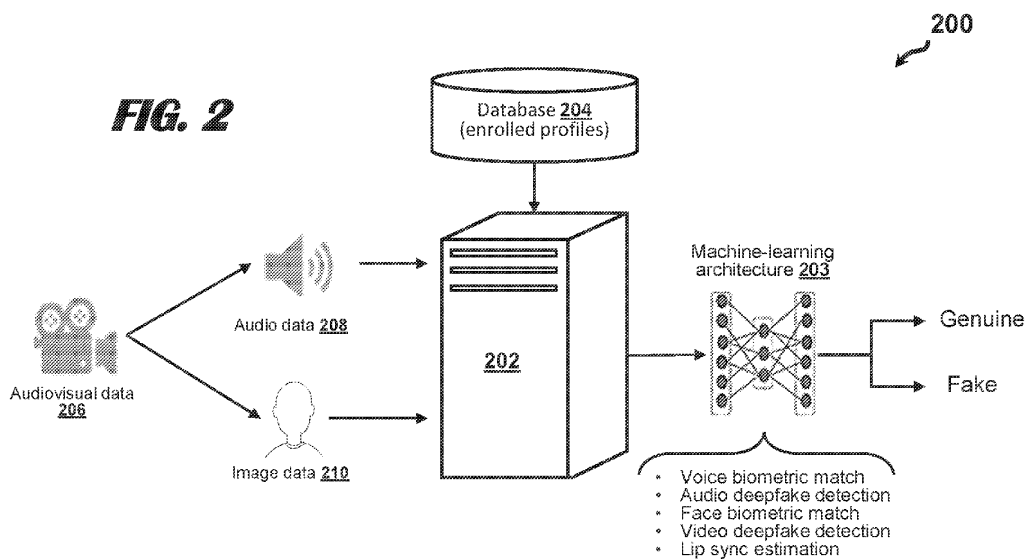
(72) Inventors: **CHEN, Tianxiang**; 817 WEST PEACHTREE
STREET, NW, SUITE 770, ATLANTA, Georgia 30308

(US). **KHOURY, Elie**; 817 WEST PEACHTREE
STREET, NW, SUITE 770, ATLANTA, Georgia 30308
(US).

(74) Agent: **SOPHIR, Eric** et al.; FOLEY & LARDNER LLP,
3000 K Street N.W., Washington, District of Columbia
20007 (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, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
ME, MG, MK, MN, 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.

(54) Title: AUDIOVISUAL DEEPFAKE DETECTION



(57) **Abstract:** The embodiments execute machine-learning architectures for biometric-based identity recognition (*e.g.*, speaker recognition, facial recognition) and deepfake detection (*e.g.*, speaker deepfake detection, facial deepfake detection). The machine-learning architecture includes layers defining multiple scoring components, including sub -architectures for speaker deepfake detection, speaker recognition, facial deepfake detection, facial recognition, and lip-sync estimation engine. The machine-learning architecture extracts and analyzes various types of low-level features from both audio data and visual data, combines the various scores, and uses the scores to determine the likelihood that the audiovisual data contains deepfake content and the likelihood that a claimed identity of a person in the video matches to the identity of an expected or enrolled person. This enables the machine-learning architecture to perform identity recognition and verification, and deepfake detection, in an integrated fashion, for both audio data and visual data.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, 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, 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 applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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