

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
06 October 2022 (06.10.2022)



(10) International Publication Number
WO 2022/212852 A1

(51) International Patent Classification:
G06F 40/30 (2020.01) *H04L 51/00* (2022.01)
G06N 3/02 (2006.01)

(21) International Application Number:
PCT/US2022/023065

(22) International Filing Date:
01 April 2022 (01.04.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/170,240 02 April 2021 (02.04.2021) US

(71) Applicant: **LIVEPERSON, INC.** [US/US]; 475 10TH AVENUE, 5TH FLOOR, NEW YORK, New York 10018 (US).

(72) Inventors: **HOWELL, Kristen**; 475 10th Avenue, 5th Floor, New York, New York 10018 (US). **WANG, Jian**; 475 10th Avenue, 5th Floor, New York, New York 10018 (US).

(US). **DUNN, Matthew**; 475 10TH Avenue, 5th Floor, New York, New York 10018 (US). **BRADLEY, Joseph**; 475 10th Avenue, 5th Floor, New York, New York 10018 (US).

(74) Agent: **JENSEN, Philip D.** et al.; POLSINELLI PC/LLP, Three Embarcadero Center, Suite 2400, San Francisco, California 94111 (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, JM, 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: DOMAIN ADAPTATION OF AI NLP ENCODERS WITH KNOWLEDGE DISTILLATION

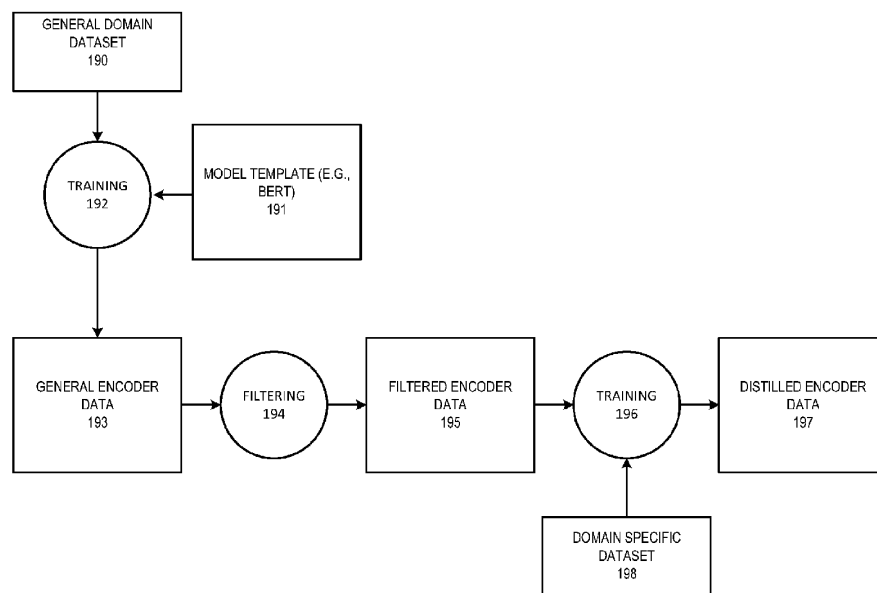


FIG. 1C

(57) Abstract: Systems, methods, devices, instructions, and other examples are described for natural language processing. One example includes accessing natural language processing general encoder data, where the encoder data is generated from a general-domain dataset that is not domain specific. A domain specific dataset is accessed and filtered encoder data using a subset of the encoder data is generated. The filtered encoder data is trained using the domain specific dataset to generate distilled encoder data, and tuning values for the distilled encoder data are generated to configure task outputs associated with the domain specific dataset.

[Continued on next page]

(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).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*