

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
14 January 2021 (14.01.2021)



(10) International Publication Number
WO 2021/006861 A1

(51) International Patent Classification:
H04M 3/42 (2006.01)

(21) International Application Number:
PCT/US2019/040722

(22) International Filing Date:
05 July 2019 (05.07.2019)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **VISA INTERNATIONAL SERVICE ASSOCIATION** [US/US]; 900 Metro Center Boulevard, Foster City, California 94404 (US).

(72) Inventors: **SONG, Hongqin**; 8732 Wafer Ash Way, Austin, Texas 78750 (US). **GU, Yu**; 5 Grove Ct., Austin, Texas 78746 (US).

(74) Agent: **JEWIK, Patrick** 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, 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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(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,

(54) Title: METHOD AND SYSTEM USING AI CALL PREDICTION AND CACHE

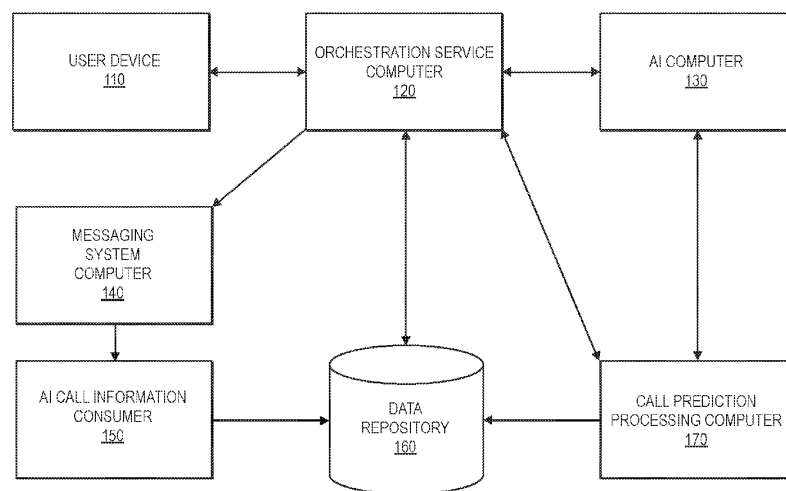


FIG. 1

(57) Abstract: Methods and systems of predicting and processing cacheable AI calls are disclosed. One method includes determining, by a call prediction processing computer, that an AI computer is operating below a threshold processor usage. The method then includes requesting, by the call prediction processing computer, a set of cacheable requests from an orchestration service computer and receiving, from the orchestration service computer, the set of cacheable requests. Then the method includes sending, by the call prediction processing computer, a cacheable request from the set of cacheable requests to the AI computer. The method then includes receiving, by the call prediction processing computer, an output. The output is generated by the AI computer based on the cacheable request. Then the call prediction processing computer stores the output in a data repository.

[Continued on next page]

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))*