



(51) International Patent Classification:

G06F 21/50 (2013.01) G06F 40/56 (2020.01)
G06F 40/30 (2020.01) G06N 7/01 (2023.01)
G06F 40/35 (2020.01)

(21) International Application Number:

PCT/EP2024/087712

(22) International Filing Date:

19 December 2024 (19.12.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/612,405	20 December 2023 (20.12.2023)	US
18/650,276	30 April 2024 (30.04.2024)	US
18/650,312	30 April 2024 (30.04.2024)	US

(71) Applicant: **BITDEFENDER IPR MANAGEMENT LTD** [CY/CY]; 59-61 Acropolis ave. 3rd floor, office 302, Nicosia, 2012 (CY).

(72) Inventors: **CERNAT, Constantin M.**; Str. Ariesul Mare, nr. 7, bl 31, ap 33, sect.6, 060792 BUCURESTI (RO).
CRACIUN, Ovidiu M.; Str. General Alexandru Cernat nr. 8, 011059 BUCURESTI (RO).

(74) Agent: **TULUCA, Doina**; Bdul. Lacul Tei 56, Bl 19, Sc B, Ap 52, sector 2, 020392 BUCURESTI (RO).

(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, 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).

(54) Title: LOGICAL LANGUAGE MODEL PROMPTS FOR FRAUD DETECTION

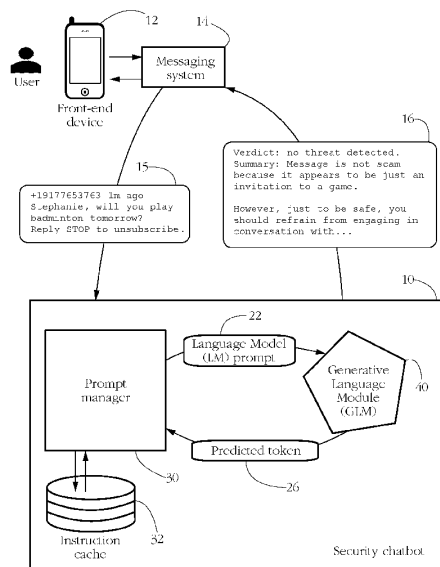


FIG. 1

(57) Abstract: In some embodiments, a security chatbot comprises a generative language module (GLM) and a prompt manager configured to dynamically update the GLM input prompt according to the output of the GLM. The input prompt instructs the GLM to carry out a task (e.g., determine whether a target message received by a user is indicative of online fraud) and in response, to output a specific flag token. In response to detecting the flag token within the output of the GLM, the prompt manager may selectively execute a code snippet identified according to the respective flag token. Executing the code snippet effectively updates the prompt, which is then fed back to the GLM for the next inference step. In contrast to conventional LLM prompting, updates to the prompt herein comprise modifications (e.g., insertion of other flag tokens, etc.) other than merely appending an inferred token to the previous prompt. Furthermore, the prompt modifications are performed inline, i.e., without submitting an entirely new prompt to the GLM.

Published:

- *with international search report (Art. 21(3))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*