



(51) International Patent Classification:

G06N 3/04 (2006.01) G06V 10/80 (2022.01)
G06N 3/08 (2006.01) G06V 10/82 (2022.01)

(21) International Application Number:

PCT/EP2022/065502

(22) International Filing Date:

08 June 2022 (08.06.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

20210100374 08 June 2021 (08.06.2021) GR

(71) Applicant: **DEEPMIND TECHNOLOGIES LIMITED**
[GB/GB]; 5 New Street Square, London EC4A 3TW (GB).

(72) Inventors: **TSIMPOUKELLI, Maria Rafailia**; 6 Pancras Square, London N1C 4AG (GB). **MENICK, Jacob Lee**; 6 Pancras Square, London N1C 4AG (GB). **CABI, Serkan**; 6 Pancras Square, London N1C 4AG (GB). **HILL, Felix George**; 6 Pancras Square, London N1C 4AG (GB).

ESLAMI, Seyed Mohammadali; 6 Pancras Square, London N1C 4AG (GB). **VINYALS, Oriol**; 6 Pancras Square, London N1C 4AG (GB).

(74) Agent: **FISH & RICHARDSON P.C.**; Highlight Business Towers, Mies-van-der-Rohe-Str. 8, 80807 Munich (DE).

(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, IQ, 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.

(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: MULTIMODAL FEW-SHOT LEARNING WITH FROZEN LANGUAGE MODELS

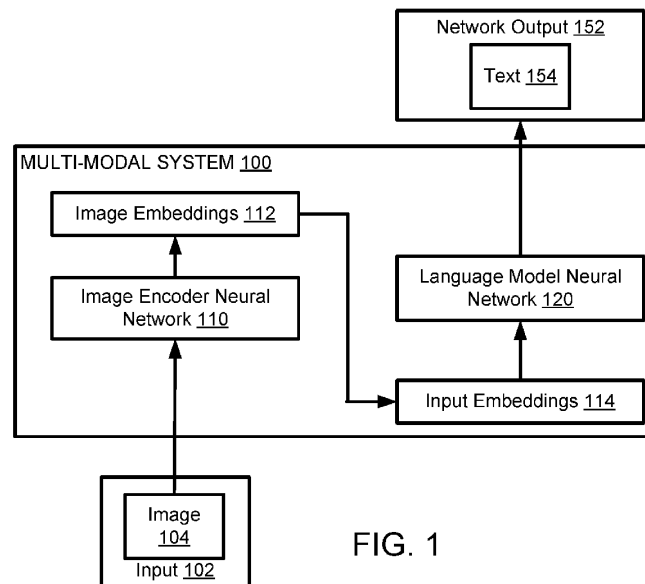


FIG. 1

(57) Abstract: Methods, systems, and apparatus, including computer programs encoded on computer storage media, for processing multi-modal inputs using language models. In particular, the inputs include an image, and the image is encoded by an image encoder neural network to generate a sequence of image embeddings representing the image. The sequence of image embeddings is provided as at least part of an input sequence to that is processed by a language model neural network.

[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))*
- *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))*