

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
12 September 2025 (12.09.2025)



(10) International Publication Number
WO 2025/188233 A1

(51) International Patent Classification:
G06F 21/55 (2013.01) *G06N 3/00* (2023.01)

(21) International Application Number:
PCT/SG2024/050129

(22) International Filing Date:
04 March 2024 (04.03.2024)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **ST ENGINEERING IHQ PTE. LTD.**
[SG/SG]; 1 Ang Mo Kio Electronics Park Road, Singapore
567710 (SG).

(72) Inventors: **DAI, Zhongmin**; c/o ST Engineering IHQ Pte.
Ltd, 1 Ang Mo Kio Electronics Park Road, Singapore
567710 (SG). **THING, Ling Ling, Vrizlynn**; c/o ST Engi-
neering IHQ Pte. Ltd, 1 Ang Mo Kio Electronics Park Road,
Singapore 567710 (SG).

(74) Agent: **YUSARN AUDREY LLC**; 4 Shenton Way,
#14-03, SGX Centre 2, 068807 Singapore (SG).

(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,

(54) Title: A SECURED ARTIFICIAL INTELLIGENCE SYSTEM AND A SECURITY ASSESSMENT SYSTEM FOR
ARTIFICIAL INTELLIGENCE MODEL

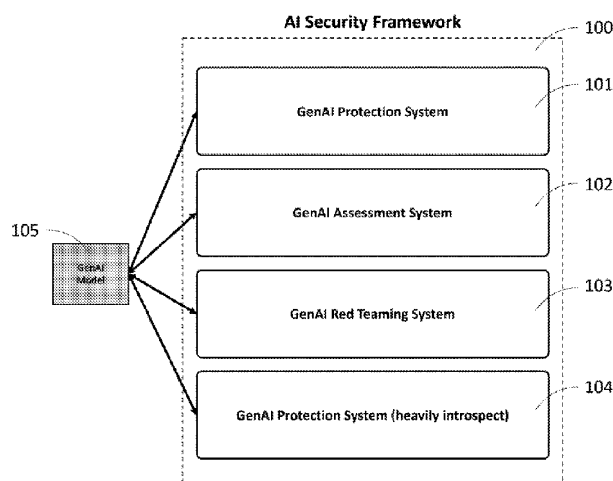


FIG. 1

(57) Abstract: The present invention introduces an artificial intelligence (AI) system that includes a model, a first model protection system, and a wrapper. The system is designed to enhance security by processing inputs and outputs through a protective layer. Specifically, inputs from users are initially directed by the wrapper to the first model protection system, which then hardens these inputs before they are fed into the model to produce outputs. Subsequently, the outputs are received by the wrapper and passed through the first model protection system again to harden the outputs before they are delivered to users. Additionally, the invention introduces a security assessment system for evaluating the potential risks associated with a target model before its deployment, comprising an assessment system, a red teaming system, and a second model protection system, collectively working to thoroughly assess and mitigate potential security threats. The present invention offers a comprehensive security framework for AI models, covering both the development and deployment phases, and addresses the escalating security challenges facing AI models.

[Continued on next page]

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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *of inventorship (Rule 4.17(iv))*

Published:

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