(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2026 年 8 月 27 日 (27.08.2026)

WIPO | PCT

(10) 国际公布号
WO 2026/174685 A1

(51) 国际专利分类号:
G06F 16/21 (2019.01) *G06F 9/455* (2018.01)

(21) 国际申请号: PCT/CN2025/098000

(22) 国际申请日: 2025 年 5 月 29 日 (29.05.2025)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202510193916.1 2025年2月21日 (21.02.2025) CN

(71) 申请人:无限星途科技私人有限公司(**X STAR TECHNOLOGY PTE. LTD.**) [SG/SG]; 新加坡, 新达城1号楼, 37-03, 淡马锡林荫道7号 038987 (SG)。

(72) 发明人:田立民(**TIAN, Limin**); 中国上海市浦东新区中国(上海)自由贸易试验区浦东大道2123号三层 200135 (CN)。 唐守旭(**TANG, Shouxu**); 中国上海市浦东新区中国(上海)自由贸易试验区浦东大道2123号三层 200135 (CN)。 张磊(**ZHANG, Lei**); 中国上海市浦东新区中国(上海)自由贸易试验区浦东大道2123号三层 200135 (CN)。 杨青相(**YANG, Qingxiang**); 中国上海市浦东新区中国(上海)自由贸易试验区浦东大道2123号三层 200135 (CN)。 吴昊晨(**WU, Haochen**); 中国上海市浦东新区中国(上海)自由贸易试验区浦东大道2123号三层 200135 (CN)。 刘兵(**LIU, Bing**); 中国上海市浦东新区中国(上海)自由贸易试验区浦东大道2123号三层 200135 (CN)。

(74) 代理人:北京品源专利代理有限公司(**BEYOND ATTORNEYS AT LAW**); 中国北京市海淀区莲花池东路39号西金大厦6层 100036 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): 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,

(54) **Title:** VERSION MANAGEMENT METHOD AND APPARATUS FOR LARGE LANGUAGE MODEL, AND DEVICE AND MEDIUM

(54) 发明名称: 大语言模型的版本管理方法、装置、设备及介质

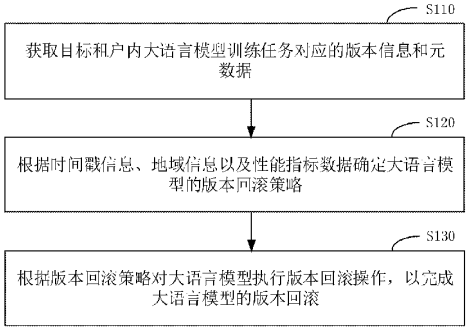


图 1

S110 Acquire version information and metadata which correspond to a large language model training task within a target tenant
S120 On the basis of timestamp information, region information and performance indicator data, determine a version rollback strategy for the large language model
S130 On the basis of the version rollback strategy, execute a version rollback operation on the large language model, so as to complete the version rollback of the large language model

(57) **Abstract:** A version management method and apparatus for a large language model, and a device and a medium. The method comprises: acquiring version information and metadata which correspond to a large language model training task within a target tenant, wherein the version information comprises timestamp information and region information, and the metadata comprises performance indicator data; on the basis of the timestamp information, the region information and the performance indicator data, determining a version rollback strategy for the large language model; and on the basis of the version rollback strategy, executing a version rollback operation on the large language model, so as to complete the version rollback of the large language model.

(57) 摘要: 一种大语言模型的版本管理方法、装置、设备及介质，其中，该方法包括：获取目标租户内大语言模型训练任务对应的版本信息和元数据；其中，版本信息包括时间戳信息和地域信息；所述元数据包括性能指标数据；根据所述时间戳信息、所述地域信息以及所述性能指标数据确定大语言模型的版本回滚策略；根据所述版本回滚策略对所述大语言模型执行版本回滚操作，以完成所述大语言模型的版本回滚。

WO 2026/174685 A1

[见续页]

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, UY, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (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)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。