

(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 3 月 5 日 (05.03.2026)



(10) 国际公布号
WO 2026/045028 A1

- (51) 国际专利分类号:
G06F 16/35 (2025.01)
- (21) 国际申请号: PCT/CN2024/140675
- (22) 国际申请日: 2024 年 12 月 19 日 (19.12.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202411212294.4 2024年8月30日 (30.08.2024) CN
- (71) 申请人: 无限光年(上海)技术有限公司(INFLY TECH (SHANGHAI) CO., LTD.) [CN/CN]; 中国上海市徐汇区虹漕路25-1号2层1325室 200233 (CN)。
- (72) 发明人: 褚巍 (CHU, Wei); 中国上海市徐汇区虹漕路25-1号2层1325室 200233 (CN)。
- (74) 代理人: 北京智信禾专利代理有限公司(CHISPO ATTORNEYS AT LAW); 中国北京市东城区崇文门外大街8号院哈德门广场西塔901 100062 (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, 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) 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,

(54) Title: DATA PROCESSING METHOD AND APPARATUS, AND MODEL OPTIMIZATION METHOD AND APPARATUS

(54) 发明名称: 数据处理方法及装置、模型优化方法及装置

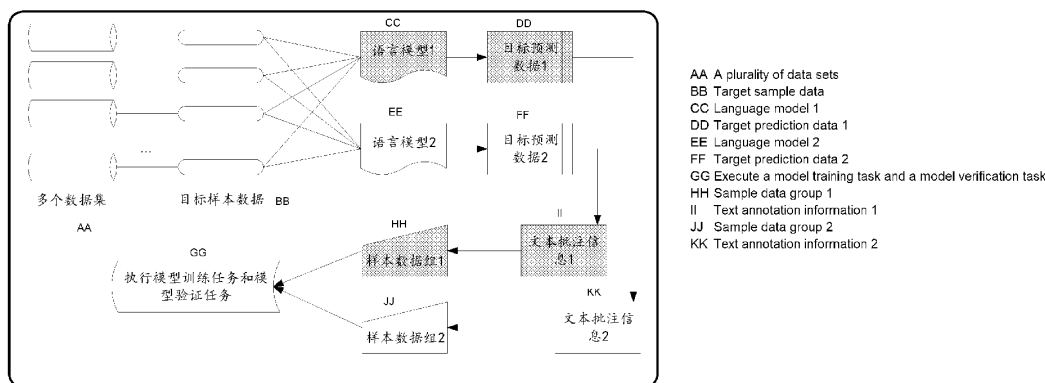


图 1

(57) Abstract: Provided are a data processing method and apparatus, and a model optimization method and apparatus. The data processing method comprises: respectively sampling in a plurality of data sets so as to determine target sample data on the basis of sampling results; inputting the target sample data into a plurality of language models for respectively processing to obtain target prediction data respectively outputted by the plurality of language models; using a preset text annotation model to perform text annotation processing on each piece of target prediction data to obtain text annotation information corresponding to each piece of target prediction data; and constructing a sample data group on the basis of the target sample data, the target prediction data, and the text annotation information, wherein the sample data group is used for executing a model training task and a model verification task.

[见续页]

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))。

(57) 摘要: 提供了一种数据处理方法及装置、模型优化方法及装置, 其中所述数据处理方法包括: 在多个数据集中分别进行采样, 根据采样结果确定目标样本数据; 将所述目标样本数据输入至多个语言模型分别进行处理, 获得所述多个语言模型分别输出的目标预测数据; 利用预设的文本批注模型对每个目标预测数据进行文本批注处理, 获得每个目标预测数据对应的文本批注信息; 基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组, 其中, 所述样本数据组用于执行模型训练任务和模型验证任务。